Surveillance of Sexually Transmitted Infections:

Exploring the Utility of Internet Search Engine Data

BY

AMY KRISTEN JOHNSON
B.A., DePaul University, 2003
M.S.W., University of Illinois at Chicago, 2006

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public
Health Sciences in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

Supriya Mehta, MHS PhD, Chair and Advisor
Ronald Hershow, MD
Frederick Kviz, PhD, Community Health Sciences
Mark Dworkin, MD
Kyle Bernstein, PhD, Centers for Disease Control and Prevention

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee- Supriya Mehta, Ronald Hershow, Frederick Kviz, Kyle Bernstein and Mark Dworkin, for all of their support. I would also like to thank Christian Stefansen, for his guidance on using Google Trends and for his previous contributions to this field of study.

I am very lucky to have many mentors. Geri Donenberg and Erin Emerson, thank you for hiring me so many years ago, for answering all my questions and for teaching me about research from proposal to dissemination. Lisa Kuhns, thank you for your unwavering support, encouragement and the advice to "just get it done." Rob Garofalo, thank you for teaching me to take risks and to dream big. Supriya Mehta, thank you for all the deadlines, reviews, emails, and conversations; for setting the bar high and always asking tough questions; also thank you for believing in this idea and encouraging me! I am forever grateful for each one of you.

To my husband, thank you for your non-stop love and support, for not letting me quit, and for all the sacrifices. I love you.

To my children, who know no other way: work hard, study, and do what you love.

To my mother and father, for making education a priority, for believing in me, and for supporting me (and –my goodness--all those lost library books!). Thank you.

This study would not have been possible without the support of partner agencies: the Chicago Department of Public Health, the National Coalition for STD Directors, and the National Association of City and County Health Officials. I would also like to acknowledge and thank the individuals who participated in the studies.

AKJ

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
CASI	Computer Assisted Self Interviewing
CDC	Centers for Disease Control and Prevention
CDPH	Chicago Department of Public health
EMR	Electronic Medical Record
GISP	Gonococcal Isolates Surveillance Project
HIV	Human Immunodeficiency Virus
MSM	Men who have sex with men
PID	Pelvic Inflammatory Disease
P&S	Primary and Secondary (Syphilis)
STD	Sexually Transmitted Disease
STI	Sexually Transmitted Infection
TB	Tuberculosis
US	United States of America
WHO	World Health Organization

SUMMARY

National surveillance of sexually transmitted infections (STIs) relies on mandatory case reporting, a system that produces data that is often incomplete and limited in scope. In addition to data quality issues, underreporting of cases as well as delays in reporting results in missed opportunities to identify and respond to trends in disease as well as limited ability to guide STI control.

As the internet is a portal for free and anonymously available health information, search engine data may provide an additional venue for surveillance efforts leading to earlier detection of trends and increased ability to monitor impact and geographic spread. Search engine data has the potential to be an efficient and economical enhancement to the established surveillance system. New surveillance methods may allow for significant improvements, particularly in timeliness.

Google Trends allows the download of de-identified search engine data trends, which can be used to investigate the implications of trends in STI-related search terms in relation to STI rates. While there is much that remains unknown, such as search engine user characteristics, content of searches, and feasibility and acceptability of integration of the new method into public health systems, it is important to explore this innovative tool and its potential application to STI surveillance.

In study 1 we determined the utility of Google Trends search data as a potential surveillance method of sexually transmitted infections via a comparative analysis of STI-related search terms by rate of specific nationally notifiable diseases. Google Trends was used to graph the popularity of the search terms over time and to map the terms by US city and state. Our preliminary study indicated there is a correlation (r = 0.72, P < 0.01) between 2011 gonorrhea rates and gonorrhea search terms by US state. Subsequent analyses investigated combinations of multiple search terms and updated national surveillance data (2013). In an updated analysis, STI surveillance data for 2013 was compared to the volume of STI search term by US city and state. Significant positive correlations were found between STI search terms and rate of disease by state for gonorrhea, chlamydia, and syphilis.

In study 2 we determined the potential for integrating internet search trend surveillance into the public health STI surveillance system by surveying a sample of health department surveillance employees to assess the feasibility and acceptability of integrating Google trends into the current surveillance system. Overall, participants reported high levels of interest in a Google Trends application for STIs, with 72% reporting being very interested. However, just over half of the participants (55%) reported Google Trends for STIs would be very useful and only 34% reported they would be very likely to integrate the new tool into their current surveillance system. Surveillance system attributes that Google Trends could provide and that were rated very important by participants were: access to data in real time (94%), improved sharing of information (84%), automatic alerts for outbreaks (84%), and increased understanding of the disease (72%). In terms of challenges and barriers to integrating Google Trends into the current surveillance system, 39% reported that their department was too busy to integrate new tools, 12% reported that the current system does not need to be modified and 8% reported that Google Trends would not be valuable. Should there be an opportunity to provide Google Trends for STI surveillance, the current study provided insight into barriers and facilitators to integration.

In study 3 we determined search engine user characteristics by surveying specific exposure groups of individuals aged 18-35 to assess the internet search behaviors related to STIs and to determine the reason, timing, and content of searches. Google Correlate was used to validate search term content. Specific exposure groups included patients at Chicago Department of Public Health STD clinics and students at the University of Illinois at Chicago. In the STD clinic sample, Non-Hispanic Blacks were less likely than Non-Hispanic Whites to search online for STI information prior to coming to the clinic. Having Government-issued or no insurance, compared to having parent's insurance, was also associated with being less likely to search the internet for STI information. Factors not related to internet search were sex, age and risk behavior (condom use, number of partners, previous STI). In the student sample, participants who had a previous STI were more likely to have ever searched the internet for STI information. The student sample had a lower overall prevalence of searching for STI information, compared to the clinic population. There were also fewer differences in the reasons for searching in terms of demographics or risk behavior, compared to the clinic sample. No demographic characteristics predicted searching for STI information; however sexual risk behaviors were related to searching for STI information in the student sample. This finding demonstrates that in a lower risk population, those most at risk for acquiring (or being exposed to) STIs are generating STI-related search terms. Participants were asked to list the terms they use to search for STI related information. On average clinic participants queries were longer compared to student queries. Clinic participants were more likely to report using search terms that were related to symptomatology such as describing symptoms of STIs, while students were more likely to report searching for general information. These differences in search terms by subpopulation have implications for STI surveillance in populations at most risk for disease acquisition.

Study results allowed us to determine the overall utility and potential application of search engine data in surveillance of reportable STIs. While Google Trends has been used to accurately predict trends in several other infectious diseases, it has not yet been applied to STIs. Additionally, internet search engine user characteristics, search patterns and search content related to STIs were previously unknown.

The results of this study substantially contributes to a gap in the literature and informs next steps ranging from a full partnership with Google and bioinformatics specialists to develop a disease specific trend page to replication of our studies in nationally representative samples. Exploring this innovative tool has the potential to bring us one step closer to real-time surveillance of STIs, enhance understanding disease under surveillance and increase the simplicity and flexibility of the current surveillance system.

1. INTRODUCTION

1.1 Background and significance

There are estimated to be nearly 20 million new diagnoses of sexually transmitted infections (STIs) each year in the United States (1). Many cases remain undiagnosed and unreported, and as a result the true burden of STIs is likely much greater (1). The purpose of STI surveillance is to estimate the morbidity and mortality of disease as well as to enhance the ability to predict and respond to patterns of disease. There are disparities in rates of STIs by age, gender, race and region, timely and accurate detection of which can support effective prevention and control.

Seventy-two percent of internet users report looking online for health information within the past year, with 77% reporting they began their session at a search engine (2). As the internet is a widely accessed portal for health information, search engine data may provide an additional venue for surveillance efforts leading to earlier detection of trends, increased ability to monitor impact and geographic spread as well as an efficient and economical addition to the already established system. New surveillance methods may allow for significant improvements, particularly in timeliness. Google Trends generates easily accessible search engine data which can be used to investigate the implications of trends in STI-related search terms in relation to STI rates.

While there is much that remains unknown, such as user characteristics and feasibility and acceptability of integration of the new method into public health systems, it is important to explore this new tool

Below, an overview is provided on the epidemiology of reportable STIs; including their morbidity, cure, and prevention. In addition a review of the primary sources of surveillance data for reportable STIs, including advantages and weaknesses is presented. Finally, a summary of an emerging method is provided: use of internet search engine data for surveillance.

1.2 Overview of reportable sexually transmitted infections: Surveillance and Epidemiology

The Centers for Disease Control and Prevention (CDC) estimates the incidence of STIs to be approximately 20 million cases each year with an estimated prevalence of 110 million infections in the United States (1). In 2012, there were over a million cases of chlamydia reported, 300,000 cases of gonorrhea and 15,000 cases of primary and secondary syphilis (1). In 2012, case reports of chlamydia, gonorrhea and syphilis accounted for more than 1.3 million cases, which was >80% of all infectious disease notification to CDC (1).

These rates represent a significant public health challenge, as most sexually active people will be infected with an STI at some point in their lives (3). Half of the new infections are found among young people ages 15-24, yet this age group represents a minority of the population who are sexually active (~25%). Approximately 1 out of 4 women aged 14-19 has been infected with at least one STI (1).

However, STIs are not limited to the young; rates are rising in older adults including the 50+ age group. Not only do STIs increase the risk for HIV infection, but if left untreated, they can also lead to reproductive health complications.

Additionally there are significant cost implications for these prevalent infectious diseases, resulting in almost 16 billion U.S. dollars in health care costs every year (3).

A notifiable disease is a disease for which regular, accurate and timely information about cases is needed for the management of the disease (e.g., prevention and control). Establishing notifiable disease reporting requirements safeguards population level health by guaranteeing the identification and follow-up of cases, as well as monitoring trends in disease to inform outbreak response. At the national level, monitoring trends in notifiable diseases enables public health authorities to identify changes in disease distribution (e.g., via age, race, sex, and geographic comparisons over time), monitor incidence and detect potential intervention points. There are only four sexually transmitted diseases that are reportable in all states and notifiable to the CDC: syphilis, gonorrhea, chancroid and chlamydia. Both healthcare providers and laboratories are required to report positive results of tests for STIs to local health departments. The local health department is then responsible for reporting to the state, which in turn gets reported to the CDC. The identification and response to changes in disease occurrence is one of the goals of the surveillance system.

In the CDC's 2012 STD Surveillance report, over 1.4 million cases of chlamydia were reported which is the largest number of cases of any condition ever to be reported to the CDC (1). Chlamydia is the most prevalent bacterial STI and since 1994 has been the most frequently reported STI. The 2012 rate per 100,000 people was 456.7, which represents a stable increase of 0.7% since 2011(1). Although at least some of the increase is due to targeted screening programs as well as increased ability to detect through improved assays (4), increases may also be due to real increases in disease incidence (4). The estimated direct and indirect cost of the disease exceeds 2 billion U.S. dollars annually, and includes costs associated with treatment of uncomplicated and advanced disease (5).

Trends in chlamydia are hard to assess with confidence for several reasons: 1) screening recommendations and practices have changed greatly over the past decade, leading to expanded testing; 2) diagnostic tests have increased in sensitivity (increased use of nucleic acid amplification tests); 3) prior to 2000 not all states had regulations in place requiring the reporting of chlamydia; 4) evolving screening recommendations have resulted in differential screening of women compared to men; however, with the availability of non-invasive methods (e.g., urine based screening) of testing this ratio is decreasing (5).

The increase in reported chlamydia infections reflects expanded screening/ testing, increased use of NAATs with greater sensitivity, as well as an increased interest in mandatory reporting from providers as well as improvements in the electronic systems used to capture data and reports.

Given the asymptomatic nature of the disease, some of the best estimates of prevalence come from population based surveys. The National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey that covers the United States population who are 14–39 years old (6). Survey results provide an important measure of representative chlamydia disease burden. In the three year time span from 2005 to 2008, the estimated prevalence of chlamydia for persons 14–39 years old was 1.5% with a 95% confidence interval of 1.2%, 1.9% (1).

To better monitor trends in disease burden, prevalence of chlamydia among persons screened in defined populations may be used, as it is not influenced by screening programs. In 2011, the median US state-specific chlamydia positivity rate among young women (15-24 years old) tested at family planning clinics was 8.3% (1). This is an almost 1% increase in a 2 year time span, as the 2009 rate was 7.5% for the same population (7).

Chlamydia is highly transmissible and has an incubation period of 7 to 21 days with a significant asymptomatic reservoir. Approximately 10% of men and from 5 to 30% of women with urogenital chlamydia infection will experience symptoms, such as genital discharge or painful urination (8, 9). In addition, symptoms may not appear until several weeks after the initial exposure (9).There is a high prevalence of co-infection in sexual partners, exceeding 50%. Re- infection is quite common due to the asymptomatic nature of the disease (10).

In 2012, the chlamydia rate in women did not increase for the first time since nationwide reporting began, though the rate in men increased 3.2% (1). From 2008-2012 the rate of infection in men increased by 25% compared to an 11% increase in women in the same time period (1). Despite accelerating rates in men, the overall rate among women remains over two times the rate in men (Figure 1). Men are often treated for chlamydia infection on the basis of presentation of a syndrome of urethritis or as a contact to a woman with chlamydia, and a definitive etiologic diagnosis is not made; these cases would not meet the surveillance definition for chlamydia and would not be reported (10).

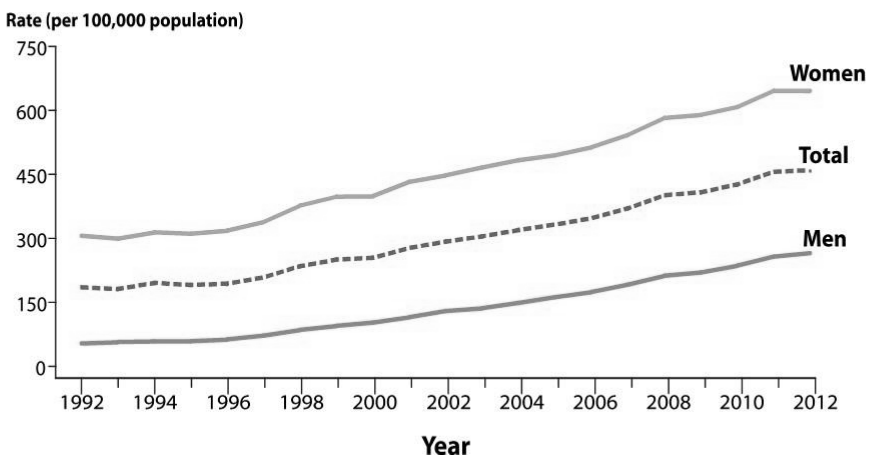
Chlamydia is most common among young people, with the prevalence among 14-24 year olds nearly three times the prevalence among 25-39 year olds (1). The prevalence of chlamydia is estimated to be around 7% for the general population in the United States with around one in fifteen young women infected (ages 14 to 19) (6, 11). Young people bear a disproportionate burden of the disease due to many synergistic factors, including: inconsistent condom use, rapid succession of sexual partners (quicker than infectivity period, increasing risk of transmission), possible cervical ectopy in young women, stigma, and barriers to education, prevention and treatment services (e.g., transportation and costs) (12-14).

The rates of chlamydia vary based on demographics, specifically among different racial/ethnic groups; the rate of chlamydia in Non-Hispanic Black women was over six times the rate of chlamydia in White women (1613.6 and 260.5 per 100,000 respectively) (1). This disparity is heightened among men, with the rate of chlamydia among Black men over eight times the rate in White men (809.2 and 95.9 cases per 100,000 males, respectively) (1). The difference between Hispanic and Whites is less pronounced, with chlamydia rates among Hispanics just over two times the rate in Whites. The chlamydia rate among Asians was lower than that of Whites. Whites have chlamydia rates at 1.6 times the rate found among Asians (1).

Chlamydia is also common among men who have sex with men, with rectal chlamydial infection ranging from 3.0-10.5% and pharyngeal infection ranging from 0.5- 2.3% (15). More detailed rates and disparities among MSM are presented in a separate section below (See section 2.1).

Chlamydia rates have increased in all regions across the Unites States from 2003 to 2012. In 2012, chlamydia was highest in the South and lowest in the Northeast, but did not vary much between the Midwest and the West (1). Chlamydia rates by state ranged from the lowest at 233.0 cases per 100,000 population in New Hampshire to the highest 774.0 cases in Mississippi (1). The 2012 rate of Chlamydia in the most populous metropolitan statistical areas (MSAs) was similar to the rate in 2011 (481.1 and 480.9 cases, respectively) (1). Counties within the United States that reported the highest chlamydia rates were located primarily in the Southeast. Seventy counties and independent cities accounted for 44% of all chlamydia cases in 2012 (1).

Figure 1: Chlamydia, rates for women, men and total 1992-2012



In the United States gonorrhea is the second most commonly reported notifiable infectious disease with around 800,000 estimated cases annually and 334,826 cases reported in 2012 (1).The estimated number of cases reflects that less than half of all infections are detected and reported. As with chlamydia, supplemental data from gonorrhea screening in multiple settings provides a comprehensive assessment of disease burden.

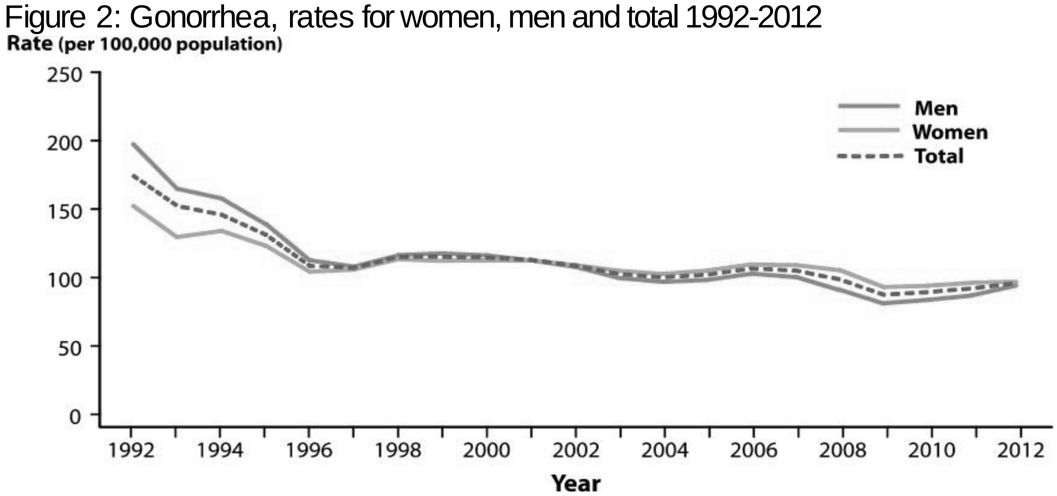
Following the National Gonorrhea control program in the mid-1970s, there was a 74% decline from 1975-1997 in the rate of reported gonorrhea, following this drastic decline was a period of plateau for a decade (Figure 2). Rates have increased slightly each year beginning in 2010, to the present increase from 2011 to 2012 by 4% (1).The rate in women (108.7 per 100,000 population) is similar to that in men (105.8 per 100,000 population) (1). Similar to chlamydia, 2012 surveillance data demonstrate a disparate burden of gonorrhea in adolescents and young adults, as well as some geographic areas and racial/ethnic groups.

The rate of gonorrhea decreased 7.5% for those 15-19 years, yet increased 3.1% for those 20-24 years old. In 2012, women 20–24 years old had the highest rate of gonorrhea (578.5 cases per 100,000 females) in comparison to all other age and/or sex groups (1). From 2011– 2012, the gonorrhea rate for women age 20-24 increased by 1.6% (1). Young men 20–24 years old had the highest rate of gonorrhea (462.8 cases per 100,000 males) compared to males of other age groups (1). During 2011– 2012, the gonorrhea rate among men, 20-24, increased 5.5%.

In the United States, in 2012, the South had the highest reported gonorrhea rate (131.9 cases per 100,000 population), followed by the Midwest (114.6), Northeast (92.6), and West (73.3) (1). The 2012 rates reflect an increase in three of the four regions of the United States: 19% in the West, 8% in the Northeast, and 3% in the Midwest (1). The South area of the US experienced a decrease of 1.4% (1). Similar to chlamydia, gonorrhea rates are highest among Blacks (462.0 per 100,000 population), 14.9 times that among Whites (31.0 per 100,000 population) (1). The gonorrhea rate among Hispanics was almost twice the rate among Whites. This disparity was greater for Hispanic men (2.2 times) than for Hispanic women (1.8 times) (1).

Gonorrhea incidence among MSM declined remarkably in response to the HIV epidemic during the 1980s but started increasing in discrete local areas in the United States in the early 1990s. The increase in STI acquisition appears to result in part from HIV “serosorting” with sex without condoms among HIV seroconcordant men (those with the same HIV-status) including the increased frequency of HIV seroconcordant sex partner recruitment for "bareback" sex (i.e., condomless sex) on dating websites and mobile applications (15). Serosorting is an effective method for HIV prevention, but it increases the transmission of STIs (15).

Figure 2: Gonorrhea, rates for women, men and total 1992-2012



The history of primary and secondary syphilis in the United States started in the 1940s with a large portion of the population impacted by the disease. The availability of penicillin caused dramatic declines in rates of syphilis (Figure 3). Rates continued to decline throughout the 1990s, reaching a low in 2000. In the late 1990s it appeared that elimination as a goal would be feasible, which led to the creation of the “National plan to Eliminate Syphilis from the United States” (16). The incidence of P&S syphilis did decline further among women and Blacks, greatly decreasing the disparity between Blacks and other racial/ethnic groups in syphilis incidence. However, since 2000 the rate of P&S syphilis has been increasing.

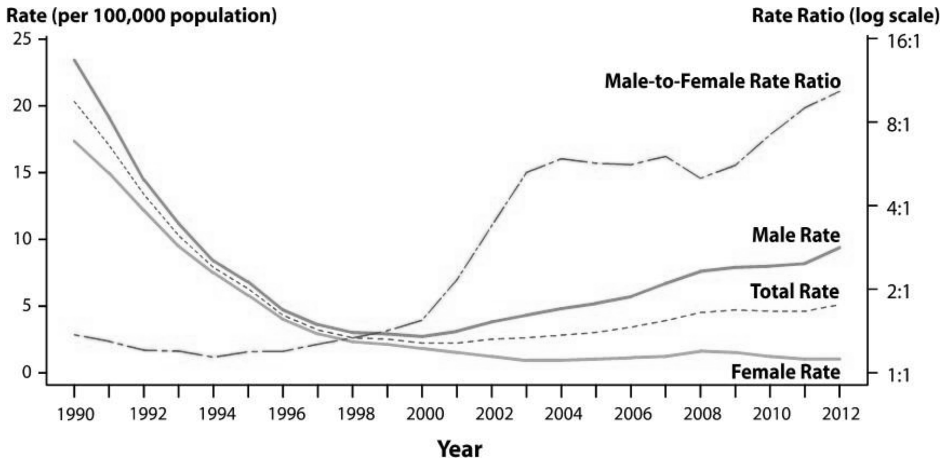
Rates remain high in some urban areas, rural areas of the south, in minority populations as well as a more recent resurgence in men who have sex with men. While the national rate remains low at 5.0 cases per 100,000 population and a total of 15,667 cases reported in 2012, in 2012 the rate among men (9.3 per 100,000) increased almost 15% from 2011 (8.1 per 100,000 population); this rate increase was specifically among men who have sex with men (1).

Trends in syphilis vary greatly between men and women, with overall increases observed predominantly among men. Outbreaks of syphilis among MSMs have been reported since 2000 characterized by high rates of HIV coinfection and high-risk sexual behaviors. Between 2001 to 2005, the rate of P&S syphilis increased 70% among men from 3.0 cases to 5.1 cases per 100,000; during the same time the overall rate declined 36% among women (1). Information on the risk behaviors of persons infected with syphilis have not been collected nationally and thus the numbers of persons with reported syphilis who are MSM is not known. The male-to- female case ratio has been used as a proxy for transmission among MSM. In 2005, an official request from the CDC went to all state health departments to begin collecting the birth sex of all sex partners for persons diagnosed with syphilis. This data remains incomplete, with the birth sex of sex partners accounted for in 82% of reported male syphilis cases in 2012 (1). Among cases of primary and secondary syphilis for which the birth sex of the sex partner was known, 75% was accounted for by male-to-male sexual behavior (MSM) (1).

After persistent declines for nearly two decades, the rate of syphilis among women increased slightly starting in 2004, however there has been a steady decline of cases from 2011 to 2012 (1). In 2012, the rate of primary and secondary syphilis was highest those aged 20–24 years which represented 14.8 cases per 100,000 population. Rates were highest among men 20–29 years, which represented an increase of 11% for those between the ages of 20-24 and an almost 16% increase for the 25-29 age group. Rates also increased, in 2012, among women aged 20–24 years and 45–54 years (1); however rates remained level or experienced decreases for women of all other age groups. Rates of P&S syphilis remained the highest among women aged 20–24 years compared to all other age groups. Rates of syphilis (primary and secondary) are highest among Blacks at a rate of over 6 times that of Whites (16.4 vs. 2.7 cases per 100,000 population) (1). The rate among Hispanics (5.7) was 2.1 times that of Whites (1). During the four year period from 2008 to 2012, the syphilis rates increased 41% among Hispanics and 21% among non-Hispanic Whites (1). The rate decreased slightly among non- Hispanic Blacks (less than 1%). The majority of all reported cases are comprised by three racial/ethnic groups: Non- Hispanic Blacks, non- Hispanic Whites, and Hispanics who made up about 95% of reported syphilis cases since 2008 (1). The South accounted for 44% of syphilis cases in the United States in 2012. During 2011–2012, the area that saw the largest increase in syphilis cases was the West with an 18% increase, followed by the Northeast (15.8%) and the South (11.3%) (1). Fifteen states accounted for 70% of all reported cases of syphilis in the United States.

The rate of P&S syphilis in 75% of the 15 states exceeded the national rate (5 cases per 100,000 population); 60% of the states were in the South region of the United States (1).

Figure 3: Primary & Secondary Syphilis, 1990-2012



2. REVIEW OF RELATED LITERATURE

2.1 Subpopulations at increased risk for sexually transmitted infections

As illustrated within the disease specific sections above, there are higher rates of reported STIs among some racial/ethnic minority groups, in specific regions of the United States, among youth, and among men who have sex with men. The spread of STIs is impacted by social, economic and behavioral factors. Sexually transmitted infections disproportionately affect disenfranchised people and their sexual networks thereby increasing the overall opportunity to acquire an infection. The following section provides a brief overview of the impact of these factors on STIs.

Sexually transmitted infections are stigmatized as they are acquired sexually. There is often shame associated with an STI diagnosis, due to community norms and perceived ability to prevent sexually transmitted infections. Stigma and shame contribute to the decreased likelihood of disclosure of disease status among sexual partners; in addition, it also may have an impact on voluntary testing.

Access to health care is essential for early detection and treatment of STIs. In addition, ongoing access to health care can provide opportunities for behavior change counseling with clinicians to reduce risk of acquiring STIs, as well as routine testing. There is an overlap between those who experience the highest rates of STIs and those who have limited access to health care.

Drug use is associated with STI incidence, particularly syphilis. Methamphetamine use can adversely affect mucous membranes, making them more susceptible to trauma and bleeding and thereby increasing the risk for STIs (17).

In addition to the biologic implications of drug use, the associated socio-behavioral factors such as sexual networks with high prevalence of STI and HIV, as well as involvement in sex work (exchanging sex for money, drugs, or other items) increases the overall risk for STIs (17).

High rates of incarceration impacts sexual networks and overall risk of acquiring STIs. Men are more than 10 times more likely to go to prison than are women, impacting the number of available men to partner with who are not incarcerated. When there is a high demand for male sexual partners there is little incentive for monogamy. Multiple and concurrent partnerships facilitate the rapid spread of STIs through networks (18).

Race and ethnicity are highly correlated with other social determinants of health such as economic factors (poverty, education). Social determinants of health can impact individuals' ability to prioritize protecting their sexual health. Those who are navigating stressful life circumstances, such as unstable housing and lack of employment have an increased risk for acquiring and transmitting STIs. Managing daily living may take priority over sexual health, such as annual STD screening, purchasing and using condoms, or mutual monogamy.

Using data from a nationally representative probability sample (The National Health and Social Life Survey) of 1,511 men and 1,921 women, Laumann et al. explored individual risk factors associated with STDs (19). Even after controlling for all individual level risk factors, Blacks were found to be five times more likely to be infected by bacterial STD than any other racial group (19).

There is an increased risk for STI acquisition in communities where STI prevalence is higher. With each sexual encounter, there is a greater chance of contact with an infected person than they would in a lower prevalence setting.

Sexual mixing patterns are segregated by race and socioeconomic status; additionally racial and ethnic minorities are less likely to mix sexually with racial groups different from their own (15). Thus STIs concentrate within racial minority populations because of assortative partner selection; for this reason alone the likelihood of Blacks having a STI is 1.3 times greater than it is for Whites (19).

Compared to women and to men who have sex with women only, men who have sex with men are at increased risk for STI acquisition (15). Though a number of individual- level sexual risk behaviors for example, higher numbers of lifetime sex partners, significantly contribute to the disparities in the overall health of men who have sex with men, other structural and environmental factors have been found to be predictive of higher rates of STIs, including HIV. In addition, MSM who experience poverty and other social determinants of health and/or if they belong to a racial/ethnic minority group are even more vulnerable to poor health outcomes.

Syphilis is the only reportable STI that includes information on the sex of partners of those infected, therefore, trends in STIs among men who have sex with men in the United States are based on findings from sentinel surveillance systems (reviewed in section 2.3).

Sexually transmitted infection testing strategies may not be the most effective for detecting infection among MSM. Anatomical testing sites for gonorrhea and chlamydia largely focus on the urethra, as opposed to the pharynx or rectum. Urethral infections are more likely to be symptomatic than pharyngeal or rectal infections and may not represent the anatomic sites MSM are likely to be infected (15).

Syphilis is more likely to be diagnosed in the secondary stage of disease in MSM, in comparison to diagnosis among men who have sex with only women, due to the primary stage chancre manifesting in the anus or rectum. As the chancre often develops in the anus or rectum, any associated rectal pain or bleeding may be confused with hemorrhoids or if asymptomatic the self-limiting chancre may go unnoticed (17).

An analysis of the MSM Prevalence Monitoring Project from 2003-04 among MSM attending the Fenway Health Clinic in Boston, found that 65% of MSM receiving an STD test were asymptomatic, 7% of asymptomatic MSM tested positive for at least one STD (20). STD prevalence among MSM who were not experiencing symptoms was: 1% urethral gonorrhea; 1.7% pharyngeal gonorrhea; 5.6% rectal gonorrhea; 2.2% urethral chlamydia and 4.3% syphilis (20). The high rates of rectal gonorrhea and syphilis in MSM who are not experiencing symptoms underscores the importance of routine screening in this population (including rectal STI testing).

2.2 **Morbidity, Cure, and Prevention**

Chlamydia, gonorrhea and syphilis are easily treated and cured if diagnosed early. However, often diagnosis is delayed due to the asymptomatic nature of disease. Symptoms need not be present for the disease to progress to serious health consequences, which are often worse for women than for men. In addition to the specific sequelae of disease, being infected with an STI increases a person’s risk for HIV infection. In men the local infection of chlamydia and gonorrhea can manifest as conjunctivitis, urethritis or proctitis resulting in the complications of epididymitis or in rare circumstances Reiter’s syndrome. Evidence suggests that chlamydia causes 35-50% of NGU in heterosexual men (15). About 70% of acute epididymitis in young men appears to be attributable to chlamydial infections (15). Chlamydia is also responsible for up to 15% of proctitis cases in men (15). The sequelae are infertility and chronic arthritis, both of which are rare and happen in less than 10% of cases (15).

The majority of urethritis is asymptomatic, however if symptoms are present they are urethral discharge and dysuria (15). The incubation period is unknown but is likely 5-10 days in symptomatic infections (15). The majority of men with gonorrhea develop symptoms within 2-5 days after gonorrhea infection; the predominant symptoms are urethral discharge or dysuria (15). Approximately a quarter of those infected will develop a scant or minimally purulent exudate (thick pus). About 5-10% of patients report not having any symptoms (15). Symptomatic anorectal gonococcal disease presents with approximately 50% symptoms of rectal pain, discharge, constipation and tenesmus (the constant feeling to pass stools) (15). If left untreated, gonorrhea can turn into a disseminated infection (spread to the blood), which can be life threatening- but is very rare only occurring in less than 3% of all cases (15).

In women uncomplicated gonorrhea and chlamydia infection have similar manifestations as with men (urethritis, dysuria) and also cervicitis and potential ascension of the reproductive tract (15). The sequelae primarily impacts reproductive functioning resulting in infertility, ectopic pregnancy, chronic pelvic pain and more rarely chronic arthritis (15). At least 15% of all American women who are infertile are so due to untreated STIs (15). Evidence suggests that up to 40% of untreated chlamydia cases will develop into pelvic inflammatory disease (PID) and that about 20% of women with PID will eventually experience infertility due to the disease (21). Most infections in women are asymptomatic, however when symptoms are present they include discharge, dysuria, and pyuria (15). There is an increased risk of upper reproductive tract damage in women with STI re-infection. Scholes et al., in 1996, found that screening can reduce more than 50% of the incidence of PID; in a randomized control trial to determine if selective screening of chlamydia could prevent PID there were 9 cases of PID among the intervention group and 33 cases among those receiving standard of care (relative risk (RR)=0.44; 95% CI: 0.20-0.90)

(21). Screening and treatment decreases the prevalence of infection in the populations and reduces the transmission of disease.

Vertical transmission via perinatal transmission results in neonatal chlamydial or gonococcal conjunctivitis in 30-50% of exposed babies (15).The local infection in neonates is conjunctivitis, pneumonitis, pharyngitis, and rhinitis; complications may be chronic lung disease and sequelae are rare if there are any at all (15). The United States Preventive Services Task Force (USPSTF) recommends gonorrhea and chlamydia screening in pregnancy for all women under age 25 to reduce risk of vertical transmission (22).

Treating infected patients prevents the onward transmission of sexually transmitted disease. The CDC recommends all sex partners should also be treated to avoid re-infection of the index patient. In addition, pregnant women should be treated for both chlamydia and gonorrhea to prevent transmission to infants during birth. Delays in receiving treatment have been associated with complications in women such as PID. Co-infection frequently occurs among patients who have gonorrhea, therefore empirical treatment for chlamydia is recommended for all. The recommended treatment is a single-dose of directly observed Azithromycin (15). Patients should be counseled to refrain from sexual activity for 7 days after the single-dose treatment regimen, or until after the completion of the 7 day regimen (15). Sex partners should be treated as well (15).

The proportion of infections caused by resistant organisms is increasing each year which causes difficulty in treating cases of gonorrhea. The Gonococcal Isolate Surveillance Project (GISP) was created in 1985 with the intent to monitor and document trends in antimicrobial susceptibilities in *N. gonorrhoeae* strains (see section 2.3. for further discussion on this sentinel surveillance program) (23).

Syphilis is divided into three clinical stages, primary, secondary and tertiary (or late) (24). Latent syphilis is a serological diagnosis where symptoms are not apparent and which is differentiated into early (less than 1 year) or late (greater than 1 year) (24). Surveillance focuses on cases of primary and secondary syphilis because it is the best indicator of incident and transmissible disease; however, there may also be a significant burden of latent or asymptomatic syphilis cases that are not captured. Transmission of syphilis is relatively inefficient, at about only 20% between an infected and uninfected sexual partner (3). The incubation period ranges between 10 and 30 days. Typically after 3 weeks a chancre develops at the site of contact (10). The chancre is painless and will heal without treatment within 1 to 6 weeks (10). About 25% of people will present with multiple lesions/chancres (24, 25). Tests for syphilis may not be positive during the primary stage Secondary syphilis occurs 3 to 6 weeks after the primary chancre appears; this systemic stage may persist for months (15). Additionally, primary and secondary stages may overlap. In over 75% of cases of secondary stage syphilis a rash develops, somewhat less common (about 50% of cases) is lymphadenopathy (15). Latent syphilis is characterized by persons with historical or serological evidence for syphilis but who have not taken treatment and who also do not have any clinical evidence of disease (15).The diagnosis of latency formally requires examination of the cerebrospinal fluid to rule out asymptomatic neurosyphilis.

Syphilis is a systemic disease; 10 to 15% of patients with primary syphilis will have cerebrospinal fluid abnormalities (25). Late complications of syphilis such as neurosyphilis and cardiovascular syphilis generally do not develop until 10-20 years after infection. Tertiary syphilis causes the principal morbidity and mortality in adults. Studies in the pre penicillin era suggest that about a third of untreated infections were followed by tertiary complications, with neurosyphilis being most common.

Penicillin, administered parenterally (via injection), is the preferred drug treatment for all stages of syphilis. The dosage and length of penicillin treatment depend on the stage and presentation of the disease. Untreated syphilis can lead to neurosyphilis and congenital defects. Untreated STIs also are a community health risk as they pose ongoing potential for infection. Untreated early-stage syphilis in pregnancy can result in death in 30-40% of cases and, if syphilis is acquired anytime in the 4 years pre-conception, it leads to infection of the fetus in an estimated 80% of all cases (26). Sexually transmitted infections are referred to as hidden epidemics due to the reluctance of many Americans to address sexual health issues in an open way (5). STIs are stigmatized due to their transmission by sexual behavior. The strongest predictors of sexual health are social, economic and environmental (as discussed in the section 2.1). Sexuality and sexual health needs to be addressed at the individual and community level to be effective. Prevention efforts could also impact prevalence of the disease, as prevalence is a function of both incidence and duration of disease. If incidence was decreased via primary prevention, and duration of infectiousness decreased via increased screening (detecting and treating) overall prevalence would decline. Controlling STIs is a core aspect of the World Health Organization's (WHO's) Global Strategy on Reproductive Health (27, 28).

Prevention can minimize the negative long-term health consequences of STIs while simultaneously reducing overall associated costs. Prevention can be framed using the Anderson-May equation for the reproductive rate (Ro=bcD) where beta is the probability of transmission, c is the number of partners and D is the duration of infectiousness; Ro is equal to the average number of secondary cases generated from a new infection if R0>1 then disease will continue to spread (15). To impact the reproductive rate we must impact one or all of the factors in the equation.

Primary prevention is designed to prevent the acquisition of the disease with strategies such as condom use, microbicides and circumcision. Secondary prevention can lessen the severity or extent of the disease. Screening programs contribute to early diagnosis and treatment which contributes to primary prevention by interrupting the transmission. Examples of secondary prevention include improved medical care via expanded screening, single dose therapies and rapid access to treatment and medical care, including partner therapy.

Primary prevention (which impacts the transmission efficiency in the reproductive rate equation) has a direct impact on individual risk, such as using barrier methods during sexual activity (condoms, diaphragms), microbicides, vaccines and male circumcision. Studies show condoms can reduce the risk of STIs by 20-60%; by limiting genital infections, condoms may also decrease the overall PID risk for women (27, 28). Laboratory studies also indicate that the female condom (a condom that is worn inside the body) is an effective barrier to viruses, including HIV, and to semen. Other barrier methods include use of the diaphragm. In observational studies, diaphragm use has been demonstrated to protect against STIs, with a significant reduction in the odds of disease acquisition (27).

Microbicides can be formulated in various different delivery methods including gels, creams, lotions, sprays, tablets or films as well as in vaginal rings. These agents are being developed and tested for both for vaginal and rectal application (27). Despite more than a decade of research, no widely available microbicide for the prevention of STIs has been distributed. However, many have been in clinical trial development. For example HPTN 035 found that microbicides were acceptable due to their ease of use and increase in sexual pleasure; decreasing acceptance was the burden of remembering to use the product and reported messiness during sex (29). Adherence was dependent on contextual and partner-related factors. Most research on microbicides has focused on preventing HIV with STI prevention as a sub-aim. Currently there are no vaccines in development or available to prevent the notifiable STIs, though modeling studies have shown that vaccines will be cost effective for curable STIs (30, 31). To date, there are vaccines for only two STIs, which are viral: hepatitis B and HPV. In 1982, the hepatitis B vaccine was introduced and in 2006 the HPV vaccine was introduced (30, 31).

Male circumcision has been shown to reduce the risk for HIV acquisition among men by an estimated 50-60%, in three randomized controlled trials; results also showed the protective effect of circumcision against certain STIs, in particular HPV and Herpes Simplex Virus Type 2 (HSV-2) (32, 33). However, in a randomized control trial, adult male circumcision was not shown to be protective against acquiring syphilis, gonorrhea or chlamydia (34).

Primary prevention strategies can also include behavioral risk reduction strategies, including strategies to reduce the overall number of sexual partners such as abstinence and mutual monogamy, reducing high risk partners, delaying the age of sexual debut, and focusing on partner selection.

All of these prevention strategies rely on the knowledge, behavior and communication skills of sexually active persons. Comprehensive sexual health education can increase these critical skills in young adults. In addition, individual and group level behavioral interventions have been shown to be efficacious in reducing acquisition of STIs. A recent meta-analysis of 29 single-session individual level STI interventions found that the overall odds of participants being infected with an STI in the intervention group was reduced by 35% relative to the control group participants (35). Implementation of STI control programs requires not only providing availability and access to these interventions, but also ensuring effective scale-up and sustainability for maximal population impact.

Secondary prevention can reduce the duration of infectiousness and decrease the overall reproductive rate, which is practical for diseases that are curable. Disease screening and treatment programs (including retesting to detect repeat infections), partner treatment, presumptive treatment, increasing access to health care and treatment guidelines are all secondary prevention strategies.

Partner notification is the process by which public health authorities help to arrange for sex partner evaluation and treatment. Data are limited regarding whether or not partner notification impacts exposure to STIs and whether or not it reduces the burden of disease within a community. Nevertheless, published evaluations of partner notification interventions highlight the important contribution this approach can make in surveillance, particularly in regard to case finding (36). In a 2002 survey of jurisdictions with high morbidity of either syphilis, HIV, chlamydia and/or gonorrhea, authors found that 89% of persons with primary or secondary syphilis were interviewed for case-finding purposes, as were 52% of those with HIV, 17% of those with gonorrhea and 12% of those with chlamydia (37). This study focused on partner notification via Disease Intervention Specialist (DIS) workers at local health departments. Other methods of partner notification include use of the internet for notification, social network based notification and expedited partner therapy. These methods are complementary approaches to case finding and use of one approach need not preclude the use of another; flexibility will increase the probability that sexual partners will be treated. However, these approaches do not facilitate the reporting of disease to the health department. Partner treatment reduces the index patient's risk for re-infection. Therefore, providers should encourage partner notification and treatment in any way that is acceptable to the index patient.

Since many infections are asymptomatic, effective control must involve periodic testing of individuals at risk. The cost of universal screening may be prohibitive, thus various approaches to defining a target population have been evaluated (35). There is a decreasing prevalence in some areas of the United States due to control programs that include clinic and community based screening. Multiple cost-effectiveness studies have shown that universal screening is preferred in settings where prevalence is above 3-7% (4). School based routine screening for STIs were shown to be cost-effective and to detect high rates of asymptomatic disease (>90%) (4, 38, 39). Overall the screening rate of women ages 15 to 24 for chlamydia is 44%; increasing this rate will increase the capacity to treat infected individuals and their partners as well as further monitor trends in disease (40). A mathematical model study by White et al., demonstrated the importance of timely delivery of effective health care in preventing the transmission of gonococcal infection (41). If gonorrhea screening and treatment covers 20% of the population, there would be an overall decline of 30% in the population (42).

In conclusion, multiple strategies are necessary to curb the incidence and prevalence of STIs. It is important to understand the burden of STIs to implement STI prevention appropriately and effectively. Accurate and timely surveillance will increase the ability to identify trends, outbreaks and clusters of specific diseases and affected populations. Rapid response to outbreaks is important for disease control, prevention of transmission, lessening the severity of disease, and potentially impacting the spread of HIV.

2.3 **Surveillance of reportable sexually transmitted infections**

Prior to proposing methods to strengthen the current surveillance system, it is important to review the history and practices of the existing system. Public health surveillance is defined as “the ongoing systematic collection, analysis, interpretation, and dissemination of health data for planning, implementation and evaluation of public health action” (43, 44). The purpose of surveillance includes: 1) assessing the health status of a population; 2) establishing public health priorities; and 3) targeting disease prevention and control activities to reduce the burden of disease.

Surveillance dates back to the first recorded epidemic in Egypt in 3180 B.C. (45).Surveillance is one of the essential public health functions to avert epidemics and control disease. Historically, three types of data/information were recorded as part of surveillance: 1) health outcomes; 2) risk factors; 3) interventions (45). This type of information remains important. Systematic reporting of various diseases started in the United States in 1874, followed by reporting in Italy in 1888 and United Kingdom two years later (1890) (45). The turn of the century brought the wide spread expansion of surveillance systems to monitor public health.

In 1942, the current-day CDC was founded as "the Office of National Defense Malaria Control Activities;" in 1946 the agency changed its name to the Communicable Disease Center (45). In 1947, the CDC acquired their first epidemiology unit by taking over a plague laboratory in San Francisco (45). As a response to repeated infections, in 1955, the Polio Surveillance Program was established in the US by the CDC (45). The national notifiable disease surveillance system in the United States monitors health conditions via reporting from states, territorial and local health departments.

From 1951 to 1985 notifiable disease data was sent to the CDC by mail and then via phone (44, 46). In 1985 the electronic surveillance project was established to collect notifiable disease data electronically; from this project the CDC developed standard protocols for states to submit their data (44, 46). In 2000, US states began to receive federal funding from the CDC to develop a plan to implement integrated electronic systems for disease surveillance (47). By 2011 every state was able to implement a disease reporting system (47).

Surveillance relies on analyzing and interpreting data from many sources across varied systems, which requires specialized skills in statistical reasoning, subject matter expertise and ability to communicate effectively across sectors. The majority of surveillance data is collected by two distinct approaches which require different types of analyses and interpretations. The first approach is representative surveys which use validated statistical methods designed to make population level inferences of public health importance. The second approach is data that comes from the public health surveillance systems that is not collected using probability sampling (e.g., case reports, syndromic surveillance).

Surveillance of STIs relies on many sources of data including: the notifiable disease surveillance system, sentinel surveillance projects, prevalence monitoring, population based surveys, and administrative/claims databases. Collaborative efforts across these systems can maximize the utility of data collected. Information from disparate sources may be able to identify patterns that individual data cannot. In addition, use of data sets collected for alternate purposes can be more cost effective than collecting new data. A brief overview of different surveillance systems and sources of STI data is presented below.

In 2005, the CDC established a sentinel surveillance system called "the STD Surveillance Network" (SSuN). SSuN is comprised of a system of local enhanced STI surveillance systems based in sites across the nation that follow common protocols (48) (example available from www.stdpreventiononline.org/assets). The purpose of SSuN is to improve the capacity to detect and monitor trends in STIs through data collection and analysis in diverse populations at risk for STIs (48). SSuN currently includes 12 collaborating local and state health departments within the network. In 2012, the network consisted of a total of 42 STD clinics within 12 states. These sites collected behavioral and demographic data from patients who presented for STI or HIV (48).

The Gonococcal Isolate Surveillance Project (GISP) is another example of a national sentinel surveillance system. Its purpose is to monitor trends in antimicrobial resistance as and susceptibilities of N. gonorrhoeae strains across the United States (See http://www.cdc.gov/STd/gisp/default.htm) (49). The GISP has demonstrated that gonococcal isolate samples from MSM are more likely to have antimicrobial resistance in comparison to isolates from other populations, like men who have sex with women (MSW) (49). The results of the GISP form the basis for gonorrhea treatment recommendations nationwide.

From 1988 to 2013, the Infertility Prevention Project (IPP) was federally funded and required states to submit data on chlamydia and gonorrhea testing to the CDC; these data were known as “prevalence monitoring data” and included basic clinical indicators as well as demographic data (See http://www.cdc.gov/std/infertility/ipa.htm). The data was to be used to guide programmatic decision-making around screening and control within organizations.

The IPP funded chlamydia and gonorrhea screening programs, including treatment services, for low- income women's healthcare clinics. Results from the IPP demonstrated that routine screening can reduce chlamydia prevalence as well as reduce the incidence of pelvic inflammatory disease (PID) (50). Sexually transmitted infection positivity data are useful locally to inform population level interventions, such as clinic-based screening recommendations. The data can also be used to identify at-risk populations (e.g., low income women) in need of prevention interventions.

Nationally representative surveys are not primary sources of surveillance data and do not have an STI focus, however data can be used for estimating the population level burden of disease. In addition, researchers and public health workers can often advocate for the inclusion of specific questions or measures in these surveys. For example, the Behavioral Risk Factor Surveillance System (BRFSS), a national probability sample of adults over 18, includes a core list of questions that every participant will receive (see http://www.cdc.gov/brfss/ for overview); there are also state- by-state questionnaires that include different questions (51). A search for sexual risk behavior or STI-related questions in the BRFSS reveals that there has not been a module on this topic in the core survey, and there is not continuity over time or region for these domains; thus it will be hard to determine trends in the sample.

Other examples of representative surveys that may be useful for STI surveillance are the National Health and Nutrition Examination Survey (NHANES) and the National Hospital Discharge Survey (NHDS). The National Health and Nutrition Examination Survey is a US-based nationally representative sample; the survey has periodically examined the rate of antibodies to HPV-16, herpes simplex virus 1 and 2 as well as chlamydia and gonorrhea. Using 2008 estimates, the prevalence of chlamydia among 18-24 year old women and men was 3.2% and 1.6%, respectively, and 0.62% and 0.32% for gonorrhea (3). The NHDS is a longitudinal survey of a probability sample of records of patients discharged from acute care hospitals in the United States. To date it has not included systematic assessment of STIs, STI symptoms, or sexual behavioral risks. Analysis of this data for diagnoses or treatment codes is possible.

Administrative datasets are not collected for the purpose of surveillance, but can be used to understand disease burden, monitor trends and ultimately help guide public health action. For example, Medicare, Medicaid and Social Security Disability data have been linked to population survey data to monitor changes in health screening and outcomes over time. With the expansion of the Affordable Care Act, Medicaid coverage is being extended to many more people including those with HIV, AIDS, hepatitis, TB and STIs. With the increase in electronic medical record use, claims databases may be able to be linked to EMRs; this may be increasingly useful for planning and evaluating public health activities in the future.

The objective of syndromic surveillance is to identify clusters of illness rapidly, before diagnoses are confirmed and reported. The purpose is to mobilize a timely response in effort to reduce the morbidity and mortality of the disease. Vaginal discharge and urethritis are the most common syndromes of STIs (as discussed in section 2.2). The World Health Organization (WHO) recommends a global strategy to respond to the STI epidemic, which includes a syndromic approach for the detection and management of abnormal vaginal discharge (52). Syndromic surveillance is not effective in monitoring disease trends for STIs, due to the high prevalence of asymptomatic disease.

Active surveillance systems require outreach on a regular basis to increase the response rate of the reporting of specific diseases, particularly prevalent diseases like chlamydia and gonorrhea; it provides the most accurate and timely information, but it is also expensive and resource intensive. The current national STI surveillance system relies on passive surveillance systems, consisting of routine notifiable disease reporting which is often limited by missing data or incompleteness in reporting. Passive surveillance suffers from inconsistent data quality and lack timeliness.

A form of passive surveillance, case reporting, is the backbone of surveillance. The advantages to case reporting are that is it legally required, the system is well established and the data quality is high for rare diseases. However, there are several disadvantages of case reporting, specifically in terms of infectious diseases. The surveillance system can be quite slow, with a range of 60-90 days for local reporting trends to multiple years for national trends.

Case reported data also depends upon the accuracy of numerous health care workers. Cases are reported from many different settings. From the 2012 STD Surveillance report, cases were reported from: private practices (30% of cases), STD clinics (8-20%), family planning clinics (10- 15%), other health clinics (about 10%), and emergency rooms/hospitals (about 5%) (1). This represents many different constituents across settings; ease of reporting and limited knowledge of when\how to report may be differential across health care setting. For example, STD specialty clinics may have different and specialized training in reporting STIs in contrast to ER workers who are general practitioners. Underreporting (via failure to report) is common and can vary by setting; this may lead to decreased sensitivity and predictive value for disease and trend detection. Failure to report can happen for many reasons including syndromic treatment without testing/diagnosis, stigma of diagnosis, and burden of reporting.

The current surveillance system has several limitations including lack of timeliness which hinders ability to respond to outbreak clusters rapidly; under-reporting impacts the overall representativeness of the data, as well as sensitivity and predictive value of the system. Strategies to address these limitations include innovative surveillance techniques that allow for the earlier detection of disease outbreaks (as near to real time as possible), do not require voluntary reporting on the part of health care providers, have simplified or automatic reporting/alert functions, use multiple sources of data and have the ability for multiple stakeholders to access data (not just public health professionals)

2.4 Sexually transmitted infection outbreak detection and response

The CDC provides guidelines for developing outbreak response plans as part of STI surveillance (53). Guidance includes: 1) developing standards for surveillance and data management, specifically, ways to identify subgroups (risk groups or by geographic areas); 2) a schedule for monitoring and review; and 3) developing a threshold that triggers a response (53). The response to exceeding the threshold will be dependent on many factors such as the risk factors and demographics of the population, geographic area, community involvement, staffing and resources available. The guidelines also suggest that STD programs and surveillance units annually review and evaluate the attributes of their systems according to the CDC Surveillance System Evaluation Guidelines (54) to maximize the ability to detect an outbreak (53).

2.5 The future of public health surveillance

Public health surveillance relies on a complex system of health informatics which includes a variety of data critical to informing public health action (55). The Outcome Assessment through Systems of Integrated Surveillance (OASIS) Project was funded from 1998-2005 as a CDC demonstration project with the goal of: "increasing the use of surveillance data to improve planning and evaluation of public health programs with a specific focus on STIs and other reproductive health outcomes" (56). A primary activity of the OASIS project was to increase the use of technological advancements such as use of GIS and geocoding. New computer technology could improve the quality, timeliness, and effectiveness of surveillance systems (44, 55, 57). Efforts to strengthen the STD surveillance system should focus on advances in technology- such as refinements of electronic databases/platforms, data systems and data integration, which will also heighten the necessity of patient confidentiality and system compliance with security (44, 57). Another technology to explore includes innovative use of the internet. Wilson, et al., reports that Internet based sources of disease information may allow for timely detection of increased rate of disease, reduction of overall cost, and increased reporting transparency (58).

Those under the age of 20 have not known life without the Internet (59). North America has the highest proportionate use of the Internet with 353 million users accounting for internet penetration of 85% (60). The vast majority of all adults (87%) and 93% of people aged 13-29 report regular internet use (2).

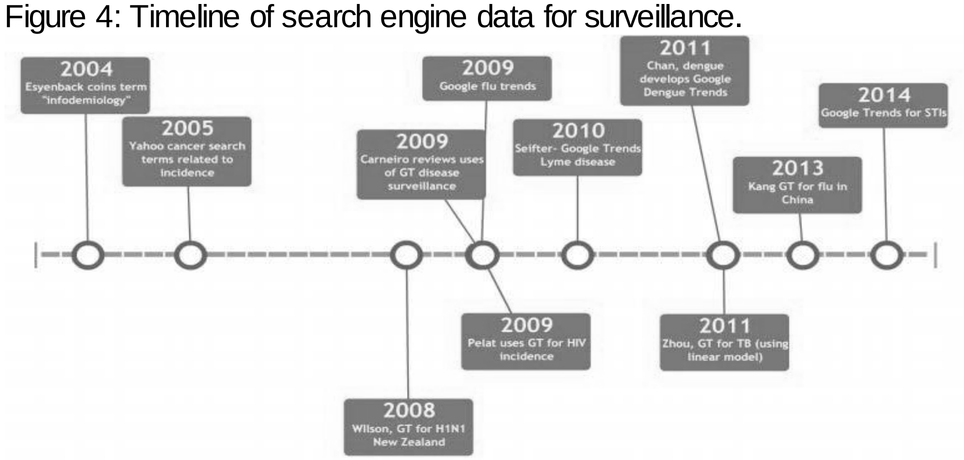
The internet is an important source of health information, as it is anonymous, low-to-no cost, and can be accessed at any time. Millions of people search online for health-related information each day (61). Annually, over 100 million people in the United States use the internet to search for medical information (61), the majority starting their search via a search engine such as Google. Of the US adult population, 78.8% have looked online for health or medical information for themselves in the past 12 months and 67.8% have looked online for someone else (61). Indeed, numerous studies have found that patients often consult the internet about health conditions prior to contacting medical professionals (62). Patients may use search engines, such as Google, to look for keywords that describe their symptoms or they may search by name of disease.

While the internet is not a “new” technology, it is relatively new to surveillance of infectious diseases. In 2004, Eysenbach coined the term “infodemiology” to describe the study of the distribution and determinants of information on the internet, and “infoveillance” to describe syndromic surveillance of disease via the internet (63). Search terms can be downloaded and analyzed over time to identify disease outbreaks or to supplement traditional surveillance methods. Patterns in search traffic can be useful in identifying trends in diseases. One of the key advantages to infodemiology is that data can be collected and analyzed in near real time.

Researchers have used internet search term data to predict trends in infectious diseases, such as influenza, Lyme disease, tuberculosis, norovirus, and chicken pox (58, 64-68). Ginsberg et al. developed the Google based application Google Flu Trends to predict influenza 7-10 days earlier than traditional surveillance systems. Ginsberg et al. found the best predictor of influenza-like-illness (ILI) was the sum of 45 search queries (65). From this groundbreaking work published in 2009 came many subsequent studies. Following the popular Google Flu trends model, influenza has been successfully monitored via search engine data and outbreaks have been predicted in not only the United States but also in China, France and Spain (69-71).

In 2009, Pelat et al., conducted a study to determine the utility of Google Trends using French language search terms for the diseases gastroenteritis, influenza and chickenpox (See figure 4) (71). Correlations with weekly incidence rates of the diseases provided by the sentinel network were calculated using Pearson correlation coefficient. Results of the chosen query terms reflected trends of incidence of both diseases (71). The highest correlation with rates of influenza-like illness was obtained with the multiple word query “grippe-aviaire-cavvin;” which translate from French as "influenza, avian and vaccine" ($r=0.82$, $p <0.01$) (71). A correlation of 0.9 was found for the French word for gastroenteritis (71). The authors concluded that in each disease, one well-constructed query was sufficient to provide time series of searches that highly correlated with incidence; illustrating patterns in search terms were related to trends in disease.

Figure 4: Timeline of search engine data for surveillance.



In 2010, Seifter et al. used Google Trends to approximate trends in Lyme disease (67). As Lyme disease has well-described seasonal and geographic patterns, it lends well to testing the utility of Google Trends. The authors found that the results for the search term “Lyme disease” reflected exposure during summer and spring months, and the search term “cough” had higher relative frequency in the winter months (67). Additionally, the states with the highest frequency for the search terms were the states where Lyme disease is endemic. While there was no information on whether these searches were being conducted prior to or after diagnoses or tests for Lyme disease, this information could be used to analyze trend data prior to the availability of surveillance data. A limitation to this study is its descriptive nature, there were no associations tested between search term volume and rate or timing of disease.

Chan et al, in 2011, evaluated whether or not web search queries are a potential data source for monitoring and detection of dengue epidemics in four different countries: Bolivia, India, Indonesia and Singapore (72). These countries were chosen as they had adequate search volume data and burden of disease. The authors fit a linear model using a time series of Google search query volume for specific dengue-related queries from 2003-2010 (72). The models were found to fit the data well (validation correlations 0.82- 0.99) (72). These data are valuable complement to assist with traditional dengue surveillance which generally suffers from a delay and informed the development of a specific Google trends application to monitor trends in dengue (see www.googletrends.org/denguetrends).

Google Dengue Trends uses search query data in real time to create an index of dengue incidence that serves as a proxy for a traditional surveillance system.

Multiple studies have shown that Google Dengue Trends correlates highly (>0.8) with actual dengue incidence. However, it seems to be a less robust indicator in areas of low incidence (as tested in 14 areas of Mexico) (73). An indicator of real-time dengue activity could serve as a prompt for further investigation and allocation of resources.

Zhou et al., also in 2011, used Google Trends to develop a syndromic approach to estimate the number of tuberculosis cases (68). Using historical CDC data on TB cases by week and quarter and TB-related symptom search terms, a model was fit to estimate national and state prevalence. Authors were able to predict cases of tuberculosis 1 to 12 weeks before traditional reporting systems. The authors suggest that while Google Trends could act as a supplement to established TB surveillance systems, there is potential for Google Trends to act as the surveillance system in international settings where there is no infrastructure or resources for TB surveillance.

Google Trends shows great promise as a timely and sensitive compliment to traditional surveillance systems. Google Trends is likely best suited for surveillance of diseases with high prevalence (70) and is more appropriate to track disease activity in countries with high internet penetration, compared to less developed countries or areas where internet penetration is low (65, 68). There are many advantages to integrating search trends data into traditional surveillance systems, including the potential for earlier detection of disease outbreaks, it is a relatively inexpensive method, it can be automated and results can be shared in real time, and it allows the public, as well as public health professionals, greater access to surveillance information (68, 74).

2.6 Internet use for sexual health information

Several studies have evaluated the percentage of US internet users searching for general health information, which includes information on STIs. Estimates range from 30-78% for the general population (75-77). In a nationally representative study conducted in 1999, 34% of White adults and 19% of African American adults reported searching online for health information. In a 2007 update of the national survey, 34% of Whites, 31% of African Americans and 20% of Hispanics reporting searching for health information online (75). It is reasonable to assume that the proportion of Americans that have searched online for health information will continue to increase over time. Increased education, White race, and lower age are all related to searching for health information online (75). In the most recent Pew Internet and American Life Study report, a nationwide survey of 3,014 adults aged over age 18 in the United States, 72% of all internet users say they looked online for health information in the past year, describing their searches as being related to serious health conditions, general health information or minor health problems (61). Seventy-seven percent of those users began their searches at search engines such as Google, Bing or Yahoo (61). An additional 13% reported beginning their search at a site with specific medical content, such as WebMD (61). Half of all health information searches are on behalf of someone else (78). Thus, using the internet to seek health information is a highly prevalent behavior for American adults.

In a Kaiser Family Foundation study in 2001, 15-24 year olds were surveyed to explore their internet behaviors in regard to health information (79). This is the only nationally representative sample of adolescents and young adults surveyed on internet health seeking behavior (79). Seventy-five percent of all young people surveyed reporting using the internet at least once to find health information, this was about the same proportion that reported ever playing games online (72%) or ever downloaded music online (72%) (79). Forty-four percent of respondents looked up information about sexual/reproductive health such as pregnancy, birth control, HIV or STDs. Not only are young people viewing health-related information, but they are also being moved to action. Four out of ten (39%) respondents reported behavior change due health information retrieved online and 14% reported seeing a medical provider because of the information they read online (79).

In a 2003, Rietmeijer et al. found that only 28% of 4,741 STD clinic patients sampled use the internet to search for information about STIs (80). Men who have sex with men (MSM) were more likely than MSW or women to use the internet to search for STI related information; Whites were more likely than Blacks (32.4% vs. 21.4%); and those who used condoms consistently versus inconsistently (33.9% vs. 27%) were more likely to use the internet to search for information (80). Nearly two-thirds (65%) of respondents were searching for general information on STDs, and 36% reported searching for information specific to HIV (36%) (80). Similar proportions reported searching for content specific to chlamydia or gonorrhea (22% and 19% respectively) (80). Study results may be somewhat outdated, over 10 years later; however the results suggest that men who have sex with men may differentially access the internet to search for STI related information. The male-to-female ratio of syphilis is very high (3); if the findings hold true in our study this may influence the overall representativeness of the syphilis search term results (may apply more directly to the populations generating the most search terms).

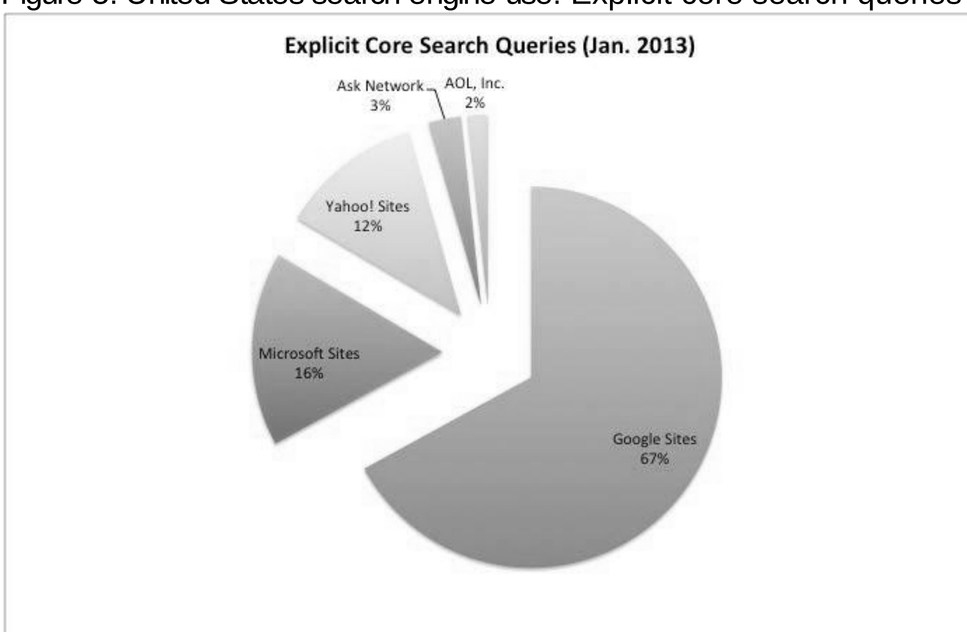
From 2008-2010, in a cross-sectional self-administered survey regarding internet use, a sample of 3,181 women ages 16-24 attending one of five publicly funded reproductive health clinics, study authors found racial and ethnic disparities in internet use for health information (76). Overall, the majority of White (92.7%) and Black (92.9%) women used the internet, compared to only 67% of Hispanic women. White women (79.2%) reported using the internet to find health information more often than either Hispanic (74.3%) or Black women (70.3%) (76). Compared to White women, Blacks and Hispanics were about 25% less likely to seek information on contraception but were approximately 40-60% more likely to seek information on pregnancy testing (76). Overall this study found a much higher percentage of young women, of all three race/ethnicities searched online for health information than previously reported. As this is a more recent study, differences in study findings may be caused by increased internet access over time and different sociodemographics of samples. Greater information seeking on pregnancy tests and STIs among Black and Hispanic women could be due to significantly higher rates of unintended pregnancy and STIs in these two groups. The study had several limitations, data was self-reported, the study was cross- sectional, and results may not generalize to other populations such as young men, older women or a higher income population. However, the results of the study inform the degree of representativeness we can expect of internet search engine surveillance, in that Hispanic young women may be less represented in comparison to Whites or Blacks. In addition, this study helps to characterize a population presumably exposed to STIs.

A primary weakness in each of these studies is that none measured the reason for internet search for STI information, whether this was directly related to STI risk or exposure, impacted health care seeking or other behaviors, or the search terms they used. Additional information characterizing populations who search for STI information is needed to interpret and make use of observed trends. Our study addresses these gaps, and will characterize the sexual risk behavior, their history of STIs, as well as how, when and why internet users conduct STI-related searches.

2.7 Google Trends: Overview

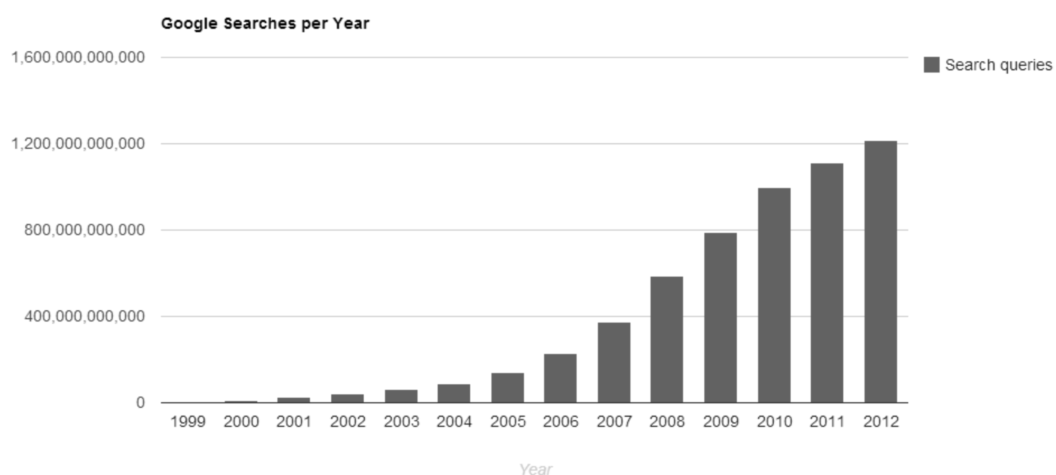
Google is the most widely used general search engine in the world. The majority of online searches in the United States begin with Google (67%, see Figure 5). When Google was founded in 1998 it was processing ten thousand searches a day; by the end of 2006 Google processed this same amount in one second. Google's Senior Vice President, Amit Singhal, reported in 2012 that Google searches over 30 trillion unique universal resource locators (URL), 3.3 billion searches per day at a speed of over 38,000 per second (81). Figure 6 displays Google's exponential growth in searches per year from 1999-2012. Google is an important portal of online health information for patients as it is often the point of entry (77,78,82,83).

Figure 5: United States search engine use: Explicit core search queries 2013 (81)



In 2008, Google launched Google Insights for Search which provided data on the search terms entered into the Google's search function. Google Insights was able to provide a visual representation of interest by region and also displayed top searches. In 2012, Google merged Google Insights for Search into Google Trends and shut down the insights page.

Figure 6: Google total searches per year, 1999-2012 *(81)*



Google Trends analyzes internet search data to record and report how many searches are completed for specific words/terms. It identifies and ignores redundant searches from unique devices by using the Internet Service Provider (ISP) address (84). Search terms need to exceed a threshold of search frequency to appear in results. Search volume is normalized and presented on a scale of 0 to 100. Volume is relative to the time period selected and is normalized via an automated procedure. Normalization of the data allows comparisons to be made without regard to total search volume. This information is displayed in a search volume index graph. There is an option to view news (e.g., media) reference volume graph as well which shows the frequency with which the web search queries appeared in Google News Stories. Google does not disclose the method for determining the search threshold nor the normalization procedures used.

Google's search algorithm can detect and correct misspellings. Google estimates between 10-20% of all searches contain a misspelled word (84). Statistical machine learning is able to process search click-throughs and provide accurate results based on previous and simultaneous searches. For example, if a user types a misspelled word into Google and doesn't click on any results, then corrects the misspelling and proceeds to click on results from the second (correct) search, Google's machine learning is adaptive and stores this information.

Figure 7: Annotated Google Trends output

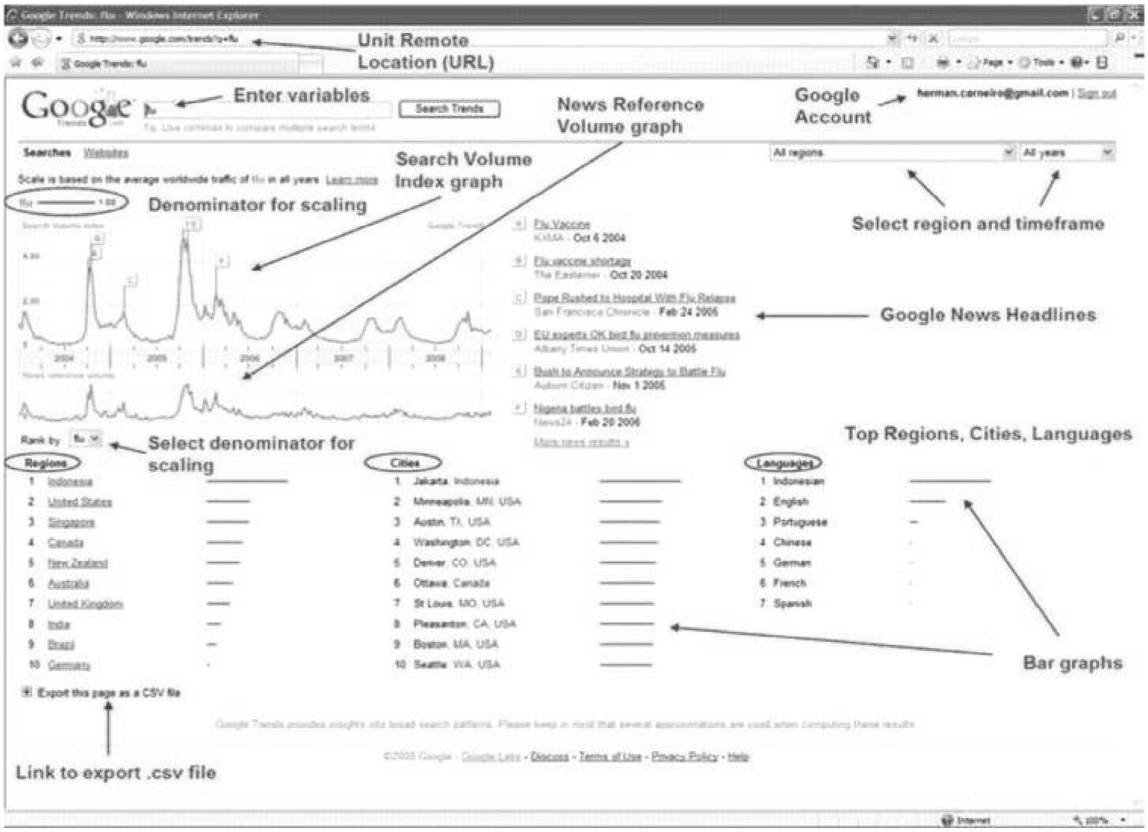


Figure 7 displays an annotated output of Google Trends for the web search term "flu" worldwide from January 2004 to March 2009 (70). Prior to using Google trends users must first create a Google account via Gmail, the free Google email service. Users can then log in to their Gmail account and access Google trends. Users enter the Web search queries in the text box next to the Google logo ("enter variable" in Figure 7) to see the relative search volume of these queries. Users also select the region (country) of interest and the timeframe of the search. The search information is displayed in a search volume index graph. Below the search volume index graph is an area that contains the news reference volume graph. This graph shows the relative frequency of the selected search term in relation to News or Media stories found on Google. When a spike is detected in the news/media graph, Google Trends labels the search volume graph with a letter and provides a link to the news media story. The top geographic areas (regions, cities) and languages with the highest relative search volume are displayed on the bottom of the webpage. Google reports that trends data is updated daily (84).

3. **STUDY 1**

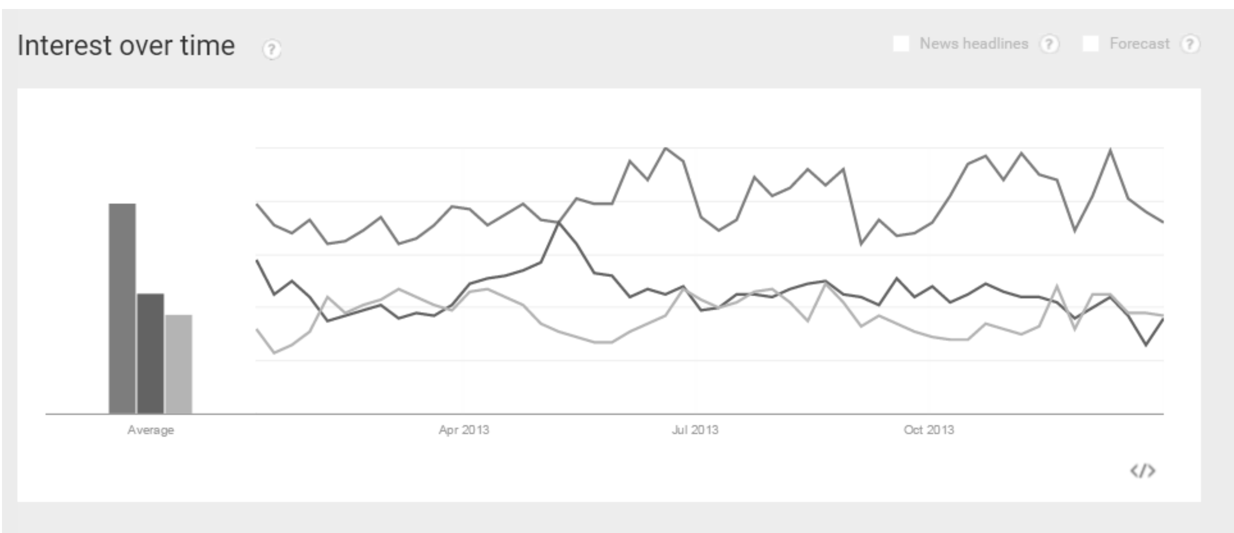
In our previous publication, we used publicly available Google Trends data to examine the association between STI search terms and rates of disease (2011) by US state (see Appendix A) (90). We found a positive and statistically significant correlation between rate of gonorrhea by state and the volume of a gonorrhea search term (90). To determine whether our previous results would be replicable, a series of STI-related search terms were compared to rate of disease by US state for 2013, the most currently available STD surveillance data. As trend data for some major metropolitan statistical areas are now available in the Google Trends functionality, we also assessed this.

We replicated the methods of our previous analysis (see Appendix A). Briefly, we entered search terms into the publicly available Google Trends website and used the export data function to download a comma separated value spreadsheet to access search volume of selected term by US state. We then compared volume to STD rates by state using the most recent national and publically available STD data (108). Bivariate correlations were used to determine the strength and direction of the relationship between search term volume and rate of disease (chlamydia, gonorrhea, and syphilis) by US city and state for 2013.

The search terms used were “STD symptoms”, “chlamydia symptoms”, “gonorrhea symptoms”, and “syphilis symptoms”. Both a broad term (“STD symptoms”) and specific terms incorporating the disease-name were used to determine if associations differed by variation of search term.

Overall the relative volume of the chlamydia search term is the greatest in 2013, while gonorrhea and syphilis are lower and similar to each other (Figure 8). This distribution replicates our study findings using 2011 STD rates (90). Figure 8 displays search volume over time for 2013 for our three specific search terms (“chlamydia symptoms,” “syphilis symptoms,” and “gonorrhea symptoms”) for the United States. We display the total search volume to demonstrate the distribution of the relative volume of each search term used.

Figure 8. Search volume of “chlamydia symptoms,” “gonorrhea symptoms,” and “syphilis symptoms” for the United States in 2013



For the general term “STD symptoms”, 28 states had search volume that exceeded the minimum threshold and was provided publicly by Google Trends. The positive association between search volume and rate of disease was statistically significant for gonorrhea (*r* = 0.70, p<0.05) and for chlamydia (*r* = 0.66, p<0.05), however it was not statistically significantly associated for syphilis (*r* =0.38, p=0.06).

When we tested the “gonorrhea symptoms” search volume against the rate of gonorrhea by state for 2013 there was a positive and significant association (*r* = 0.69, p<0.05). In addition, we also found positive and significant associations with “chlamydia symptoms” and rate of chlamydia by state for 2013 (*r* = 0.58, p<0.05) and “syphilis symptoms” by rate of syphilis (*r* = .48, p<0.05).

These positive and significant correlations can be visualized by comparing search trend volume intensity mapped to STD rates mapped to state. There is an overlap between the states that have high levels of STI-related search volume and high rates of disease (figure 9); this was true for chlamydia, gonorrhea and syphilis (evidenced by the positive and significant *r* between search term volume and rate of disease). For example, figure 9 depicts search volume for “STD symptoms” by US state, with darker color indicating increase search volume (A), also depicted for comparison is the CDC generated gonorrhea rate per 100,000 by US state (B) also with darker colors indicating increased rates.

The search term “chlamydia symptoms” was positively associated with chlamydia rates by city (*r* = 0.69, p<.05) for 2013. In contrast, “gonorrhea symptoms” was not associated with chlamydia rate by city (*r* = 0.01, p=0.98) nor was “syphilis symptoms” with rate of syphilis by city (*r* = -0.06, p=0.89).

In this updated analysis, both the broad term “STD symptoms” and terms using specific disease name were associated with rates of chlamydia, gonorrhea and syphilis; this is consistent with our previous findings (90). The correlation coefficients are similar compared to the associations detected using the 2011 data; however, using 2013 data we also detected associations with chlamydia and syphilis. The broad STD term was not associated with syphilis by state, only the specific “syphilis symptom” term was associated with rate of syphilis by state.

Figure 9. Google Trends generated United States map of search volume by state for "STD Symptoms" (A) and Gonorrhea rate per 100,000 (B), 2013

A.

B.

In refining the geographic unit from US state to city, there was less information available from Google Trends. While the chlamydia specific search term was related to volume by city there were only 10 cities with volume that exceeded the threshold for public display. Google Trends does not publish search volume for geographic regions unless it exceeds a specific (unpublished) threshold. The search volume was only available for 8 cities for the gonorrhea term and 7 cities for the syphilis term. Google has opened a developer's application for Google Trends for health researchers to access data that is not available on the public page. Gaining access to the Google Trends health researchers console will allow a more accurate investigation of whether or not an association is present between search volume and rates of STIs on a city-level.

These results are the first step in understanding the overall utility of Google Trends as an additional surveillance tool and resource for monitoring trends and outbreaks of disease. The findings have implications for building multiple query prediction models, in that specific terms may have more weight in predicting trends in disease (109). In addition, user search content may vary by disease information sought. It is important to understand and characterize the content of internet searches in order to build the most reliable, sensitive and specific predictive models to aid STI surveillance.

4. **STUDY 2.**

National surveillance of sexually transmitted infections (STIs) relies on mandatory case reporting, a system that produces data that is often incomplete and limited in scope. In addition to data quality issues, underreporting of cases as well as delays in reporting results in missed opportunities to predict and respond to trends in disease as well as decreased ability to guide STI control.

As the internet is a portal for free and anonymously available health information, search engine data may provide an additional venue for STI surveillance efforts leading to earlier detection of trends and increased ability to monitor impact and geographic spread. Search engine data has the potential to be an efficient and economical enhancement to the established surveillance system (64, 74).

Google Trends allows the download of de-identified search engine data trends, which can be used to investigate the implications of trends in STI-related search terms in relation to STI rates. In our previous study, we demonstrated STI-related search terms were statistically significantly and positively correlated with STI rates by US city and state (90). While further in-depth analyses are needed to investigate and replicate this association, it remains unknown whether or not Google Trends is acceptable and perceived to be useful to current STI surveillance workers. The current study sought to provide information about integrating Google Trends from the perspective of STI surveillance employees.

The potential for integrating internet search trend data into the STI surveillance system was assessed through an online cross-sectional survey of a convenience sample of US health department STI surveillance employees. The survey was designed to generate information on the feasibility, desirable attributes, and potential implementation process necessary to integrate search engine data into the current system (see appendix B). Key questions in the survey were designed to capture barriers and facilitators as well as overall interest in Google Trends. Prior to data collection the survey was pilot tested with two participants who are involved with STI surveillance. The Institutional Review Board from the University of Illinois at Chicago reviewed the study and approved it as exempt.

The National Association of County and City Health Officials (NACCHO) and the National Coalition of STD Directors (NCSD) distributed the survey link to their members via email in April 2015. The National Coalition of STD Directors included a short description of the survey and a link to participate in their email newsletter, which is sent on a bi-monthly basis to their membership. As NACCHO's membership encompasses health officials in domains outside of STIs, the survey description and link was distributed via email to a subcommittee focused on HIV, AIDS and STIs as well as their general membership.

Surveys were administered using an online platform called Qualtrics. Qualtrics is an Application Service Provider (ASP) with a platform for creating and distributing online surveys (96). The survey opened with an eligibility screener to determine if a participant was: 1) 18 years of age or older; and 2) currently employed in a US STI surveillance unit.

Data analysis was conducted in Stata version 13 (110). Univariate analysis was used to describe the population surveyed and to present general descriptive information of the variables collected. The outcomes of interest, anticipated utility, and willingness to integrate Google Trends were compared by demographic and situational or environmental factors in bivariate analysis using chi-square tests or Fisher's exact test (when cell sizes were less than 5) to compare differences in proportion.

The survey link was included in four email newsletters, sent on separate dates in April 2015. On average, across the 4 emails, 28% of recipients (about 306) opened the email containing the link to the survey.

A total of 77 respondents, 25% of all who opened the email, completed the eligibility screening questions between April 6th and April 30th 2015. All respondents were 18 years old or older and 83.1% were currently employed in a surveillance unit in a health department in the United States. Thirteen respondents did not meet the eligibility criteria and were exited out of the survey. Of the 64 respondents who were eligible to participate, 91% completed the entire survey. The six respondents who did not complete the entire survey dropped out at various points, resulting in a rate of 10-50% completion. Survey completion took a median of 10 minutes (IQR=5-10).

The majority of participants were aged 25-44 (55%), 70% were female, 85% reported having a graduate degree and 40% reported being epidemiologists (TABLE I). Survey respondents were geographically diverse with the highest percent from the West (29%) and the fewest from the South (21%).

TABLE I. PARTICIPANT AND HEALTH DEPARTMENT CHARACTERISTICS, N=64

Characteristic	N (%)
Age	
18-24	1 (1.7)
25-34	19 (32.8)
35-44	13 (22.4)
45-54	14 (24.1)
55-34	11 (19.0)
Sex	
Female	14 (69.5)
Male	18 (30.5)
Education	
College graduate	9 (15.5)
Graduate degree	49 (84.5)
Job title	
Epidemiologist	23 (39.7)
Manager	16 (27.6)
Director	9 (15.5)
Assistant	2 (3.4)
DIS	3 (5.2)
Nurse	3 (5.2)
Chief or commissioner	1 (1.7)
Deputy	1 (1.7)
How many people are employed in your division or unit of the health department?	
1-9	10 (20.4)
10-24	18 (36.7)
25-50	7 (14.3)
51-75	3 (6.1)
75+	11 (22.4)
How long have you been in your current position?	
Less than 1 year	10 (20.4)
1-5 years	25 (51)
6-10 years	4 (8.2)
11-15 years	3 (6.1)
15+ years	6 (12.2)
Geographic area	
Northeast	14 (24.1)
West	17 (29.3)
Midwest	15 (25.9)
South	12 (20.7)

TABLE I. PARTICIPANT AND HEALTH DEPARTMENT CHARACTERISTICS, N=64 (continued)

Characteristic	N (%)
Preferred learning style for new technology	
Webinars	48 (75)
Small group sessions	38 (59.4)
Conference events	28 (43.8)
One on one sessions	26 (40.6)
Talking to a coworker	18 (28.1)
Large group sessions	13 (20.3)
Hard copy literature	15 (23.4)
Health Department STI outbreak & response	
Method to detect	
Syphilis	49 (76.6)
Gonorrhea	41 (64.1)
Chlamydia	35 (54.7)
Outbreak response plan	
Syphilis	44 (68.8)
Gonorrhea	31 (48.4)
Chlamydia	19 (29.7)
Responded to outbreak in past 3 years	
Syphilis	26 (40.6)
Gonorrhea	17 (26.6)
Chlamydia	8 (12.5)

The majority of the sample was well connected to technology: 91% indicated that they used applications (“apps”) on a smartphone or tablet a few days a week or every day. Over half (56%) of participants had heard of Google Flu Trends prior to taking the survey. Overall participants reported high levels of interest in a Google Trends application for STIs, with 72% reported being very interested. However, just over half (55%) of participants reported Google Trends would be very useful and only 34% reported they would be very likely to integrate the new tool into their current work.

In terms of challenges and barriers to integrating Google Trends into the current surveillance system, 39% reported that their department was too busy to integrate new tools, 12% reported that the current system does not need to be modified and 8% reported that Google Trends would not be valuable. Surveillance system attributes that were rated very important were: access to data in real time (94%), improved sharing of information (84%), automatic alerts for outbreaks (84%), support from administration for advanced techniques (75%), and increased understanding of the disease (72%). About half of participants reported having a person ready to assist with technical difficulties (53%) and systems that do not increase workload (51%) were very important.

We hypothesized demographic and individual level factors may be associated with higher levels of interest, perceived utility and willingness to integrate. Females more frequently indicated they were very interested in integrating Google Trends compared to males (TABLE II). Participants of the oldest aged category (55-64) were less likely to report they are very interested in Google Trends. Those employed less than five years were more likely to report being very interested in Google Trends. A greater proportion of respondents from the South region were very interested and reported very useful. Those who reported it would be “very likely” that Google Trends could improve outbreak response were more likely to report being “very interested” and that Google Trends would be “very useful”, but were not more likely to report being “very interested” in integrating Google Trends into current surveillance systems.

In terms of environmental factors (Table II), participants who agreed that their department is too busy to integrate new tools were less likely to report being “very interested”; likewise those who agreed that the current system did not need change were less likely to be very interested and less likely to perceive Google Trends as very useful. However, those who agreed it was very important to increase understanding of the disease under surveillance were more likely to report being very interested in Google Trends. Those who agreed it was very important for in-person technical assistance reported less interest in Google Trends. While not statistically significant, if respondents reported not having an outbreak response plan for reportable STIs they were more likely to be interested in Google Trends. Similarly having no method to detect outbreaks in STIs was associated with being more willing to integrate Google Trends into the current system. No other factors were associated with being more willing to integrate Google Trends into the current system

TABLE II. CHARACTERISTICS REALTED TO INTEREST IN AND PERCEIVED UTILITY OF GOOGLE TRENDS APPLIED TO SURVEILLANCE, N =64

	Interest		*p-value	Utility		p-value
	Somewhat	Very		Somewhat	Very	
Demographic characteristics Sex						
Male	4 (33.3)	8(66.7)	0.54	6 (28.6)	15 (71.4)	0.78
Female	9 (24.3)	28 (75.7)		7 (25.0)	21 (75.0)	
Age						
18-34	5 (26.3)	14 (73.7)	0.01	9 (47.4)	10 (52.6)	0.32
35-44	1 (10.0)	9 (90.0)		4 (40.0)	6 (60.0)	
45-54	0	10 (100)		2 (20.0)	8 (80.0)	
55-64	6 (60.0)	4 (40.0)		6 (60.0)	4 (40.0)	
Job title						
Manager	2 (14.3)	12(85.7)	0.34	9 (64.3)	5(35.7)	0.16
Epidemiologist	7 (35.0)	13(65.0)		7 (35.0)	13(65.0)	
Other surveillance worker	3 (20.0)	12 (80.0)		5 (33.3)	10 (66.7)	
Education						
College graduate	1 (14.3)	6 (85.7)	0.50	2 (28.6)	5 (71.4)	0.41
Graduate Degree	11 (26.2)	31 (73.8)		19 (45.2)	23 (54.8)	
Length of time employed						
Less than 1 year	0	10 (100)	0.10	2 (20.0)	8 (80.0)	0.29
1-5 years	7 (28.0)	18 (72.0)		12 (48.0)	13(52.0)	
5+ years	5 (38.5)	8 (61.5)		6 (46.2)	7 (53.8)	
Region						
Northeast	4 (33.3)	8 (66.7)	<0.01	6 (50.0)	6 (50.0)	0.30
West	4 (30.8)	9 (69.2)		5 (38.5)	8 (61.5)	
Midwest	3 (21.4)	11 (78.6)		8 (57.1)	6 (42.9)	
South	1 (10.0)	9 (90.0)		2 (20.0)	8 (80.0)	
Situational and structural factors						
Likelihood GT could improve outbreak response						
Somewhat likely	18 (51.4)	17 (48.6)	<0.01	27 (77.1)	8 (22.9)	<0.01
Very likely	0	28 (100)		1 (3.6)	27 (96.4)	
My department is too busy to integrate new tools						
Agree	12 (48)	13 (52.0)	0.01	13 (52.0)	12 (48.0)	0.44
Disagree	6 (15.8)	32 (84.2)		15 (39.5)	23 (60.5)	
Current system does not need to be modified						
Agree	7 (100)	0	<0.01	7 (100)	0	<0.01
Disagree	11 (19.6)	45 (80.4)		21 (37.5)	35 (62.5)	
Using a system that does not increase my workload						
Somewhat important	9 (30.0)	21 (70.0)	0.75	17 (56.7)	13 (43.3)	0.09
Very important	9 (24.5)	25 (73.5)		12 (35.3)	22 (64.7)	
Increase my understanding of disease under surveillance						
Somewhat important	9 (50.0)	9 (50.0)	0.02	10 (55.6)	8 (44.4)	0.30
Very important	9 (19.6)	37 (80.4)		19 (41.3)	27 (58.7)	
Support from administration						
Somewhat important	4 (25.0)	12 (75.0)	0.75	5 (31.3)	11 (68.7)	0.19
Very important	14 (29.2)	34 (70.8)		24 (50.0)	24 (50.0)	
Tools that have automatic alerts						
Somewhat important	6 (46.1)	7 (53.9)	0.14	6 (46.2)	7 (53.8)	0.98
Very important	12 (25.0)	36 (75.0)		22 (45.8)	26 (54.2)	
Having someone to help						
Somewhat important	12 (40.0)	18 (60.0)	0.08	15 (50.0)	15 (50.0)	0.53
Very important	6 (19.4)	25 (70.6)		13 (41.9)	18 (58.1)	

* Chi-square p-value. Fisher's exact test when cell N <5.

Strategies to address limitations of the current surveillance system include innovative surveillance techniques that allow for the earlier detection of disease outbreaks, have simplified or automatic reporting/alert functions, use multiple sources of data and have the ability for multiple stakeholders to access data (not just public health professionals) (44, 55, 57, 58). Results from this survey indicate high levels of agreement on these enhanced features from a geographically diverse sample of STI surveillance workers.

The overall sample size was smaller than anticipated with a response rate of approximately 25%, thus results will not be reflective of the entire target population and need to be interpreted with caution. As the predictive capability of Google Trends for reportable STIs has not yet been demonstrated, this survey may have been premature and respondents may not have had enough information about the application of Google Trends to STI surveillance to inform their responses. Despite the small sample size, respondents were geographically diverse having similar proportions of respondents in each of the four regions of the United States. Results will be biased if surveillance workers who perceive a benefit to integrating new technology into the current surveillance system were more likely to participate compared to workers without this same belief.

Our previous study demonstrated an association between STI-search terms and rates of STIs by state; however more complex analysis is needed to determine if the association remains at smaller geographic units (e.g., zip code) and by user demographics (90). Should there be an opportunity to utilize Google Trends for STI surveillance, the current study provided insight into preferred learning methods, barriers and facilitators to integration. With a high level of reported interest and the preferred learning method of webinar, this would potentially facilitate the scale up and adoption of

Google Trends for STIs. However, further research is needed to understand the practical application and integration of Google Trends into STI surveillance from a worker perspective.

5. **STUDY 3a.**

The current national surveillance system for sexually transmitted infections (STIs) has several limitations including lack of timeliness, which impedes rapid response to outbreaks, as well as under reporting which impacts the overall representation, sensitivity and predictive value of the system. Internet search engine data may provide an additional venue for STI surveillance efforts (63, 64, 70). As STIs are stigmatized, many may query the internet prior to getting tested or seeking treatment. Improvements to the surveillance system via search engine data could lead to earlier detection of outbreaks and trends in disease.

Google Trends generates easily accessible and publicly available search engine data which can be used to investigate trends in STI-related search terms in relation to STI rates. Search terms can be downloaded and analyzed over time to identify disease outbreaks or to supplement traditional surveillance methods. Patterns in search traffic can be useful in identifying trends in diseases. There are several advantages to integrating internet search trends data into traditional surveillance systems, including the potential for earlier detection of disease outbreaks, it is a relatively inexpensive method, it can be automated and reported in real time, and it allows the public as well as public health professionals to have greater access to surveillance information (68, 74).

In our previous study, we demonstrated that STI-related search terms were statistically and significantly positively correlated with STI rates by US state (90). However, the characteristics of internet search engine users who are generating the STI-related search terms remain unknown.

Several studies estimate the percent of US internet users searching for general health information to be between 30-78% (75-77). Individuals may turn to Google, using keywords to describe their experienced symptoms or to identify symptoms of diseases they may have.

While some studies have examined STI-related internet searches (75, 76, 80), none measured the reason for internet search for STI information, whether this was directly related to STI risk or exposure, impacted health care seeking or other behaviors, or the search terms they used. Additional information characterizing populations who search for STI information is needed to interpret and make use of observed trends, especially as it applies to using Google Trends for surveillance. We characterized internet search engine use among to determine the reason and timing of STI-related searches.

Two populations of interest were sampled to describe internet user characteristics and STI search behavior: 1) patients attending Chicago Department of Public Health (CDPH) STD Clinics; and 2) students at the University of Illinois at Chicago (UIC). The STD clinic population is a higher risk population; as they are seeking services at an STD clinic they are likely to be exposed to STIs. The student population represents a lower risk (for STIs) group compared to the STD clinic patients.

The Institutional Review Boards (IRBs) of the University of Illinois at Chicago and Chicago Department of Public Health approved the study. The study was anonymous and no identifying information was asked or recorded. Participants were eligible for the study if they: 1) were between the ages of 18-35 (inclusive); 2) could speak and understand English; 3) were receiving services at the STD clinic (clinic population only); and 4) were able and willing to provide oral informed consent. Survey completion took between 5 and 10 minutes and participants did not receive compensation.

At the STD clinics, patients' registration information was used as a primary screening tool to determine if they were eligible based on age and receiving services at the clinic; if these two criteria were met, study staff screened for eligibility and if eligible, conducted study procedures (informed consent, survey administration). The majority of students were recruited from a general University event that draws over 1000 students, students were also recruited from the University Wellness Center waiting room. Students were approached (in both the event and Wellness Center) and asked if they wanted to screen for participation; if eligible, staff conducted study procedures.

Based on previous studies, we estimated the prevalence of STI-related internet search engine use to be 30-49% (76, 80, 101). Using a conservative estimate of outcome prevalence of 30%, 307 subjects were needed in each sub sample (clinic and student) to have $\geq$80% power to detect a difference in exposure (e.g., past history of STIs, greater than high school education) of 15%, for exposure prevalence ranging from 10-50% with a two-sided test of significance.

All survey data was collected via paper-pencil format and was subsequently entered into a database and analyzed by the first author. Data collection included demographic, sexual risk behavior, and internet search behavior (see Appendix C for survey). The STD Clinic and student surveys were nearly identical, with the removal of one clinic specific question from the student survey (e.g., "why did you come to the clinic today?").

All data analysis was carried out in Stata version 13 (110). The goals of the analyses were: (1) to identify factors associated with using the internet for STI related information, and (2) to describe reasons for searching for STI information online. The primary outcome variable for the clinic population is searching for STI related information online prior to coming to the STI clinic, dichotomized as yes vs. no; the primary outcome for the student population is ever searching for STI information online, dichotomized as yes vs. no. Frequencies and medians are used to describe the sample and bivariate analysis was used to detect differences between categorical variables and outcomes using chi-square test. Due to lack of model convergence under the log-binomial model, Poisson regression with robust variance estimator was used to estimate incidence rate ratios (IRRs). Variables significant at the $p<0.20$ level were entered into a multivariable model using forwards stepwise entry. Those variables with a p-value <0.20 were maintained in the multivariable model. All two-way interactions were tested.

At the clinic, 612 patients were screened for eligibility, of those 472 (77%) were eligible. Of those who were ineligible, 95% were too old to participate in the study. Of those who were eligible, 452 (96%) consented to participate in the study, and 446 (98%) completed the survey. The 6 subjects who did not complete the survey after consenting were called for their medical visit and did not return to complete the survey. Three hundred and thirty-one students were approached for participation, 289 (87%) were screened for eligibility, 282 (98%) were eligible and 279 (99%) consented and completed the survey.

A total of 446 participants completed the survey at one of three Chicago Department of Public Health STD clinics over a four month time period (TABLE III). The median age of participants was 24 years old, 11% aged 18-19, 41% aged 20-24, 31% aged 25-29 and 16% aged 30-35. Just over half (51.4%) of the participants identified as Non-Hispanic Black, followed by Non-Hispanic White (29.5%) and Hispanic/Latino of any race (19.1%). More males (57.3%) than females (42.7%) completed the survey. The majority of the sample had at least a high school diploma, with 32% reporting some college and 29.6% reporting a college or graduate degree. Over one-third (37%) of participants reported having no insurance coverage, with lower proportions reporting parent's insurance plan (20.5%), government plan (23.2%) and private insurance (19.1%). Previous STI diagnosis was common (50.8%) and the most common reason for visit was having STI symptoms (57.4%) followed by regular testing (42.9%), and partner recommendation to get tested (32.3%) (multiple answers possible). Nearly three-fourths (72%) reported 2 or more sex partners in the past 6 months.

Almost all participants (98.7%) reported access to the internet, with 91.3% of clinic participants accessing the internet on their smart phone or tablet (TABLE IV). The majority (93.7%) reported previously searching for medical topics online. Eight-nine percent of participants endorsed ever using the internet to search for STI information.

TABLE III. CLINIC PARTICIPANT CHARACTERISTICS OVERALL AND BY OUTCOME

		*Searched for STI information prior to clinic visit		
	Total N=446	Yes, n=252	No, n=161	Chi-square
	n (%)	n (%)	n (%)	
Age, in years				
18-19	52 (11.7)	32 (65.3)	17 (34.7)	0.46
20-24	185 (41.5)	101 (56.7)	77 (43.3)	
25-29	138 (30.9)	83 (64.8)	45 (35.2)	
30-35	71 (15.9)	36 (62.1)	22 (37.9)	
Race/Ethnicity				
NH White	127 (29.5)	87 (73.1)	32 (26.9)	<0.01
NH Black	221 (51.4)	96 (47.8)	105 (52.2)	
Hispanic/Latino, any race	82 (19.1)	55 (71.4)	22 (28.6)	
Sex				
Male	255 (57.3)	143 (61.9)	88 (38.1)	0.68
Female	190 (42.7)	109 (59.9)	73 (40.1)	
Sexual Behavior				
MSW	192 (44.1)	108 (62.8)	64 (37.2)	0.90
WSM	138 (31.7)	81 (60.4)	53 (39.6)	
MSM & MSWM	56 (12.9)	31 (58.5)	22 (41.5)	
WSW & WSWM	49 (11.2)	26 (57.8)	19 (42.2)	
Highest level of education				
Less than HS	32 (7.2)	9 (37.5)	15 (62.5)	0.01
HS graduate	137 (30.9)	71 (58.2)	51 (41.8)	
Some college	143 (32.3)	77 (55.4)	62 (44.6)	
College/graduate degree	131 (29.6)	94 (74.6)	32 (25.4)	
Insurance Status				
Parents plan	90 (20.5)	62 (73.8)	22 (26.2)	0.01
Government plan	101 (23.2)	51 (53.7)	44 (46.3)	
Private plan	83 (19.1)	50 (66.7)	25 (33.3)	
No coverage	161 (37.0)	83 (55.7)	66 (44.3)	
Reason for clinic visit				
STI symptoms				
Yes	253 (57.4)	149 (62.9)	88 (37.1)	0.36
No	188 (42.5)	101 (58.4)	72 (41.6)	
Partner told you				
Yes	142 (32.2)	85 (63.0)	50 (37.0)	0.56
No	299 (67.8)	165 (60.0)	110 (40.0)	
Test regularly				
Yes	189 (42.9)	102 (57.3)	76 (42.7)	0.18
No	252 (57.1)	148 (63.8)	84 (36.2)	

TABLE III. STD CLINIC PARTICIPANT CHARACTERISTICS OVERALL AND BY OUTCOME (continued)

	Total	**Total Yes,**	**Total Yes,**	**Total**
	n (%) n	n (%)	n (%)	n (%)
STI				
Yes	222 (50.8)	109 (54.5)	91 (45.5)	0.01
No	215 (49.2)	138 (67.0)	38 (33.0)	
Number of Sex partners past 6m				
1 partners	117 (26.8)	67 (61.5)	42 (38.5)	0.91
2-3 partners	177 (40.6)	100 (60.2)	66 (39.8)	
4+ partners	142 (32.6)	81 (62.8)	48 (37.2)	
Condom used last sex				
Yes	165 (37.7)	96 (61.5)	60 (38.5)	0.84
No	273 (62.3)	152 (60.6)	99 (39.4)	

*33 missing responses- does not apply due to skip pattern if no internet or did not use internet for sexual health searches

*MSW= Men who have sex with Women; WSM=women who have sex with men; MSM= men who have sex with men; WSW=women who have sex with women; MSWM= men who have sex with women and men; WSWM= women who have sex with women and men

TABLE IV. INTERNET ACCESS AND BEHAVIOR BY STUDY SAMPLE, CLINIC AND STUDENT

	STD Clinic, N=446 n	**Students, N=279 n**
Access to the internet	440 (98.7)	277 (99.3)
Access internet on smart phone or tablet	407 (91.3)	268 (96.8)
Access internet on laptop or desktop	305 (68.4)	207 (74.7)
Access internet at home	421 (94.4)	273 (98.6)
Access internet at friends/family's	187 (42.0)	199 (71.8)
Access internet in public spaces	255 (57.2)	230 (83.0)
Ever searched for medical topics	418 (93.7)	206 (74.4)
Frequency of search past 3m		
1-2 times/month	262 (62.7)	134 (65.0)
1x/week	85 (20.3)	39 (18.9)
2-3x/week	38 (9.0)	23 (11.2)
Daily	36 (8.6)	9 (4.4)
Reason for search		
School	62 (14.8)	79 (38.3)
Work	48 (11.5)	48 (23.3)
Condition you have	315 (75.4)	129 (62.6)
Condition someone else has	185 (44.3)	84 (40.8)
Ever search for STI information	400 (89.7)	181 (65.3)
Ever search for genital health information	343 (76.9)	85 (30.7)
Most recent search for STI information*		
Today	123 (29.8)	8 (4.4)
1-3 days ago	100 (24.2)	46 (25.1)
Within the past week	47 (11.4)	42 (23.0)
Within the past 2 weeks	25 (6.1)	16 (8.7)
Within the past month	42 (10.2)	13 (7.1)
More than a month ago	76 (18.4)	74 (40.4)
When do you search for STI information*		
Before having unprotected sex	101 (24.4)	115 (62.8)
After having unprotected sex	262 (63.4)	101 (55.2)
After experiencing symptoms	338 (81.8)	91 (49.7)
Before getting tested	297 (72.0)	60 (32.8)
After getting diagnosed	176 (42.6)	50 (27.3)
Why do you search for STI information*		
Want to learn more	348 (84.3)	166 (90.7)
Thought you might have a STI	330 (79.9)	77 (42.0)
Wanted to prevent STI	296 (71.7)	81 (44.3)
Wanted to know how to treat STI	373 (78.2)	62 (33.9)

*clinic sample, n=413 and student sample, n=183 (due to skip pattern)

Overall, 57% of clinic participants reported searching for STI-related information on the internet prior to coming to the current clinic visit: 27.5% reporting they searched that day, 22.4% within the past 3 days, and 10.5% within the past week. In terms of when they searched (multiple responses possible), 75.8% reported searching after experiencing symptoms, 66.6% before getting tested, 58.7% after having unprotected sex, 53.3% after getting diagnosed and 22.7% before having unprotected sex. Reasons for the search included (multiple responses possible): wanting to learn more (78%), thought they had an STI (74%), wanted to know how to treat an STI (72.4%), and how to prevent getting an STI (66.4%). Almost all participants indicated using Google to access STI information (99.5%).

Participants were asked about their reasons for conducting the STI search. Overall there was a high level of endorsement for all reasons given, including searching to learn more about STIs (78%), searching because you might have an STI (74%), searching to know how to treat an STI (72%) and searching to learn how to prevent an STI (66%). Age and sex were not associated with STI-related internet search prior to the clinic visit, though non-Hispanic Blacks were less likely to report using the internet prior to the clinic visit (48%) compared to Non-Hispanic Whites (73%) and Hispanic Latinos (71%) ($p<0.01$; TABLE III). Those with higher education and parent's insurance status were also more likely to conduct an STI-related internet search prior to the clinic visit. Sexual behavior (condom use, number of sex partners, birth sex of sex partner) and reason for clinic visit was not associated with the STI-related internet search prior to the clinic visit, though those reporting a previous STI diagnoses were less likely to search (55%) compared to those without a previous STI diagnosis (67%) ($p=0.01$).

Although the association between prior STI diagnosis and STI-related internet search prior to clinic visit were inversely related (IRR 0.81; 95% CI: 0.69 - 0.95), this association was attenuated and no longer statistically significant after controlling for race, education and insurance status (TABLE V). Due to a priori hypothesis, previous STI was maintained in the multivariable model despite lack of significance at $p<0.20$. In the multivariable model, after controlling for previous STI, education and insurance status, Non-Hispanic Blacks remained less likely than Non-Hispanic Whites to search the internet prior to coming to the clinic (adjusted IRR [aIRR] 0.75; 95% CI: 0.61-0.92). Compared to having parent's insurance, those with government insurance (aIRR 0.80) or no insurance (aIRR 0.77) were both statistically significantly less likely to conduct STI-related internet search. As being on parent's insurance is likely related to age, we ran the model with and without controlling for age. Insurance was a significant predictor regardless of whether age was in the model. As age did not contribute to the overall model fit, it was removed in final analyses.

An interaction between sex and insurance was detected, in models stratified by sex and controlling for previous STI, race/ethnicity and education. Among females, compared to having parents' insurance, those reporting government insurance, private insurance and no insurance were each associated with being less likely to conduct an STI-related internet search prior to the clinic visit, whereas for men insurance status did not affect internet search behavior (TABLE VI).

TABLE V. FACTORS PREDICTING INTERNET SEARCH PRIOR TO CLINIC VISIT IN CLINIC PATIENT POPULATION, N=382

Characteristic	IRR (95% CI)	aIRR (95 % CI)
Previous STI diagnosis	0.81 (0.69, 0.95)	0.97 (0.82, 1.15)
Race/ethnicity		
NH White	Ref	Ref
NH Black	0.65 (0.54, 0.78)	0.75 (0.61, 0.92)
Latino/Hispanic	0.98 (0.82, 1.17)	1.11 (0.94, 1.32)
Education		
Less than HS	Ref	Ref
HS graduate	1.55 (0.91, 2.66)	1.29 (0.76, 2.19)
Some college	1.48 (0.86, 2.53)	1.32 (0.78, 2.24)
College Graduate +	1.99 (1.17, 3.37)	1.66 (0.98, 2.82)
Insurance status		
Parent’s Insurance	Ref	Ref
Government Insurance	0.73 (0.58, 0.91)	0.80 (0.64, 1.01)
Private Insurance	0.90 (0.74, 1.12)	0.84 (0.68, 1.04)
No Insurance	0.75 (0.62, 0.91)	0.77 (0.63, 0.95)

A total of 279 students completed the survey. Over half (51%) were aged 18-19, 45% were aged 20-24 and 3% were 25-29; the median age of participants was 19 years old (TABLE VII). By race/ethnicity, subjects were 23% Non-Hispanic white, 26% Non- Hispanic black, 33% Latino (any race) and 18% Asian. More females participated (55%) than males (45%). Seventeen percent of participants reported a history of a STI diagnosis and 37% reported not using a condom the last time they had sex. In terms of number of sex partners in the past 6 months, the majority reported 1 partner (46%), with 22% reporting 2 partners and 21% reporting 3 or more partners. Internet access was ubiquitous, with only one participant reporting not having access (TABLE IV). The majority of students accessed the internet on handheld devices like a smart phone or tablet (97%).

TABLE VI. INSURANCE STATUS STRATIFIED BY SEX, CLINIC PATIENT POPULATION

	Among Males	**Among Females**	**Breslow-Day**
Insurance status			0.028
Parent's Insurance	Ref		
Government insurance	1.06 (0.78, 1.44)	0.62 (0.45, 0.87)	
Private insurance	0.96 (0.72, 1.29)	0.76 (0.55, 1.04)	
No insurance	0.85 (0.63, 1.13)	0.71 (0.52, 0.97)	

Nearly two-thirds (65%) reported ever using the internet to search for STI-related information. Most student participants reported searching for STI information more than a month ago (37%). In terms of timing of search behavior (multiple responses possible), over half of participants reported searching for STI information before having unprotected sex (58%). Eighty-three percent of participants reported the reason for searching was to learn more about STIs, 39% reported searching because they thought they might have an STI and 40% reported searching to learn how to prevent an STI. All participants (100%) endorsed using Google to search for STI information.

TABLE VII. STUDENT PARTICIPANT CHARACTERISTICS OVERALL AND BY OUTCOME, N=279

		Used the internet to search for STIs*		
	Total N=279	**Yes, n=181**	**No, n=161**	**Chi-square p-value**
	n (%)	n (%)	n (%)	
Age, in years				
18-19	144 (51.6)	94 (65.7)	49 (34.3)	0.38
20-24	185 (45.5)	80 (63.5)	77 (36.5)	
25-29	8 (2.9)	7 (87.5)	1(12.5)	
Race/Ethnicity				
NH White	64 (22.9)	42 (65.6)	22 (34.4)	0.94
NH Black	73 (26.2)	49 (68.1)	23 (31.9)	
Hispanic/Latino, any race	91 (32.6)	58 (64.4)	32 (35.6)	
Asian	51 (18.3)	32 (62.8)	19 (37.3)	
Sex				
Male	126 (45.2)	88 (70.4)	37 (29.6)	0.11
Female	153 (54.8)	93 (61.2)	59 (38.8)	
Sexual Behavior				
MSW	86 (43.2)	64 (74.4)	22 (25.6)	0.41
WSM	81 (40.7)	56 (69.1)	25 (30.9)	
MSM & MSWM	15 (7.5)	13 (86.7)	2 (13.3)	
WSW & WSWM	17 (8.5)	14 (82.4)	3 (17.6)	
Highest level of education				
HS graduate	32 (11.5)	24 (77.4)	7 (22.6)	0.10
Some college	231 (82.8)	144 (62.6)	86 (37.4)	
College/graduate degree	16 (5.7)	13 (81.3)	3 (18.7)	
Insurance Status				
Parents plan	93 (34.4)	67 (72.8)	25 (27.2)	0.20
Government plan	33 (12.2)	21 (63.6)	12 (36.4)	
Private plan	144 (53.3)	88 (61.5)	55 (38.5)	
Previous STI				
Yes	34 (17.0)	31 (91.2)	3 (8.8)	0.01
No	166 (83.0)	117 (70.5)	49 (29.5)	
Sex partners past 6m				
0 partners	21 (10.3)	14 (66.7)	7 (33.3)	0.73
1 partners	94 (46.1)	72 (76.6)	22 (23.4)	
2 partners	46 (22.5)	34 (73.9)	12 (26.1)	
3+ partners	43 (21.1)	30 (69.8)	13 (30.2)	
Condom used last sex				
Yes	126 (63.0)	88 (69.4)	38 (30.2)	0.08
No	74 (37.0)	60 (81.1)	14 (18.9)	

*2 missing responses- does not apply due to skip pattern if no internet or did not use internet for sexual health searches

*MSW= Men who have sex with Women; WSM=women who have sex with men; MSM= men who have sex with men; WSW=women who have sex with women; MSWM= men who have sex with women and men; WSWM= women who have sex with women and men

The majority of students reported that they searched online for STI information because they wanted to learn more (83%). Forty percent of student participants reported searching because they wanted to prevent STIs, with fewer reporting they searched because they thought they had an STI (39%) or because they wanted to know how to treat an STI (31%).

There were no differences detected in the distribution of age, race/ethnicity, or sex and prevalence of using the internet to search for STI information (TABLE VII). Sexual behavior was not associated with internet search prevalence, however more MSM/MSWM (86%) and WSW/WSWM (82%) endorsed searching for STI information compared to MSW (74%) and WSM (69%), this difference was not statistically significant and there were small numbers of MSM and WSW. Those with a previous STI were more likely to have searched for STI information online compared to those with no STI history (91% vs. 70%, p=0.01). Those who did not use condoms during their most recent sex act were more likely to have searched for STI information online compared to those who did use condoms (81% vs. 69%, p=0.08). In the multivariable model for the student sample, after controlling for condom use, participants who had a previous STI were 1.28 times more likely to have ever searched the internet for STI information (aIRR 1.28; 95% CI: 1.10- 1.47) (TABLE VIII). In the multivariable adjusted model, those who used condoms during their most recent sex encounter were less likely to search the internet for STI information (aIRR 0.87; 95% CI: 0.75 - 1.02).

TABLE VIII. FACTORS PREDICTING SEXUALLY TRANSMITTED INFECTION RELATED INTERNET SEARCHES IN STUDENT POPULATION, N=200

Characteristic	IRR (95% CI)	aIRR (95% CI)
Previous STI	1.29 (1.12, 1.49)	1.28 (1.10, 1.47)
Condom use	0.86 (0.73, 1.01)	0.87 (0.75, 1.02)

The proportion of those who ever searched the internet for STI information was greater than expected based on the previous literature: 90% among STD clinic patients and 65% among students (75, 76, 80). In addition our local samples did not differ in Internet access by race/ethnicity; therefore findings on differential search prevalence by race cannot be explained by lack of or decreased access to the internet for particular subgroups.

Non-Hispanic Blacks were significantly less likely to search for STI information prior to coming to the clinic; this has important implications for the representativeness of using Google Trends for surveillance, as Non-Hispanic blacks may be underrepresented. However, it is possible that Non-Hispanic Blacks search post-diagnosis which would increase the representativeness of Google Trends. Further research should be conducted comparing our study results to internet search behavior post-STI diagnosis. Google Trends may be able to refine algorithms to account for under representation by subgroups based on monitoring the contribution of specific queries within a predictive model; this feature would increase the representativeness of the surveillance capabilities. There are differences in search terms that subpopulations use to access STI information online, which can help inform and develop a tailored multi-query search term model to predict disease.

Contrary to previous literature, we found that Non-Hispanic Whites and Latinos had similar rates of searching online for STI information in the STD clinic population. In a cross-sectional self-administered survey of over 3,000 women aged 16-24 attending publicly funded reproductive health clinics from 2008-2010, Non-Hispanic Black and Latino women were more likely to search for STI information compared to Whites (76). Further research is needed to investigate if there are differences in search behavior between race/ethnicities, as this could result in a group being over or under represented in future Google Trends models aimed to predict STI rates.

Previous studies found MSM were more likely to search for STI information compared to women or to men who have sex with women. We did not find this to be true in our clinic sample; in the student sample, although we had low numbers of MSM, they were more likely to search for STI topics compared to WSM and MSW, though this difference was not statistically significant. Men who have sex with men (MSM) may be more aware of risk for disease transmission due to targeting of specific HIV-prevention education, thus they may be more likely to search for STI-related information compared to WSM and MSW. There may have been too little variability in our high-risk population to detect a difference in search behavior by sexual identity/behavior.

The student sample had a lower overall prevalence of searching for STI information, compared to the clinic population. No demographic characteristics predicted searching for STI information; however sexual risk behaviors were related to searching for STI information. This finding demonstrates that even in a lower risk population (represented by students) those most at risk for acquiring or being exposed to STIs are more likely to be generating STI-related search terms. Based on these findings, it is likely that models for predicting STIs based on search terms would capture those at risk or exposed to STIs (vs. other general internet users).

This study was designed to characterize search engine users who generate STI-related terms. There were differences in the distributions of search frequency and reason for search by demographics and risk profile. This helps us better understand how, why and when subpopulations use STI-related search terms. These study results have the potential to be used to inform STI outbreak detection in specific subpopulations based on tailored queries in advanced models using Google Trends.

Study results need to be interpreted with the following limitations in mind. Individuals attending the STD clinic were not selected randomly from the entire STD clinic population, but rather were a convenience sample of individuals attending the clinic on the days/times that study staff was present. The student survey may also have unmeasured selection and sampling bias. Although interviewers were trained to approach students as they were walking by the table, they may have perceptions of who is “easier” to engage and therefore may have been more likely to recruit a specific gender, perceived race or age. Our study did not randomly select from the entire pool of students at the University, but rather relied on a convenience sample of students who were attending the campus-sponsored event. Participation bias may be present if those who are more likely to use internet based search engines were also those more likely to participate (due to greater interest in the topic), though participation among eligible subjects was high (96% and 98%). While we targeted a younger age group due to general STI burden and high likelihood of internet use, we may have excluded those who are most at risk for syphilis (older population) and they are therefore underrepresented in our results. All information was self-reported; if participants differentially reported their exposures, such as behavioral risks, a response bias will be present which could result in misclassification and will influence our measures of association.

The study was cross-sectional and used a convenience sample, therefore, it is not representative of all internet users or all persons with STIs. While a nationally representative survey would be ideal, surveying select exposure groups enabled us to document and analyze similarities and differences in STI-related internet search behavior in two select groups of users. The study targeted the population most affected by STIs who are also most likely to use the internet. Results inform the characteristics of those who are likely to generate STI search terms which will impact use of search engine data for surveillance.

Our study results illustrate how, why, when and if a high risk and a lower risk population conduct STI- related internet searches. This information is important for understanding and improving the predictive capacity of using internet based search engine data for STI surveillance. This information has the potential to be used to inform outbreak detection in specific subpopulations at most risk for acquiring STIs based on established baseline thresholds of queries in an advanced algorithm of Google Trends. This is the first study to document the characteristics of user generated search engine data related to STIs. The survey should be replicated to confirm findings and with a nationally representative sample, targeting those at most risk of acquiring STIs.

6. **STUDY 3b**

The internet is an important source of health information, as it is anonymous, low-to-no cost, and can be accessed at any time. Millions of people search online for health-related information each day, most starting their search via a search engine such as Google (61). Search terms can be downloaded and analyzed to detect patterns in relation to rates of disease to test the hypothesis that increases in specific search terms may be related to increases in rates of disease. Ginsberg et al. developed the Google based application Google Flu Trends which can predict influenza 7-10 days earlier than traditional surveillance systems (65). From this groundbreaking work published in 2009 came many subsequent studies. Following the popular Google Flu trends model, influenza has been successfully monitored via search engine data and outbreaks have been predicted in not only the United States but also in China, France and Spain (69,70). In addition, multiple infectious diseases have been successfully predicted using Google Trends, such as dengue, West Nile Virus, tuberculosis, and more recently HIV (69- 72).

In our preliminary study, we used publicly available Google Trends data to examine the association between sexually transmitted infection (STI) search terms and rates of disease by US state (90). We found a positive and statistically significant correlation between rate of gonorrhea by state and gonorrhea search term (90). Our follow-up study sought to characterize internet users and STI-related search behavior.

We found differences in user characteristics related to STI search prevalence, specifically that demographics were important predictors in a high-risk sample and sexual risk behavior was predictive in a low-risk sample (See study 3a). Here we present the analysis of reported search terms to ascertain whether or not there are similarities or differences in the specific terms individuals use to access STI information. If there are differences in the content of the search terms, by demographics or risk behavior, specificity can be increased in the application of Google Trends for STI surveillance.

Briefly, subjects were recruited from public STD clinics and from a University event. Recruitment was restricted to subjects aged 18-35 who were English speaking. Survey completion took between 5 and 10 minutes and participants did not receive compensation. Data collection included demographic, sexual risk behavior and internet search behavior. The Institutional Review Boards (IRBs) of the University of Illinois at Chicago and the Chicago Department of Public Health approved the study (methods described in detail in Study 3a).

Data was collected via in-person interview and recorded verbatim by trained research assistants and the study PI. Open ended responses were solicited by the question, "Please tell us what words or phrases you used to find the information." All data was entered by the PI and later verified by a research assistant. Data files were imported into Dedoose, an online cloud-based mixed methods analysis program that specializes in allowing multiple users code and display qualitative and quantitative data (106). A strength of Dedoose is its ability to integrate mixed methods analysis, utilizing "descriptors" or quantitative data within the program.

Qualitative coding procedures followed grounded theory approach in which the codes applied reflect the content of the data rather than prior hypotheses (107). The goal of using grounded theory approach was to highlight and explore similarities and differences within the data.

The first rater read all response excerpts and developed a preliminary codebook consisting of 12 codes that emerged from the data. The second rater read all responses and then viewed the preliminary codebook. We used multiple raters and the process of multiple coding to enhance subjectivity of analysis (111). As the dataset was relatively small, the entire dataset was coded by both raters. After each rater reviewed the dataset the raters met to discuss adding, deleting or modifying codes (deleted 2 codes, added 2, modified 1); raters also discussed and documented their understating of each code. For example, the difference between applying the code “transmission” versus applying the code “prevention” was discussed. The process of concordance allows insights to the data which helps refine the coding frame (107, 111).

The Dedoose training module was used to randomly select 20 percent of all excerpts for both samples. Both raters coded these excerpts for training purposes and to establish a baseline reliability score. In the majority of cases the disagreement involved omission of code and was agreed upon once reviewed. Including the passages where omission occurred the kappa coefficient was 0.81, representing a very good level of agreement; when the omissions were not included the kappa increased to 0.92, an excellent level of agreement reflecting very few disagreements in coding.

Both raters then coded the entire dataset for both samples. The dataset was compared for reliability, any divergence (n=38) was discussed and reconciled. Divergence in general resulted from omission of codes, for example omitting a code in an excerpt that already had a code applied. After all divergences were reconciled the data was analyzed to detect themes and patterns by descriptors (e.g., response patterns in males compared to females). After all codes were analyzed, we selected only those with frequency of occurrence of at least 10% within each sample to compare between descriptors. However, we also present a summary of the codes that applied to a lower proportion of the sample to present the depth and breadth of the dataset fully.

Bivariate analysis was conducted using Pearson's chi-square and Fisher's exact test (when cell sizes contained fewer than 5 observations) to detect differences between the presence and absence of codes between samples, as well as the presence or absence of codes within samples by descriptors (e.g., age, sex). In effort to triangulate responses within the samples, we compared the qualitative assigned codes with the quantitative responses to the following statement "I am going to read a list of reasons people use the internet to search for information about STIs or STDs. In general, did you look for information because…" with the following yes/no categories: a) you wanted to learn more about STIs; b) you thought you might have an STI; c) you wanted to know how to prevent getting an STI; d) you wanted to know how to treat an STI; d) you wanted to find a place to get tested for STIs.

Google Correlate was used to identify correlated search terms in the United States with those search terms identified in the analysis. Google Correlate is a publicly available data tool that is a part of the Google Trends package. It enables users to enter search terms and to find queries with high correlations of the entered terms. Google Correlate is used to build multiple search query models in effort to refine and enhance the specificity of predictive models using Google Trends (112).

Overall, 446 subjects were recruited from the public STD clinics and 279 students were recruited. STD clinic patients were 57 male, median age 24, 30% non-Hispanic White, 51%, non-Hispanic Black, and 19% Latino/Hispanic of any race. Students were 54% female, median age of 19, 32% Hispanic/Latino, 26 % Non-Hispanic Black, 22% Non-Hispanic White, and 18% Asian.

Twelve codes emerged from the dataset in roughly two major content areas: seeking information about STIs using general terms (including prevention, testing and treatment) and accessing information based on symptoms (TABLE IX). One code “sex education” was only applied to the student excerpts. The term “education” occurred in 11% of the student excerpts and none of the clinic excerpts.

The clinic sample length of query ranged from 4 to 230 characters, with a median of 50 characters; the student sample ranged from 4 to 116 characters with a median of 41. The average English language query on Google is estimated at 20 characters or approximately 4-5 words (81, 113), thus our participants reported longer than average search queries. There were no demographic differences detected in the student sample in regard to length of query, however, in the clinic sample all respondents who reported queries over 150 characters were also Non-Hispanic Black.

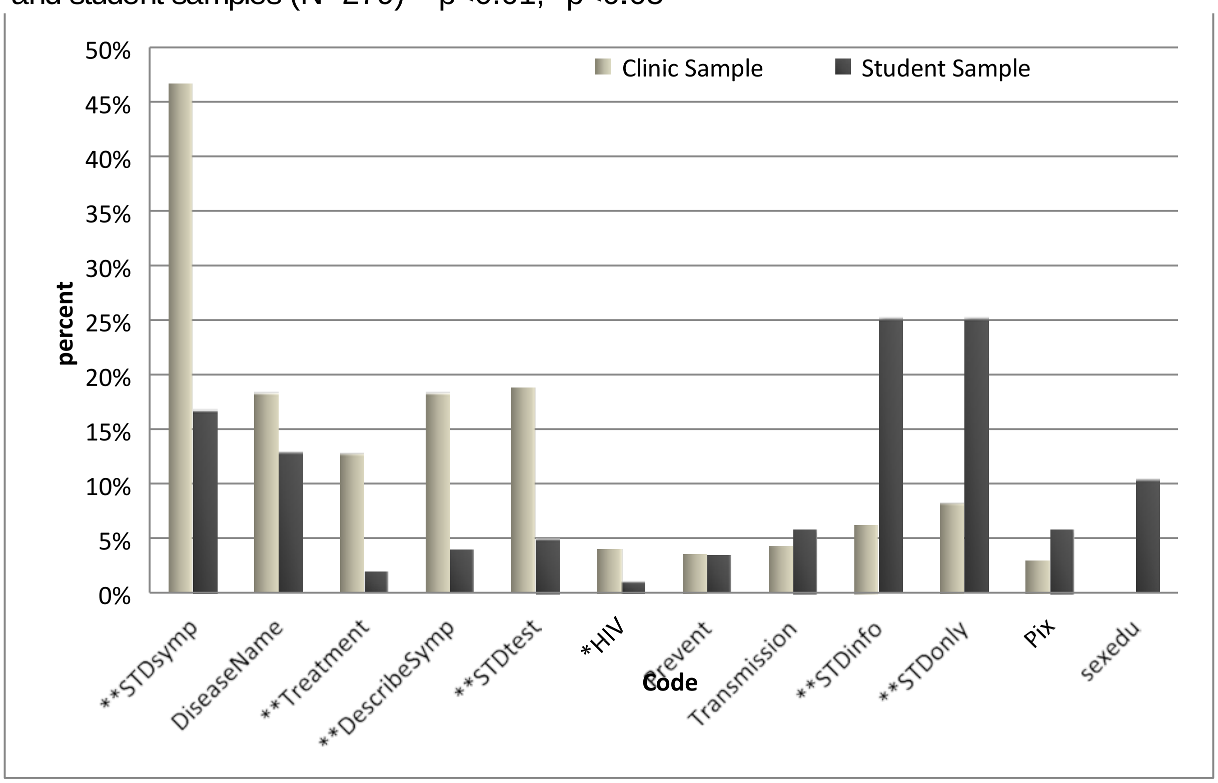
TABLE IX. SEARCH TERM CODES AND QUOTES FROM CLINIC AND STUDENT SAMPLE

Code	Description	STD clinic patient quote	Student quote
Describesymp	Description of STD symptoms	"abdominal pain, fishy odor, change in discharge"	"Itching down there"
DiseaseName	Use of STD by name	"Syphilis"	"Gonorrhea"
HIV	Mention of HIV or AIDS	"HIV and STD "HIV test clinic"	"HIV and STD"
Pix	Google images, searching for pictures of STDs	"How does STD look?"	"Google Images STD"
Prevent	Preventing STDs	"How safe are condoms"	"I want to know about STD before I have sex I look online"
STDinfo	General STD information	"What is STI mean?"	"Health information for STD"
STDonly	Using the word STD, STI or phrase Sexually Transmitted Infection no other search terms	"STD"	"Sexually transmitted disease"
STDsymp	Using the phrase STD symptoms without an actual description of symptoms	"STD symptoms"	"Symptoms of STD"
STDtest	Looking for STD testing sites or information about testing procedures	"What can I do to treat chlamydia?"	"Treatment options for STDs"
Treatment	Treatment or medication information	"What can I do to treat chlamydia?"	"Treatment options for STD"
Transmission	How are STDs transmitted	"How do you spread STD?"	"STD risk how do I get one?"
SexEdu	Using the term education within the search	N/A	"Sex education online"

*"SexEdu" code only applied to student sample

Nearly half (47%) of the clinic sample reported using the search term “STD Symptoms” compared to 17% of the student sample ($p<0.01$; Figure 10). In addition, clinic participants were significantly more likely than student participants to report describing STI-related symptoms, or searching using words related to treatment and testing. Student participants were significantly more likely than the clinic sample to report general terms, using “STD” as their only search term (26% vs. 8%; $p<0.01$) or searching for general STI information (26% vs 6%, $p<0.01$). The clinic sample was more likely to report seeking information based on symptoms, reflected in codes for describing symptoms and STD symptoms, compared to the student sample, which was more likely to search for general information.

Figure 10. Frequency of qualitative code occurrence applied to search content in clinic (N=446) and student samples (N=279) **$p<0.01$; *$p<0.05$



The top 5 codes that exceeded the 10% threshold are presented in Table X for the clinic participants and Table XI for the student participants. The top codes for the clinic participants were codes that encompassed describing STI symptoms, using STI disease names, using the phrase “STD symptoms”, searching for STD testing information and/or searching for STD treatment. The top codes for the student participants were codes that encompassed searching for STD information, using the term “STD” as the only search term, using the phrase “STD symptoms”, searching for sexual education and/or searching using STI disease names. Two of the content areas using “STD symptoms” and searching using STI disease names were found in both the clinic and student samples among the top reported search queries.

In the clinic sample, females were more likely to report describing specific STI related symptoms than were males (p=0.04; TABLE X). In addition, Non-Hispanic Blacks (22%) were more likely than both Non-Hispanic Whites (11%) and Hispanic/Latinos of any race (12%) to report searching for STI information by describing symptoms (p=0.01). Those who were Hispanic/Latino were more likely to report using the search term STD symptoms compared to other races (p=0.04). Those who were 25-29 years old were more likely to use the STD disease name in their search compared to any other age category (p=0.01). Finally, those who used condoms were more likely than those who did not use condoms to report searching for STD test information (22% vs.14%, p=0.02).

In the student sample, there were no differences in occurrence of the codes by race or sex (TABLE XI). Compared to other age groups, those who were older (25-29 years old) were more likely to report they searched using the term “STD Symptoms”

TABLE X. TOP QUALITATIVE CODES BY CLINIC SAMPLE CHARATERISTICS

		Describesymp	Disease Name	STD symp	STD Test	STD treatment
	N	n (%)	n (%)	n (%)	n (%)	n (%)
Sex						
Male	255	36 (14)*	44 (17)	108 (42)	46 (18)	28 (11)
Female	190	41 (22)	33 (17)	85 (45)	32 (17)	26 (14)
Race						
NH White	127	14 (11)ƚ	14 (11)*	55 (43)*	29 (23)	13 (10)
NH Black	221	49 (22)	49 (22)	91 (41)	21 (10)	27 (12)
Hispanic/Latino, any	82	10 (12)	12 (15)	40 (49)	24 (29)	12 (15)
Age						
18-19 years old	52	7 (13)	6 (11) ƚ	26 (50)	11 (21)	6 (12)
20-24 years old	185	40 (22)	33 (18)	81 (44)	27 (15)	24 (13)
25-29 years old	138	22 (16)	30 (22)	59 (43)	30 (22)	11 (8)
30-35 years old	71	8 (11)	8 (11)	27 (38)	10 (14)	13 (18)
Sexual Behavior						
MSW	192	29 (15)	34 (18)	76 (40)	34 (18)	18 (9)
WSM	138	31 (22)	23 (17)	60 (43)	23 (17)	20 (14)
MSM	56	6 (11)	10 (18)	29 (52)	9	10 (18)
WSW	49	10 (20)	10 (20)	24 (49)	8	6 (12)
Previous STD						
Yes	222	42 (19)	46 (21)	99 (45)*	32 (14)*	29 (13)
No	166	35 (16)	31 (14)	91 (42)	42 (20)	25 (12)
Number of sex						
1 partners	117	27 (23)	22 (19)	53 (45)	20 (17)	21 (18)
2-3 partners	177	30 (17)	30 (17)	74 (42)	34 (19)	18 (10)
4 or more partners	142	20 (14)	25 (18)	64 (45)	22 (15)	15 (11)
Condom use						
Used condom	165	25 (15)	33 (20)	71 (43)	37 (22)*	20 (12)
Did not use condom	273	52 (19)	44 (16)	119 (44)	38 (14)	34 (12)

* p<0.05; ƚ p<0.01

MSW= Men who have sex with Women; WSM=women who have sex with men; MSM= men who have sex with men; WSW=women who have sex with women; MSWM= men who have sex with women and men; WSWM= women who have sex with women and men

TABLE XI. TOP QUALITATIVE CODES BY STUDENT SAMPLE CHARACTERISTICS

		STDinfo	**STDonly**	**STDsymp**	**SexEdu**	**diseasenam**
	N	n (%)	n (%)	n (%)	n (%)	n (%)
Sex						
Male	126	26 (21)	22 (17)	20 (16)	10 (8)	9 (7)
Female	153	23 (15)	29 (19)	14 (9)	16 (10)	16 (10)
Race						
NH White	64	12 (19)	8 (13)	8 (13)	4 (6)	6 (9)
NH Black	73	13 (18)	14 (19)	9 (12)	8 (11)	9 (12)
Hispanic/Latino, any race	91	15 (16)	18 (20)	13 (14)	6 (7)	6 (7)
Asian	51	10 (20)	11 (22)	4 (8)	3 (6)	5 (10)
Age						
18-19 years old	144	26 (18)	29 (20)*	18 (13)*	11 (8)	12 (8)
20-24 years old	185	24 (13)	20 (11)	13 (7)	10 (5)	14 (8)
25-29 years old	8	1 (12)	2 (25)	3 (37)	0	0
Sexual Behavior						
MSW	86	19 (22)	16 (19)	17 (20)*	5 (6)	4 (5)
WSM	81	13 (16)	17 (21)	7 (9)	9 (11)	9 (11)
MSM	15	5 (33)	4 (27)	1 (7)	2 (13)	3 (20)
WSW	17	6 (35)	4 (24)	0	2 (12)	1 (6)
Previous STD						
Yes	34	6 (18)	4 (12)	3 (9)	5 (15)	6 (18)*
No	166	37 (22)	37 (22)	22 (13)	13 (8)	12 (7)
Number of sex partners						
0 partners	21	5 (24)*	4 (19)	3 (14)	0	2 (9)
1 partner	94	19 (20)	24 (26)	7 (7)	12 (13)	8 (9)
2 partners	46	3 (6)	12 (26)	10 (22)	3 (7)	2 (4)
3 or more partners	43	2 (5)	9 (21)	4 (9)	4(9)	7 (16)
Condom use						
Used condoms	126	27 (21)	29 (23)	16 (13)	10 (56)	10 (8)
Did not use condoms	74	18 (24)	13 (18)	9 (12)	8 (44)	7 (9)

*p<0.05

MSW= Men who have sex with Women; WSM=women who have sex with men; MSM= men who have sex with men; WSW=women who have sex with women; MSWM= men who have sex with women and men; WSWM= women who have sex with women and men (p=0.01) or that they only used the term “STD” (p=0.04). In terms of sexual behavior, men who have sex with women (MSW) were the most likely to report searching for STI information by typing “STD symptoms” (p=0.04), there were no other differences in code occurrence by sexual behavior. Those who reported having a previous STI were more likely to search using a disease name (p=0.05). Finally, those with zero sex partners in the past 6 months, were more likely to report searching for general STD information (p=0.02).

The presence or absence of all qualitative codes was compared with participant responses for the reasons they searched for the STD information. Nine of the 11 codes applied to the clinic sample had no statistically significant relationship with the quantitative question and 9 of the 12 codes applied to the student sample had no statistically significant relationship. In the clinic sample, the code encompassing specific sexually transmitted disease name was associated with endorsing searching to learn more (p=0.01). The codes reflecting searches related to transmission and prevention of STIs were associated with endorsing searching to learn how to treat an STI and how to prevent an STI (p=0.01 and p= 0.03, respectively). In the student sample, the code for description of specific STI-related symptoms was related to endorsing searching because you think you have a STD and for treatment information (Fisher’s exact p=0.04; p=0.02). The code for searches for STD test information was associated with endorsing searching to find a place to get tested (p=0.02) and the code for searching for STD treatment was associated with endorsing searching to learn how to treat an STD (p=0.01).

The direct term “STD symptoms” was reported by 47% of the clinic sample and 17% of the student sample. When entered into Google Correlate, the term “STD symptoms” generated many STI related terms with high levels of correlation (0.87-0.94). The top 20 terms that were associated with “STD symptoms” are displayed in TABLE XII. Five of the twenty terms are not directly related to STIs (“how to talk to women”, “estrogen pills”, “pregnant symptoms” “talk to women” and “first trimester symptoms”). As disease name was often stated as a term used in searching for STD information (19% clinic; 13% student), “Chlamydia” was entered into Google Correlate. Sixteen of the top 20 search terms that were correlated with “STD symptoms” were also highly correlated with “Chlamydia.” Two of the terms that were generated by “STD symptoms” and not directly related to STIs (“how to talk to women”, “first trimester symptoms”) were not correlated with the search term “Chlamydia.” Search terms related to STD test, discharge (“thick discharge” “white discharge” “creamy discharge”), “STD” or STD treatment had few terms correlated that were related to STI information. For example, the term “STD” generated “gonorrhea” as the third highest correlated term at 0.924, however all of the other top 20 terms were unassociated with STIs. In addition, participant-generated search terms related to "sex education" and using the phrase "sexually transmitted infection" did not result in any Google correlated STI related terms.

TABLE XII. GOOGLE CORRELATE RESULTS FOR THE UNITED STATES: TOP REPORTED SEARCH TERMS GENERATED BY CLINIC AND STUDENT SAMPLES

	STD symptoms	Chlamydia	STD test	Thick discharge	STD treatment	STD
Chlamydia	0.943	--	--	--	--	--
STD symptoms in men	0.935	0.919	--	--	--	--
Chlamydia treatment	0.906	0.918	--	--	0.815	--
Chlamydia symptoms	0.903	0.900	--	--	--	--
Chlamydia in men	0.902	0.913	--	--	--	--
How to talk to women	0.900	--	--	--	--	--
Signs of STD	0.898	0.929	--	0.944	--	--
STD symptoms in women	0.893	--	--	--	--	--
Gonorrhea symptoms	0.889	0.984	0.854	--	0.807	0.924
Estrogen pills	0.888	0.922	--	0.931	--	--
Pregnant symptoms	0.882	0.925	--	0.965	--	--
Talk to women	0.881	0.902	--	--	--	--
Cure chlamydia	0.879	0.932	--	--	--	--
Treat chlamydia	0.878	0.938	--	--	--	--
Syphilis symptoms	0.878	--	--	--	--	--
Discharge	0.877	0.940	--	--	--	--
Thick discharge	0.875	0.917	--	--	--	--
Gonorrhea	0.872	0.941	--	--	--	--
First trimester symptoms	0.871	--	--	--	--	--
Milky white discharge	0.871	0.905	--	0.955	--	--

Terms in italics not directly related to STIs
Table reports Pearson correlation coefficient.
*top codes- sex education, sexual health education, sexually transmitted infection, and STI did not produce any STI-related search terms

Using grounded theory approach, two main themes emerged: searching to find general STD information and searching based on symptoms. The themes emerged from the dataset, yet are intuitive based on our two samples. The assumption that a lower risk sample would search for more general information held true in our results, as did a higher risk sample searching based on symptoms of disease. There was an overlap in the highest frequency codes across samples of searching by disease name and using the phrase “STD symptoms”. In the student sample, risk behavior was associated with being more likely to have these codes applied, whereas in the clinic sample there were no associations found by demographic or risk behavior characteristics. This highlights that there is an overlap in the populations, although students as a whole are lower risk, at the individual level there is variation in exposure to STIs as well as search queries.

We also noted that the query length median of our sample was double the average length of a Google query. This discrepancy may be due to data collection methods; participants were asked to report what they type to find STD information, they were not directly observed. Google has an autocomplete feature which suggests the end of a search based on the first letters and words typed in. Thus, participants may have used more words to describe their search query then they would have if they were actually typing into Google. Additionally, we did not assess search process and patterns or type of device used to access information. All of these factors are not only important in understanding how users access STI-information online, but also in refining search terms used for predicting trends in disease.

Our efforts to triangulate the data were inconclusive, likely due to data collection methods and the multiple response format of the quantitative question. Participants were able to endorse multiple reasons why they searched for STI information and they were asked for their search terms generally. However, when associations were detected, they were between codes and responses that were congruent (i.e., code for STD test correlated with search reason “search for STD test information”). Future studies should consider a multi-stage approach to data collection in which qualitative responses are recorded and coded to inform the quantitative questions, which should then validate the coding scheme and content.

Google Correlate results produced 16 STI-related and strongly correlated search terms for two of the top reported terms (STD symptoms and disease name). The two terms were reported by both samples and validated using Google Correlate as a national sample. These terms could be used to build a multi-query predictive model for STI rates, by US state. Lampos et al. discovered a multi-query model enhanced the performance of the predictive capabilities of the Google Flu Trends model (109). Accounting for variation within clusters of terms, weighting specific terms (to control for their overall contribution to the model), as well as supervising the machine learning (e.g., not completely automating the procedure) significantly improves inference (109). The combination of search query data and disease data leads to better trained models for prediction (109).

Study results should be interpreted in light of some limitations. All information was self-reported, including sexual risk behavior and previous STI diagnosis. Surveys were confidential and anonymous, in effort to increase the validity of self-reported data.

The study was cross-sectional and used convenience sampling, therefore, it is not representative of all internet users or all persons with STIs. The study did target the population most affected by STIs who are also most likely to use the internet (individuals ages 18-35). The primary weakness of previous studies analyzing the relationship between search trend data and disease rates is the absence of measuring and describing the content of disease-specific internet searches. Our study addresses this gap and characterizes internet users as well as the content of their STI-related searches. The mixed method approach allowed us to examine the content of queries from various angles leading to a more comprehensive understanding of search behavior. In addition, the combination of a high risk and lower risk sample allowed us to generate a rich heterogeneous qualitative data set (111).

This study described and analyzed search term query contents by content and sample characteristic. As models for disease monitoring are developed based on search term volume, it is important to understand if there are differences or similarities in search behavior and content in subpopulations that may be at risk for disease acquisition. Results from this study will help inform using Google Trends for STI surveillance by building queries specific to those most at risk of disease. Future studies should incorporate direct observation of search behavior and examine the influence of search patterns and Google's autosuggest function, as well as testing predictive models using multiple correlated search queries.

7. **CONCLUSION**

7.1 **Public health implications**

We conducted a multi-faceted evaluation of the utility of internet search terms to STI surveillance via Google Trends. Each approach had a different perspective- testing the overall utility of Google Trends applied to STIs via a comparative analysis, assessing the feasibility and acceptability of integrating Google Trends into current surveillance from a worker perspective, and characterizing internet users in terms of frequency and reason for STI-related search as well as describing and analyzing the content of searches. Traditional surveillance of STIs relies on mandatory case reports and suffers from many issues including incomplete data and delays in reporting. Using search trend data may provide an additional tool to enhance surveillance of STIs to monitor and predict outbreaks and trends in disease, which may lead to advances in STI prevention and control.

In study 1, we assessed the potential utility of Google Trends as a surveillance method by conducting a comparative analysis of STI-related search terms and rate of STIs by U.S. state for two different time periods (2011 and 2013). Statistically significant positive associations were detected at both the state and the city level, indicating that increases in search volume are associated with increases in STI rates. These findings are the first step in determining the utility of using search engine trends for STI surveillance and demonstrate the potential for using Google Trends as an additional surveillance tool.

In study 2, we examined how internet search trend data may be integrated into the existing STI surveillance system by conducting an online survey with STI surveillance employees.

Our results provide insight into barriers and facilitators to integration with the current system. With a high level of reported interest and the reported preferred learning method of webinar, the scale up and adoption of Google Trends for STIs is achievable and may be implemented cost-effectively.

In study 3, we determined search engine user characteristics by surveying specific exposure groups of individuals aged 18-35, a population with highest rates of internet use and generally highest risk for STIs, to assess internet search behaviors related to STIs and to determine the reason (information seeking, treatment, prevention), timing (prior to or after exposure, onset of symptoms, etc.) and content of STI-related internet searches. We validated findings of the content reported by our sample by searching for correlated terms via Google Correlate with data from the United States. The application of Google Trends to disease monitoring is dependent on search term volume, therefore it is important to understand if there are differences or similarities in search behavior and search content in subpopulations that may be at risk for disease acquisition. These differences may allow for the detection of outbreaks in specific subgroups, as well as ensure overall representativeness of a multiple query model.

7.2 **Methodological issues**

In study 1, we used publicly available annual STD rates by state, published by the CDC. This data includes reported cases, but does not include cases that may have been empirically treated or undiagnosed cases, therefore the rates are an underestimate of the true disease burden. While our search terms were significantly and positively associated with rate of disease by state in two separate time periods, further analysis is needed to determine the predictive capabilities of the search terms. Future studies should seek partnerships with local health departments to access updated surveillance data by week or month, opposed to annual rates, in order to refine and specify prediction models based on search terms.

In study 2, we conducted an online survey of STI surveillance workers in the United States. Advantages to this strategy include access to broadly geographically distributed populations who may be difficult to otherwise reach, time savings for researchers and participants, direct data entry, ease of administration and data analysis, and cost-savings (114). Disadvantages include sampling concerns (selection bias, potential lack of sampling frame), lack of generalizability, and lower response rates compared to in-person surveys (114). We used an online survey to try to capture a national sample as cost-effectively as possible. Our survey was completed by approximately 25% of those who received the survey link. Methods to improve response rate include personalized invitations, follow-up reminder prompts, and phone call follow-up with option to complete the survey in-person. Prior to advocating for integration of Google Trends into the current surveillance system, it is important to get a more representative view on barriers and facilitators to technology integration, as well as investigate how Google Trends can complement existing surveillance systems.

In study 3, the majority (85%) of student surveys were conducted at a one-day event over a time span of 6 hours, with 15% of surveys conducted with students accessing services at the university Wellness Center. Wellness Center and event participants were not significantly different in terms of exposure (sexual risk behavior) or outcome. While participation rates were high, we have no information about non-participants and therefore cannot assess if they differ significantly from participants on exposure or outcome status. Findings from our survey will not generalize to all students.

Both the STD clinic survey and the student survey were conducted via interviewer-administration in a paper-pencil format. However, the survey was anonymous and all study staff were trained to collect data in a non-judgmental manner (practicing discussing sexual partners and STD diagnoses). The interviewer-administration format allowed us to standardize data collection (interpretation of question, recording accuracy) but may have reduced overall perception of privacy and increased underreporting of behaviors thought to be socially undesirable. Alternatives to this method would have been to employ CASI; this would also allow an embedded search task to monitor and record exact search behavior.

In order to facilitate reproducible and consistent results, we included a detailed account of our methods for study 1, including the search terms that were used with Google Trends as well as the years of the publicly available data (which allows replication); however the internet is a rapidly changing environment and as users of a public interface provided by a private company (Google) we have no control over their updates. Since the onset of this study (approximately 2 years) there have been numerous updates and upgrades to Google Trends; this did not impact our results when we updated our analyses with most currently available surveillance data.- However there is potential that the application could refine its capabilities and alter future results despite duplication of methods.

In study 3, qualitative coding was conducted by two independent raters. Kappa coefficients were high, indicating high levels of inter-rater agreement of the applied qualitative codes. While the codes emerged from grounded theory approach, excerpts were relatively short and did not take extensive interpretation to code.

Multiple staff conducted STD clinic and student user characteristic surveys. All staff were trained and certified to conduct research with human subjects, and participated in mock sessions in which they practiced recruitment, eligibility screening, informed consent and data collection under direction and observation of the study PI. This was intended to ensure data quality and consistent data collection across interviewers. All staff was also periodically observed by the study PI. The PI collected data at least one shift a week to ensure study procedures were acceptable to the population and staff of the clinics. The study PI and staff conducted surveys at one of the three STD clinic locations. In analyzing the study outcome by interviewer, it was detected that two interviewers had significantly lower rates of outcome compared to other interviewers. However, closer inspection identified this was a result of the study site: across all study sites Non-Hispanic Blacks had a lower overall prevalence of the outcome and the two interviewers were based at a site that enrolled exclusively Non-Hispanic Black patients. As there were few refusals, comparison of refusal rate by interviewer did not provide any additional information as to whether or not bias was introduced by study staff. Six of the seven trained staff participated in both the STD clinic and student surveys, one staff member only collected data at the STD clinics. There were no differences in refusal rates or outcome by interviewer in the student sample.

We conducted an interim analysis of the STD clinic sample and due to both the outcome and exposures being higher than anticipated (based on literature review), we increased the overall sample size to ensure we could detect a statistically significant difference in exposures hypothesized to be related to outcome. This was a part of the data quality control and assurance conducted, and the amendment to increase sample size was approved by the UIC and CDPH IRBs.

As studies 2 and 3 used convenience samples, we cannot generalize findings to the broader population. For example, our findings from study 2 do not generalize to all surveillance workers. In aim 3, we sampled two different exposure groups in order to compare findings and present a broader scope of information on the subject. While a nationally representative survey would be ideal, surveying select exposure groups enabled targeted characterization of users. We targeted the population most affected by STIs who are also most likely to use the internet (STD health clinic clients aged 18-35), as well as a lower risk comparison population (students 18-35). While we targeted a younger age group due to general STI burden and high likelihood of internet use, we may have excluded those who are most at risk for syphilis and they will be underrepresented in our results.

Participants for studies 2 and 3 were not selected at random, which introduces the possibility of selection bias due to study design. In addition to the bias from our sampling design, unmeasured non-response bias and self-selection bias may be present in our results (introduced by participants). Those who participated in our surveys may be inherently different than those who did not participate. Of those who were eligible we do not have further information on those who chose not to participate, although participation rates in the clinic and student survey were high overall (96% and 98%).

For study 2, the overall participation rate of 25% was low, but within the average range of an online survey. Average response rates of online surveys are estimated to range from 20-34% (115). There are several strategies to improve online response rates, including multiple follow-ups with non-responders, utilizing short surveys, and providing incentives. We were able to send the link to the survey four separate times, though twice the link went to the identical population and we had no control over follow-up with non-responders. In addition, the multiple times the survey was sent was not a reminder or a follow-up specifically, but a general probe for participation. While the survey was short (~25 questions) there were no incentives offered for completion. One of the participating organizations offered to survey their membership in person at a national meeting, which may lead to increased response. However, the recommended follow up is a random sample of surveillance workers within jurisdictions that have the highest rates of chlamydia, gonorrhea and syphilis in 2013 (most currently available national surveillance data), serving populations of at least 100,000 residents.

While there are a number of organizations that conduct prevention and control activities, health departments conduct the majority of the work and are the basis of the national surveillance system. The NACCHO 2013 survey found that 64% of 1,971 local health departments reported providing testing for STIs. Additionally, the likelihood that health departments provide screening and treatment of STIs increases with increasing population size (91). For example, 92% of health departments that served a population greater than 500,000 persons reported providing screening for STIs, compared to 50% for those serving populations less than 50,000 persons (92). For jurisdictions of 100,000- 499,999 persons, 87% provided screening for STIs (91). Taken together, this information supports the approach of selecting the cities with the highest rates, sufficient population density, and health departments that may benefit most from Google Trends as an additional STI surveillance method.

Participants for study 2 were recruited via the email they signed up with to receive organizational newsletters, presumably their work email address at the health department. The link to the survey was not in a direct email to the participant, but rather embedded in a newsletter. Employees may have decided to participate based on whether they perceive their work environment to be supportive of and/or interested in integrating new technology. Finally, results may be biased as those who choose to participate may be different from those who do not. For example, employees may be more likely to participate if they perceive a benefit to integrating new technology into the current surveillance system compared to those who do not think this is an endeavor worth pursuing. If this were true our results would be biased in favor of implementing the new surveillance tool.

7.3 **Policy implications**

While STI case reporting is mandatory, surveillance needs to also detect and monitor increases in disease, particularly outbreaks. Using Google Trends has the potential to heighten STI surveillance capabilities to inform real-time detection of outbreaks and trends earlier than local or national systems with the benefit of including potential non-cases (those who are not yet diagnosed, or might be infected but not tested). If Google Trends is integrated into the current surveillance system, there may be increased opportunities to respond to trends in disease. The current surveillance system relies on confirmed cases; data from Google Trends could add information about non-diagnosed individuals or non-reported cases. The combination of increased ability to detect outbreaks and better understanding of the disease may increase outbreak response and control of disease. As our survey highlights, many STI surveillance workers report their departments did not have STI outbreak detection methods in place, and even more did not have a response plan in place. Thus, public policy needs to support - and public health workers need to advocate for low-cost, rapid, flexible and innovative strategies to respond to outbreaks of disease.

7.4 **Future research**

The next steps in exploring the utility of Google Trends applied to STI surveillance are: analytic models using search terms and STI rate and replicating the surveillance worker survey.

It is essential to perform further analytic techniques exploring the predictive capacity for trends in STIs using the search terms we identified in our user survey. Access to the Google Trends application program interface (API) allows health researchers access to data that is not publicly available. While internet user information remains unavailable from Google, access to the Trends API allows the query and download of search trends which do not exceed the public threshold. Partnerships with local health departments should be established to access STI data on a weekly basis in order to conduct a time series analysis with search trend data. Access to this level of STI data would allow the application of newer methods of now casting and supervised machine learning models to our assessment of utility. Engaging an investigator trained in bioinformatics or computer science with expertise in these methods will be essential in furthering our work, specifically to train a multi-query model, such as elastic net used in influenza prediction (109). Elastic net incorporates multiple queries and allows for adjustment in the contribution of each term, so that a spike in volume in one specific term does not inflate the prediction of disease (109). Finally, partnerships with alternate search engines should be established. While Google maintains the majority of search engine traffic (64% in 2015), Bing! from Microsoft accounts for 21% and Yahoo! for 13% of US searches (81). It is important to try to triangulate search volume patterns from the most popular sources.

As our survey on STI surveillance workers had a low response rate, the survey should be repeated with funding for incentives as well as direct follow-up with non-responders. Results of the study would inform the scale up and adoption of Google Trends for STI surveillance. In addition, our survey did not assess the ways in which surveillance workers could integrate the new tool into their current workflow. Thus, future data collection efforts should assess the integration of Google Trends into current surveillance systems.

Google Trends is an innovative tool that has the potential to monitor disease burden, detect trends and predict change, adding a flexible, adaptable and real-time element to traditional surveillance of nationally notifiable STIs. Results from this study demonstrate that: Google trends is statistically significantly and positively associated with reported STI rates at a city and state-level, and this was confirmed in two independent time periods; those at risk for and with STIs commonly conduct STI-related internet searches, primarily using Google; the local sample used multiple queries that are highly correlated with specific STI search terms on a national level; from a worker perspective, there is an interest and potentially a need for the application of internet-based search engine query to augment STI surveillance. Future research is needed to determine the predictive capacity and most practical application of Google Trends for reportable STI surveillance.

APPENDICES

APPENDIX A

Wolters Kluwer Health, publisher of the Journal of Sexually Transmitted Infections, allows for the reproduction in full of previously published manuscripts in the author's thesis or dissertation. There is no copy right release needed. The policy can be found in full at :

http://journals.lww.com/stdjournal/_layouts/15/1033/oaks.journals/rightsandpermissions.aspx

APPENDIX A (continued)

NOTE

A Comparison of Internet Search Trends and Sexually Transmitted Infection Rates Using Google Trends

Amy K. Johnson, MSW and Supriya D. Mehta, MHS, PhD

Abstract: Google Trends was used to determine the relationship between sexually transmitted infection (STI)–related search engine trends and STI rates. Trends seem to be similar to the relative rates of STIs and to regional differences in rates. Search engine trends are an innovative tool to integrate into STI surveillance.

In 2009, researchers at Google partnered with the Centers for Disease Control and Prevention to establish the first use of Internet search engine data to predict trends in infectious disease.[1] By using highly correlated search engine terms for flu-like symptoms, Ginsberg et al.[1] were able to predict flu outbreaks 1 to 2 weeks ahead of the Centers for Disease Control and Prevention's US Influenza Sentinel Provider Surveillance Network. The results of this project made international headlines; subsequently, Google has made the program, "Google Trends," available for public use. To date, Google Trends has successfully predicted trends in West Nile virus, norovirus, varicella, influenza, and HIV.[1–4]

In the United States, Internet use is largely ubiquitous with 78% of all adults and 95% of teenagers endorsing regular Internet use.[5] In addition, search engine use tops the list of most popular online activities, with 92% of adults reporting use of Internet search engines like Google.[6] In 2013, results from the Internet and American Life project showed that 72% of American Internet users have gone online specifically to seek health information; 77% of those users state that they started their research with a search engine such as Google.[6] The Internet is a portal for users who are seeking health information that is easily and anonymously accessible. There is a potential for search engine data to provide an additional venue for surveillance efforts, allowing for earlier detection of trends in disease and increased ability to monitor overall impact and examine the geographic spread of disease. Most recently, in 2013, Jena et al.[2] used search engine data to model HIV incidence by US state. As surveillance for notifiable sexually transmitted infections (STIs) evolves, the Internet may play a key role. This exploratory study sought to determine whether or not STI-related searches on Google Trends would be related to trends in rates of nationally notifiable STIs.

Google Trends analyzes Internet searches to tally how many searches are completed for the terms entered, relative to the total number of searches on Google over a selected period. Google limits redundant searches from unique devices, as identified by Internet protocol address. In addition, search terms need to exceed a threshold of search frequency to appear in the results.[7] Search data are normalized by an automated procedure of dividing sets of the data by a common variable, thus allowing comparisons to be made without regard to total search volume. For example, if 2 regions show the same percentage for a search term, it does not mean that their absolute volumes are the same; rather, the data can be compared equally because it has been normalized by the total volume from each region.[7] Search volume is presented on a scale of 0 to 100 to reflect interest over time for the relative period of interest selected.[7] The values for the same search term may vary by the selected period; to compare these values, they must be scaled on the same period.[7]

Search terms were selected based on an initial probe of the terms gonorrhea, syphilis, and chlamydia to determine the most effective choice based on correlations above 80%. Google Trends generates correlated search terms for any term entered. The lower bound of a strong correlation was chosen as a cutoff point to examine any extraneous terms that might be correlated with our selected terms. The following terms were selected for our final analysis: "Gonorrhea symptoms, chlamydia symptoms, syphilis symptoms." After selecting the search terms, Google Trends was used to graph the popularity of search terms over time from 2005 through 2011 and to map by US state. Centers for Disease Control and Prevention STI surveillance data from 2011, the most recently available year of surveillance data, were compared to the graph for each term.[8] Using the export data function, we downloaded response volume of search data as a continuous variable by state in comma separated value format from Google Trends. Bivariate correlations were used to assess the strength of the relationship between frequency of search term and rate of disease by US state for rates in 2011. Google Trends and STI data were analyzed in Stata/SE 13 for Windows.[9]

Each search term was highly correlated (>0.80) with many other like terms. For example, "gonorrhea symptoms" was correlated with: gonorrhea in men, gonorrhea STD, facts gonorrhea, gonorrhea women, and gonorrhea treatment. The term "chlamydia symptoms" was the most commonly searched of the 3 STI terms across all years (Fig. 1). Search trends showed relative increases in search terms overall. The biggest increases in chlamydia search terms occurred early in 2009 and again toward the end of 2010, whereas increases over time for gonorrhea and syphilis search terms were more gradual and closely paralleled each other.

The frequency of the search terms relative to all other searches was greatest in states where STI rates are highest, for example, in the southwestern states (Fig. 2A–C). Pearson correlation coefficient was used to explore the association between STI rates and Google search term frequency by state for 2011. Owing to missing data at the state level for search terms, syphilis trends could not be tested. The gonorrhea search term was positively and statistically significantly associated with gonorrhea rates in 2011 ($r = 0.72$, $P < 0.001$); however, the chlamydia search term did not show an association with chlamydia rates in 2011 ($r = 0.06$, $P = 0.65$).

Although the search term of chlamydia was unexpectedly not associated with STI rates by state and we were unable to assess the correlation for syphilis, trends in the graphical display

From the Division of Epidemiology and Biostatistics, University of Illinois at Chicago School of Public Health, Chicago, IL

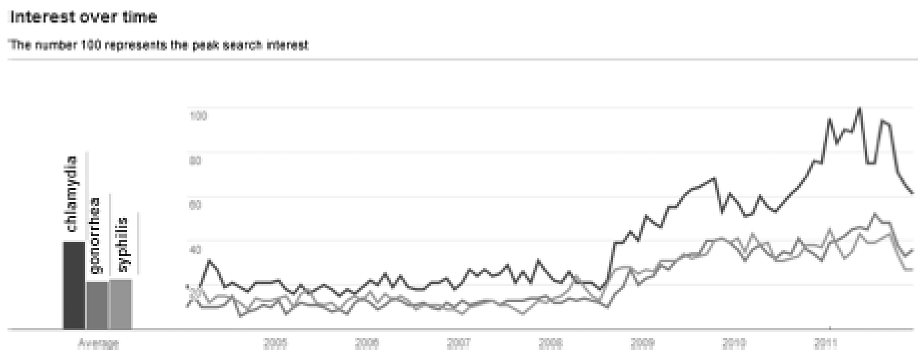
Conflict of interest: None declared.

Correspondence: Amy K. Johnson, MSW, Epidemiology and Biostatistics Division, School of Public Health, University of Illinois at Chicago, 1603 W. Taylor St, Room 966, Chicago, IL 60612. E-mail: akristen@uic.edu.

Received for publication August 13, 2013, and accepted October 21, 2013.

DOI: 10.1097/OLQ.0000000000000065

APPENDIX A (continued)



Chlamydia = "Chlamydia symptoms" search term.
Gonorrhea = "Gonorrhea symptoms" search term.
Syphilis = "Syphilis symptoms" search term.

Figure 1. Trend in Google search terms over time.

of the data seem to be similar to overall prevalence curves for each STI. The lack of association between chlamydia search term and state rate of disease may be due to the short period of data analyzed. Because screening for chlamydia is much more common relative to screening for gonorrhea, if most chlamydia cases are detected asymptomatically, this may explain the lack of correlation between search terms regarding chlamydia symptoms and reported chlamydia rates. Although rates of syphilis

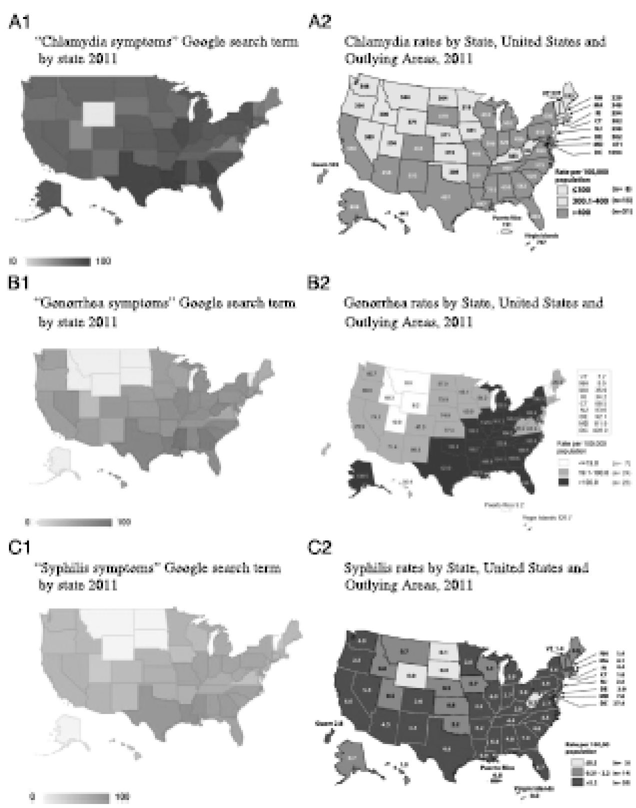


Figure 2. Search term frequency and STI rate by state.

APPENDIX A (continued)

remain steady overall, rates among women are declining and rates among men, particularly men who have sex with men, are increasing; this, along with greater awareness, may be related to the relative increase in syphilis search term frequency. We were surprised by the similarity in relative interest for gonorrhea and syphilis symptoms, despite the much higher rates of gonorrhea. This may reflect differences in information seeking by disease, symptomatic versus asymptomatic detection, or awareness or perceived severity of disease.

Because this is an ecological data analysis, findings cannot be applied at the individual level; for example, the trends are at a population level and cannot conclude that only STI-infected individuals are, in fact, those who are generating all STI-related search terms. There is uncertainty about the cause of trends in search terms (e.g., actual increases in infection, news or media attention) and when they occur (e.g., before STI diagnosis or after). The current study also does not account for differences in access to computer/Internet by region.

Data with actual usage and more precise time intervals, as well as time series analysis, will enable statistical assessment of whether such data may be used to predict epidemics. Working with Google to enhance user flexibility in terms of export data files, options for data intervals, enhanced geographic capabilities, and increased transparency in the methods of forecasting will be essential for future analysis of search engine data. Although a direct export function of search trend data exists, the user needs more flexibility including options with data available at shorter intervals, such as weekly or monthly, as well as information about missing data. In addition, geographic limitations are present; the distribution of search volume is unclear beyond the top cities within states, and there is no ability to modify the display maps (e.g., mapping by category or cutoff). Forecasting ability is somewhat limited; the specific methods that Google uses to forecast trends are not reported on the Google Trends Web site.

Search engine trends are an innovative tool to integrate into real-time surveillance of STIs. Health departments, who have access to weekly rates, may be able to forecast trends to anticipate increases in STIs or epidemics. Public health professionals should work with Google to refine trending tools for infectious disease surveillance. Future directions should include collaborations with health departments, as well as increased specificity of search terms to match subpopulation data for surveillance trends.

REFERENCES

1. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2009; 457:1012.
2. Jena AB, Karaca-Mandic P, Weaver L, et al. Predicting new diagnoses of HIV infection using Internet search engine data. Clin Infect Dis 2013; 56:1352–1353.
3. Pelat C, Turbelin C, Bar-Hen A, et al. More diseases tracked by using google trends. Emerg Infect Dis 2009; 1327–1328.
4. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. Can Med Assoc J 2009; 180:829–831. doi: 10.1503/cmaj.1090215.
5. Who's Online. 2011. Washington, DC: Pew Research Center. Available at: http://www.pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx. Accessed July 22, 2013.
6. Fox S. Pew Internet: Health. 2013. Available at: http://www.pewinternet.org/Commentary/2011/November/Pew-Internet-Health.aspx. Accessed July 22, 2013.
7. Google Trends How To. Available at: https://support.google.com/trends/topic/13975?hl=en&ref_topic=13762. Updated 2013. Accessed July 3, 2013.
8. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2011. Atlanta, USA: Department of Health and Human Services.
9. StataCorp. Stata statistical software: Release 13. College Station, TX: StataCorp LP, 2013.

APPENDIX B

programming document, survey will be collected online

Health Department Survey

Consent screen

Opening screen:

“Thank you for your interest in our survey. We would like to know your opinions and thoughts about integrating new technology into current public health practice, specifically surveillance of sexually transmitted diseases. Your responses will remain confidential and anonymous. The survey should take 5-10 minutes to complete."

1. How often do you use apps on a smartphone or tablet for personal use?

Everyday	01
A few days per week	02
About once a week	03
Less than once a week	04
Does not apply, I do not use apps	05

2. Have you heard of Google Flu Trends?

Yes	0 1
No	0 0

(On screen break) Google Flu Trends is an application created in partnership by researchers at the CDC and engineers at Google. Google Flu Trends uses search engine data to predict outbreaks of influenza 7-10 days earlier than traditional surveillance methods.
If a similar application called Google STI Trends existed that could help predict trends and identify outbreaks in reportable sexually transmitted infections earlier than traditional surveillance methods, how … **(repeats on each screen break)**

APPENDIX B (continued)

3. useful would this tool be for you?

Not at All useful	Somewhat useful	Very Useful
1	2	3

4. interested would you be in using Google STI trends?

Not at All interested	somewhat interested	Very interested
1	2	3

5. likely would your department be to integrate Google STI Trends into current surveillance?

Not at All likely	somewhat likely	Very likely
1	2	3

6. likely is it that a tool like Google STI Trends could improve response to outbreaks?

Not at All likely	somewhat likely	Very likely
1	2	3

There may be challenges or barriers to implementing a new method into an existing system. Please indicate your level of agreement or disagreement with the following items.

7
.

	Strongly Agree	Agree	Disagree	Strongly Disagree
a. My department is too busy to integrate new tools	1	2	3	4
b. The current system does not need to be modified	1	2	3	4
c. I do not believe Google STI trends will be valuable	1	2	3	4

APPENDIX B (continued)

What is your preferred learning style for learning a new surveillance tool or technique? (select all that apply)

Webinars	01
Large group sessions	02
Hardcopy literature	03
Professional conference events	04
One-on-one sessions	05
Talking to a coworker	06
Small group sessions	07
Other (please specify)	08
[text box]	

If 1 means not at all important and 5 means very important, how would you rate the importance of the following surveillance related tasks?

8. Using a systems that does not increase my work load

Not At all Important	Somewhat Important	Very Important
1	2	3

9. Surveillance data that are readily accessible in real time

Not At all Important	Somewhat Important	Very Important
1	2	3

10. improved sharing of information (e.g., with stakeholders, community members)

Not At all Important	Somewhat Important	Very Important
1	2	3

11. increase my understanding of disease under surveillance

Not At all Important	Somewhat Important	Very Important
1	2	3

12. Support from administration to incorporate advanced surveillance tools

Not At all Important	Somewhat Important	Very Important
1	2	3

APPENDIX B (continued)

13. Surveillance tools that have automatic (program generated) alerts for potential important changes in disease occurrence.

Not At all Important	Somewhat Important	Very Important
1	2	3

14. Having a specific person readily available to assist me with technological difficulties

Not At all Important	Somewhat Important	Very Important
1	2	3

We are interested in understanding more about the response threshold for outbreaks in reportable STIs. These next questions are about outbreak response.

15. Does your health department have a method in place to detect outbreaks for the following diseases?

a. Syphilis Yes No

If Yes, please briefly describe: (<text box>)

b. Gonorrhea Yes No

If Yes, please briefly describe: (<text box>)

c. Chlamydia Yes No

If Yes, please briefly describe: (<text box>)

16. Does your health department have an outbreak response plan for the following diseases?

a. Syphilis	Yes	No
b. Gonorrhea	Yes	No
c. Chlamydia	Yes	No

17. in the past 3 years, has your health department actively responded to outbreaks for

APPENDIX B (continued)

The following diseases?

a.	Syphilis	Yes	No
b.	Gonorrhea	Yes	No
c.	Chlamydia	Yes	No

If yes for any 18a-c,
18. Briefly describe the outbreak response:

The next questions are about your background.

19. What is your birth sex?

Male	01
Female	02
Refused	98

20. How old are you?

18-24	01
25-34	02
35-44	03
45-54	04
55-64	05
65+	06
Refused	98

21. What is the highest grade or level of formal education you have completed? Would you say…

Less than high school graduate,	01
High school graduate/GED,	02
College graduate, or	03

APPENDIX B (continued)

Post graduate degree?	04
Refused	98

These next questions are about your current job at the local health department.

22. Roughly, how many people are employed in your division or unit of the health department?

1-9	01
10-24	02
25-50	03
51-75	04
75+	05

23. How many of these (response to #4) employees report directly to you?

[numeric response]

24. From the list below, please select the title which best describes your current position.

Chief or commissioner	01
Deputy	02
Director	03
Manager	04
Epidemiologist	05
Assistant	06
Disease Intervention Specialist	07
Other (please specify)	08

25. How long have you been in your current position?

Less than 1 year	01
1-5 years	02
6-10 years	03
11-15 years	04
15+ years	05
Refused	98

APPENDIX B (continued)

26. What geographic region do you live in?

Northeast	01
West	02
Midwest	03
South	04
Refused	98

That was the last question. Thank you for participating in the survey.

APPENDIX C

Study ID ______________ Date________________________
Clinic site ID ___________ Interviewer ID _______________

I-HEALTH Project

1. Do you have access to the Internet?

Yes 01
No .... 02→***skip to** Q10*

1b. How do you access the Internet?

Interviewer: Read list and circle all that apply

Handheld personal device,
such as a smart phone or tablet 01
Lap top or desktop computer 02
Other___________________________________ 04

1c. Where do you access the Internet?

Interviewer: Read list and circle all that apply

At home 01
At friend's or family's homes 02
In public spaces, such as school, café or library 04

2. Have you ever searched the Internet to find information about health or medical topics?

Yes 01
No .... 02→***SKIP toQ3***

2b. In the past 3 months, how often have you used the internet to find information about health or medical topics?

1-2 times per month 01
Once a week 02
2-3 times a week 03
Daily 0

APPENDIX C (continued)

2c. Why did you search for health or medical topics? (**Please select all that apply**)

School assignment 01
Work related 02
Condition you have 04
Condition someone you know has 08
Other________________ 16

3. Have you ever searched the Internet to find information about sexually transmitted infections (STIs or STDs), which includes infections like chlamydia, gonorrhea, syphilis, herpes and genital warts?

Yes 01
No 02

4. Have you ever searched the Internet to find information about genital health, which includes information about vaginal or penile discharge, genital ulcers, or burning/itching when you pee?

Yes 01
No 02

Interviewer **if BOTH Q3 and Q4 = NO skip to Q10**

5. Did you use the internet to search for information about sexually transmitted infections before coming to the clinic today?

Yes 01
No 02

6. When was your most recent internet search for STI information?

Today 01
1-3 days ago 02
Within the past week 03
Within the past 2 weeks 04
Within the past month 05
More than one month ago 06

7 In general when do you use the internet to search for information about sexually transmitted infections?

	Yes	No	refuse
a. Before having unprotected sex	1	2	8
b. After having unprotected sex	1	2	*8*
c. After experiencing symptoms (burning/itching)	1	2	*8*
d. Before getting tested for STIs	1	2	*8*
e. After getting diagnosed with a STI	1	2	8
f. Any other reasons?	1	2	*8*

APPENDIX C (continued)

8. I am going to read a list of reasons people use the internet to search for information about STIs or STDs. In general, did you look for information because

	Yes	No	refuse
a. You wanted to learn more about STIs	1	2	8
b. You thought you might have an STI	1	2	*8*
c. You wanted to know how to prevent getting an STI	1	2	*8*
d. You wanted to know how to treat an STI	1	2	*8*
e. You wanted to find a place to get tested for STIs	1	2	*8*
f. Any other reasons?	1	2	*8*

Specify:__________________________________

8. Please tell me all the different types of websites you use to access information on STIs.

	Yes	No	refuse
a. Google	1	2	8
b. Other search engine (NOT google)	1	2	*8*
c Medical or health website, such as WebMD, Mayoclinic		2	*8*
d. Government website, such as CDC, NIH	1	2	*8*
e. This STD clinic website or CDPH website	1	2	*8*
f. Social media such as Facebook, Twitter	1	2	*8*
g. General websites such as Wikipedia	1	2	*8*
h. Any other type of website?	1	2	*8*

Specify:__________________________________

9. Please tell us what words or phrases you used to find the information.

____

____

____

APPENDIX C (continued)

This next section asks about your background.

10. How old are you? ____________ *Years*

11. What is your birth sex?

Male 01
Female 02

12. What is your current gender identity?

Male 01
Female 02
Transgender (male-to-female) 03
Transgender (female-to-male) 04
Other 05

13. Do you consider yourself to be Hispanic or Latino?

Yes 01
No 02

14. What is your racial background? Please tell me all that apply.

White, ... 01
Black or African American, ... 02
Asian, ... 04
Native Hawaiian or other Pacific Islander, or 08
American Indian or Alaskan native 16

APPENDIX C (continued)

15. What is the highest grade or level of formal education you have completed? Would you say…

Less than high school graduate, ... 01
High school graduate/GED, ... 02
Some College, ... 03
College graduate, or 04
Post graduate degree? 05

16. Do you think of yourself as (choose only one)…

Heterosexual or straight, ... 01
Bisexual, ... 02
Homosexual , gay, or lesbian 03
Other? ***(Please specify)*** 04
____

17. Are you currently a student?

Yes 01
No 02

18. Are you currently employed?

Yes 01
No .... 02 **skip to19**

18b. Are you employed part time or full time?

Part time (0 to 32 hours a week) 01
Full time (+32 hours a week) 02

19. What is the ZIP code where you live? ____________________

APPENDIX C (continued)

20. How confident are you filling out medical forms by yourself? Would you say…

Extremely, ... 01
Quite a bit,.. 02
Somewhat, ... 03
A little bit, or 04
Not at all 05

21. Do you have health insurance?

Yes, parents' insurance plan.................................... 01
Yes, government insurance
(Medicaid, medicare, etc)... 02
Yes, private insurance.. 03
No coverage of any type .... 04 **skip to22**
Don't know .... 05 **skip to22**

21b. If YES, would you be willing to use your health insurance for today's visit?

Yes .... 01 **skip to22**
No 02

21c. If NO, because (choose all that apply)

I do not want my insurance company to know....... 01
Insurance might send records home 02
I do not want my parents/spouse/significant other to know............... 04
Usual doctor might send records home................. 08
I cannot afford the co-pay/deductible 16
My insurance will not cover this visit...................... 32
Other:______________________ 64

APPENDIX C (continued)

22. Why did you come to the clinic today?

Interviewer: Read list and circle all that apply

You have symptoms you think are from an STD 01
A partner told you to get tested because of possible STD exposure 02
You get screened for STDs or HIV regularly 04
A medical reason that is not related to STDs or sexual health 08
Other (please specify) 16

____

Finally I'm going to ask you some questions about your sexual behavior. Remember all of your responses are confidential and you can skip any question you don't want to answer.

23. Think about the last time you had vaginal or anal sex. Was a condom used?

Yes 01
No.. 02
Refuse *8*
I have never had sex .... **END SURVEY**

24. How many sex partners (people you have had either vaginal or anal sex with) have you had in the past 6 months? ____

number of sex partners

Refuse.. *8*

25. In your lifetime, have you had sex with…

Only men.. 01
Only women 02
Both men and women 03
Refuse.. *8*

APPENDIX C (continued)

26. Has a doctor, nurse or other health care provider ever told you that you had an STI or an STD?

Yes 01
No 02

If NO, →**END**

26b. Which STI?

Interviewer: Read list and circle all that apply

Gonorrhea 01
Chlamydia 02
Syphilis 04

HPV or genital warts 08
Herpes 16
Trichomoniasis 32
Other___________________________________ 64
Not sure/Don't remember name 97

That was my last question. Thank you for participating in the survey.

APPENDIX D

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice
Initial Review (Response To Modifications)

March 3, 2015

Amy Johnson, MSW, BA
Epidemiology and Biostatistics
1603 W Taylor
M/C 923
Chicago, IL 60612
Phone: (773) 544-0131 / Fax: (312) 922-2916

RE: **Protocol # 2015-0192**
"I-Health"

Dear Ms. Johnson:

Your Initial Review (Response To Modifications) was reviewed and approved by the Expedited review process on February 26, 2015. You may now begin your research.

Please note the following information about your approved research protocol:

Protocol Approval Period:	February 26, 2015 - February 26, 2016
Approved Subject Enrollment #:	307
Additional Determinations for Research Involving Minors:	These determinations have not been made for this study since it has not been approved for enrollment of minors.
Performance Sites:	UIC
Sponsor:	None
PAF#:	Not applicable

Research Protocol:

a) I-HEALTH: A brief survey examining Internet search prevalence in an STD clinic patient population in Chicago; Version 1, 12/29/2014

Recruitment Materials:

a) I-Health Study eligibility screener; Version 2.0, 02/19/2015

APPENDIX D (continued)

b) I-Health Study recruitment script; Version 1.0, 02/19/2015

Informed Consent:

a) Waiver of Signed Consent Document (documentation) and an alteration of consent has been granted under 45 CFR 46.117; minimal risk

b) Subject information sheet, I-Health Study; Version 2.0

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific categories:

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
02/10/2015	Initial Review	Expedited	02/11/2015	Modifications Required
02/25/2015	Response To Modifications	Expedited	02/26/2015	Approved

Please remember to:

→ Use your **research protocol number** (2015-0192) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the OPRS website under:
"UIC Investigator Responsibilities, Protection of Human Research Subjects"
(http://tigger.uic.edu/depts/ovcr/research/protocolreview/irb/policies/0924.pdf)

Please note that the UIC IRB has the right to seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 996-9299. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Anna Bernadska, M.A.
IRB Coordinator, IRB # 2
Office for the Protection of Research Subjects

APPENDIX D (continued)

Enclosures:

1. **Informed Consent Document:**
    a) Subject information sheet, I-Health Study; Version 2.0
2. **Recruiting Materials:**
    a) I-Health Study eligibility screener; Version 2.0, 02/19/2015
    b) I-Health Study recruitment script; Version 1.0, 02/19/2015

cc: Ronald C. Hershow, Epidemiology and Biostatistics, M/C 923
Supriya Mehta, Faculty Sponsor, Epidemiology and Biostatistics, M/C 923

APPENDIX D (continued)

Department of Public Health

CITY OF CHICAGO

January 27, 2015

Amy Johnson
Principal Investigator
University of Illinois at Chicago
1603 W. Taylor St.
Chicago, IL 60612

Dear Ms. Johnson:

Thank you for submitting proposal entitled *"I-Health: A brief survey examining internet search prevalence in an STD clinic patient population in Chicago"*, which is assigned **CDPH #15-01**. This proposal was reviewed and approved by Expedited Review.

Waiver of written informed consent is approved: there is minimal risk to the subjects, and a written consent could possibly lead to identification.

This approval has a contingent requirement that you forward a summary or copy of your findings to the chairperson of the IRB.

This approval is only valid for one year. You are required to submit an annual progress report at the time of your yearly anniversary. Should you not file this report, the IRB will suspend your research project. All changes in the protocol or consent form and any serious and unexpected adverse events must be reported to the Board immediately.

Sincerely,

Arthur Kohrman, M.D.
Chair

cc: IRB Files

APPENDIX D (continued)

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

March 30, 2015

Amy Johnson, MSW, BA
Epidemiology and Biostatistics
1603 W Taylor
M/C 923
Chicago, IL 60612
Phone: (773) 544-0131 / Fax: (312) 922-2916

RE: **Research Protocol # 2015-0372**
"Project Tech Launch: Survey of STD Surveillance Workers Outbreak Response and New Technology"

Dear Ms. Johnson:

Your Claim of Exemption was reviewed on March 30, 2015 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)]. You may now begin your research

Your research may be conducted at UIC and with Adults.

The specific exemption category under 45 CFR 46.101(b) is:

2 Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have

APPENDIX D (continued)

responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.
2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:
   a. The researchers affiliation; UIC, JBVMAC or other institutions,
   b. The purpose of the research,
   c. The extent of the subject's involvement and an explanation of the procedures to be followed,
   d. Whether the information being collected will be used for any purposes other than the proposed research,
   e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
   f. Description of any reasonable foreseeable risks,
   g. Description of anticipated benefit,
   h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
   i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
   j. A statement that the UIC IRB/OPRS or JBVMAC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:
→Use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

APPENDIX D (continued)

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 996-2014 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Sandra Costello
Assistant Director, IRB # 2
Office for the Protection of Research Subjects

cc: Ronald C. Hershow, Epidemiology and Biostatistics, M/C 923
Supriya Mehta (faculty advisor), Epidemiology and Biostatistics, M/C 923

CITED LITERATURE

1. Centers for Disease Control and Prevention. 2012 sexually transmitted diseases surveillance report. 2014.

2. Pew Research Center. Internet user demographics. 2013.Retrieved August 10, 2014 , from http://www.pewinternet.org/data-trend/teens/internet-user-demographics/.

3. Satterwhite CL, Torrone E, Meites E, et al. Sexually transmitted infections among US women and men: Prevalence and incidence estimates, 2008. *Sex Transm Dis*. 2013; 40(3):187-193.

4. Low N, Bender N, Nartey L, Shang A, Stephenson JM. Effectiveness of chlamydia screening: Systematic review. *Int J Epidemiol*. 2009; 38(2):435-448.

5. Center for Disease Control and Prevention. Incidence, prevalence and cost of sexually transmitted infections in the united states. 2014. Retrieved August 2014, from http://www.cdc.gov/std/stats/STI-Estimates-Fact-Sheet-Feb-2013.pdf.

6. Allsworth JE, Ratner JA, Peipert JF. Trichomoniasis and other sexually transmitted infections: Results from the 2001-2004 national health and nutrition examination surveys. *Sex Transm Dis*. 2009; 36(12):738-744.

7. Centers for Disease Control and Prevention. 2009 sexually transmitted disease surveillance report.. 2010.

8. Farley TA, Cohen DA, Elkins W. Asymptomatic sexually transmitted diseases: The case for screening. *Prev Med*. 2003;36(4):502-509.

9. Korenromp EL, Sudaryo MK, de Vlas SJ, et al. What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? *Int J STD AIDS*. 2002;13(2):91-101.

10. Nelson KE, Williams CFM. Infectious disease epidemiology: Theory and practice. In: *Infectious disease epidemiology: Theory and practice.* Jones and Bartlett Publishers; 2007.

11. Centers for Disease Control and Prevention (CDC). CDC grand rounds: Chlamydia prevention: Challenges and strategies for reducing disease burden and sequelae. *MMWR Morb Mortal Wkly Rep*. 2011;60(12):370-373.

CITED LITERATURE (continued)

12. Cunningham SD, Kerrigan DL, Jennings JM, Ellen JM. Relationships between perceived STD-Related stigma, STD-Related shame and STD screening among a household sample of adolescents. *Perspectives on sexual and reproductive health*. 2009;41(4):225-230.

13. Kraut-Becher JR, Aral SO. Gap length: An important factor in sexually transmitted disease transmission. *Sex Transm Dis*. 2003;30(3):221-225.

14. Singer A. The uterine cervix from adolescence to the menopause. *BJOG: An International Journal of Obstetrics & Gynecology*. 1975;82(2):81-99.

15. Holmes KK. *Sexually transmitted diseases.* McGraw-Hill Medical; 2008.

16. Centers for Disease Control and Prevention. The national plan to eliminate syphilis from the United States. *Atlanta, GA:* 2006.

17. Peterman TA, Heffelfinger JD, Swint EB, Groseclose SL. The changing epidemiology of syphilis. *Sex Transm Dis*. 2005;32:S4-S10.

18. Thomas JC, Levandowski BA, Isler MR, Torrone E, Wilson G. Incarceration and sexually transmitted infections: A neighborhood perspective. *Journal of Urban Health*. 2008;85(1):90-99.

19. Laumann EO, Youm Y. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. *Sex Transm Dis*. 1999;26(5):250-261.

20. Mimiaga MJ, Helms DJ, Reisner SL, et al. Gonococcal, chlamydia, and syphilis infection positivity among MSM attending a large primary care clinic, Boston, 2003 to 2004. *Sex Transm Dis*. 2009;36(8):507-511.

21. Scholes D, Stergachis A, Heidrich FE, Andrilla H, Holmes KK, Stamm WE. Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. *N Engl J Med*. 1996;334(21):1362-1366.

22. Koumans EH, Rosen J, van Dyke MK, et al. Prevention of mother-to-child transmission of infections during pregnancy: Implementation of recommended interventions, United States, 2003-2004. *Obstet Gynecol*. 2012;206(2):158.

CITED LITERATURE (continued)

23. Kirkcaldy RD, Kidd S, Weinstock HS, Papp JR, Bolan GA. Trends in antimicrobial resistance in Neisseria gonorrhoeae in the USA: The gonococcal isolate surveillance project (GISP), January 2006-June 2012. *Sex Transm Infect*. 2013;89 Suppl 4:iv5-10.

24. Brown DL, Frank JE. Diagnosis and management of syphilis. *Am Fam Physician*. 2003;68(2):283-290.

25. Garnett GP, Aral SO, Hoyle DV, Cates Jr W, Anderson RM. The natural history of syphilis: Implications for the transmission dynamics and control of infection. *Sex Transm Dis*. 1997;24(4):185-200.

26. Centers for Disease Control and Prevention. STD surveillance retrieved March 23, 2013 from http://www.cdc.gov/std/stats11/toc.htm. Updated 2012.

27. Minnis AM, Padian, NS. Effectiveness of female controlled barrier methods in preventing sexually transmitted infections and HIV: Current evidence and future research directions. *Sex Transm Infect.* 2005:81 (3): 193-200

28. Hood JE, Hogben M, Chartier M, Bolan G, Bauer H. Dual contraceptive use among adolescents and young adults: Correlates and implications for condom use and sexually transmitted infection outcomes. *J Fam Plann Reprod Health Care*. 2014;40(3):200-207.

29. Abdool Karim SS, Richardson BA, Ramjee G, et al. Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women. *AIDS*. 2011;25(7):957-966.

30. Gottlieb SL, Low N, Newman LM, Bolan G, Kamb M, Broutet N. Toward global prevention of sexually transmitted infections (STIs): The need for STI vaccines. *Vaccine*. 2014;32(14):1527-1535.

31. Zimet GD, Mays RM, Fortenberry DJ. Vaccines against sexually transmitted infections: Promise and problems of the magic bullets for prevention and control. *Sex Transm Dis*. 2000;27(1):49-52.

32. Wawer MJ, Makumbi F, Kigozi G, et al. Circumcision in HIV-infected men and its effect on HIV transmission to female partners in Rakai, Uganda: A randomised controlled trial. *The Lancet*. 2009;374(9685):229-237.

CITED LITERATURE (continued)

33. Mattson CL, Campbell RT, Bailey RC, Agot K, Ndinya-Achola J, Moses S. Risk compensation is not associated with male circumcision in Kisumu, Kenya: A multi- faceted assessment of men enrolled in a randomized controlled trial. *PLoS One*. 2008;3(6):e2443.

34. Mehta SD, Moses S, Agot K, et al. Adult male circumcision does not reduce the risk of incident neisseria gonorrhoeae, chlamydia trachomatis, or trichomonas vaginalis infection: Results from a randomized, controlled trial in Kenya. *J Infect Dis*. 2009;200(3):370-378.

35. Eaton LA, Huedo-Medina TB, Kalichman SC, et al. Meta-analysis of single-session behavioral interventions to prevent sexually transmitted infections: Implications for bundling prevention packages. *Am J Public Health*. 2012;102(11):e34-44.

36. Hogben M. Partner notification for sexually transmitted diseases. *Clin Infect Dis*. 2007;44 Suppl 3:S160-74.

37. Centers for Disease Control and Prevention. "Recommendations for partner services programs for HIV infection, syphilis, gonorrhea, and chlamydial infection." *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control* 57, no. RR-9 (2008): 1.

38. Cohen DA, Nsuami M, Etame RB, et al. A school-based chlamydia control program using DNA amplification technology. *Pediatrics*. 1998;101(1):e1-e1.

39. Wang LY, Burstein GR, Cohen DA. An economic evaluation of a school-based sexually transmitted disease screening program. *Sex Transm Dis*. 2002;29(12):737- 745.

40. Ahmed K, Scholle S, Baasiri H, et al. Chlamydia screening among sexually active young female enrollees of health plans-united states, 2000-2007. *Morb Mortal Weekly Rep*. 2009;58(14):362-365.

41. White PJ, Ward H, Cassell JA, Mercer CH, Garnett GP. Vicious and virtuous circles in the dynamics of infectious disease and the provision of health care: Gonorrhea in Britain as an example. *J Infect Dis*. 2005;192(5):824-836.

42. Kretzschmar M, van Duynhoven YT, Severijnen AJ. Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations. *Am J Epidemiol*. 1996;144(3):306-317.

CITED LITERATURE (continued)

43. Van Beneden C, Lyndfield R. Public health surveillance for infectious diseases. *Principles and practice of public health surveillance*. 2010:236-254.

44. Choi BC. The past, present, and future of public health surveillance. *Scientifica*.;2012.

45. Declich S, Carter AO. Public health surveillance: Historical origins, methods and evaluation. *Bull World Health Organ*. 1994;72(2):285.

46. Morse SS. Public health surveillance and infectious disease detection. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*. 2012;10(1):6-16.

47. Rolka H, Walker MDW, English MR, Katzoff MJ, Scogin G, Neuhaus E. Analytical challenges for emerging public health surveillance. *CDC's Vision for Public Health Surveillance in the 21st Century*. 2012;61:35.

48. Rietmeijer CA, Donnelly J, Bernstein KT, et al. Here comes the SSuN: Early experiences with the STD surveillance network. *Public Health Rep*. 2009;124(Suppl 2):72.

49. Kirkcaldy RD, Kidd S, Weinstock HS, Papp JR, Bolan GA. Trends in antimicrobial resistance in Neisseria gonorrhoeae in the USA: The gonococcal isolate surveillance project (GISP), January 2006–June 2012. *Sex Transm Infect*. 2013;89(Suppl 4):iv5-iv10.

50. Gerberding TM. Report to congress: Infertility and prevention of sexually transmitted diseases. 2004.

51. National Center for Chronic Disease Prevention and Health Promotion. Behavioral Risk Factor Surveillance System retrieved June 27th 2014 from http://www.cdc.gov/brfss/. Updated May 8, 2014.

52. May L, Chretien J, Pavlin JA. Beyond traditional surveillance: Applying syndromic surveillance to developing settings–opportunities and challenges. *BMC Public Health*. 2009;9(1):242.

53. Centers for Disease Control and Prevention. Program operations guidelines for STD prevention: Outbreak response plan. 2011.

CITED LITERATURE (continued)

54. German RR, Lee L, Horan J, Milstein R, Pertowski C, Waller M. Updated guidelines for evaluating public health surveillance systems. *MMWR Recomm Rep*. 2001;50(RR- 13):1-35.

55. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. *Am J Epidemiol*. 1994;140(5):383-397.

56. Gaffga NH, Samuel MC, Stenger MR, Stover JA, Newman LM. The OASIS project: Novel approaches to using STD surveillance data. *Public Health Rep*. 2009;124(Suppl 2):1.

57. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the web for public health surveillance. *N Engl J Med*. 2009;360(21):2153-2157.

58. Wilson K, Brownstein JS. Early detection of disease outbreaks using the internet. *Canadian Medical Association Journal*. 2009;180(8):829-831.

59. Jenkins WD, Wold B. Use of the internet for the surveillance and prevention of sexually transmitted diseases. *Microb Infect*. 2012;14(5):427-437.

60. Internet world stats. Retrieved May 7, 2014 from http://www.internetworldstats.com/stats.htm 2012.

61. Fox S. Pew internet: Health 2013. Retrieved July 22, 2013 from http://www.pewinternet.org/Commentary/2011/November/Pew-Internet-Health.aspx.

62. Finney Rutten, Lila J., et al. "Picking up the pace: changes in method and frame for the health information national trends survey (2011–2014)." *Journal of health communication* 2012:17 (8): 979-989..

63. Eysenbach G. 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*. 2009;11(1):e11.

64. Eysenbach G. Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. 2005:244-248.

65. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457:1012.

CITED LITERATURE (continued)

66. Jena AB, Karaca-Mandic P, Weaver L, Seabury SA. Predicting new diagnoses of HIV infection using internet search engine data. *Clinical infectious diseases*. 2013;56(9):1352-1353.

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

68. Zhou X, Ye J, Feng Y. Tuberculosis surveillance by analyzing google trends. *Biomedical Engineering, IEEE Transactions on*. 2011;58(8):2247-2254.

69. Gu H, Chen B, Zhu H, et al. Importance of internet surveillance in public health emergency control and prevention: Evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. *Journal of medical Internet research*. 2014;16(1).

70. Carneiro HA, Mylonakis E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*. 2009;49(10):1557-1564.

71. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron AJ. More diseases tracked by using google trends. *Emerging Infectious Diseases*. 2009: 15 (8):1327.

72. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance. *PLoS neglected tropical diseases*. 2011;5(5):e1206.

73. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of internet-based dengue query data: Google dengue trends. *PLoS neglected tropical diseases*. 2014;8(2):e2713.

74. Zhou X, Shen H. Notifiable infectious disease surveillance with data collected by search engine. *Journal of Zhejiang University SCIENCE C*. 2010;11(4):241-248.

75. Miller EA, West DM, Wasserman M. Health information websites: Characteristics of US users by race and ethnicity. *J Telemed Telecare*. 2007;13(6):298-302.

76. Laz TH, Berenson AB. Racial and ethnic disparities in internet use for seeking health information among young women. *J Health Commun*. 2013;18(2):250-260.

77. Lenhart A, Purcell K, Smith A, Zickuhr K. Social media & mobile internet use among teens and young adult millennials. *Pew Internet & American Life Project*. 2010.

CITED LITERATURE (continued)

78. Sadasivam RS, Kinney RL, Lemon SC, Shimada SL, Allison JJ, Houston TK. Internet health information seeking is a team sport: Analysis of the pew internet survey. *Int J Med Inf*. 2013;82(3):193-200.

79. Rideout, V. Generation Rx. com: How young people use the Internet for health information. A Kaiser Family Foundation Survey. 2001:1-46.

80. Rietmeijer CA, Bull SS, McFarlane M, Patnaik JL, Douglas JM. Risks and benefits of the internet for populations at risk for sexually transmitted infections (STIs): Results of an STI clinic survey. *Sex Transm Dis*. 2003;30(1):15-19.

81. Google search statistics. Retrieved on May 17, 2014, from http://www.internetlivestats.com/google-search-statistics/.Updated 2014.

82. Buhi ER, Daley EM, Fuhrmann HJ, Smith SA. An observational study of how young people search for online sexual health information. *Journal of American college health*. 2009;58(2):101-111.

83. Jones RK, Biddlecom AE. Is the internet filling the sexual health information gap for teens? An exploratory study. *J Health Commun*. 2011;16(2):112-123.

84. Google trends how to. Retrieved on July, 3, 2013, from https://support.google.com/trends/topic/13975?hl=en&ref_topic=13762. Updated 2013.

85. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. . 2003.

86. Sprent P, Smeeton NC. *Applied nonparametric statistical methods.* CRC Press; 2007.

87. Cohen J, Cohen P, West SG, Aiken LS. *Applied multiple regression/correlation analysis for the behavioral sciences.* Routledge; 2013.

88. Kassner M. Search engine bias: What search engines are telling you (and what they're not). Retrieved August 19th, 2014, from http://www.techrepublic.com/blog/it-security/search-engine-bias-what-search-results-are-telling-you-and-what-theyre-not/. Updated 2013.

89. Shewan D. How google hummingbird changed the future of search. Retrieved August 19th, 2014, from http://www.wordstream.com/blog/ws/2014/06/23/google-hummingbird. Updated 2014.

90. Johnson AK, Mehta SD. A comparison of internet search trends and sexually transmitted infection rates using google trends. *Sex Transm Dis*. 2014;41(1):61-63.

91 National Association of County and City Health Officials. National profile of local health departments. 2013: 38.

92. Brownson RC, Reis RS, Allen P, et al. Understanding administrative evidence-based practices: Findings from a survey of local health department leaders. *Am J Prev Med*. 2014;46(1):49-57.

93. Jacobs JA, Clayton PF, Dove C, et al. A survey tool for measuring evidence-based decision making capacity in public health agencies. *BMC Health Serv Res*. 2012;12:57- 6963-12-57.

94. Carr JL, Sheffield PE, Kinney PL. Preparedness for climate change among local health department officials in New York state: A comparison with national survey results. *J Public Health Manag Pract*. 2012;18(2):E24-32.

95. Buehler JW, Sonricker A, Paladini M, Soper P, Mostashari F. Syndromic surveillance practice in the united states: Findings from a survey of state, territorial, and selected local health departments. *Advances in Disease Surveillance*. 2008;6(3):1-20.

96. Qualtrics Software. Qualtrics security white paper. 2014.

97. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *MIS quarterly*. 2003:425-478.

98. King WR, He J. A meta-analysis of the technology acceptance model. *Information & Management*. 2006;43(6):740-755.

99. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: Toward an integrative view. *Information & Management*. 2006;43(3):350-363.

100. Szklo M, Nieto FJ. *Epidemiology: Beyond the basics.* Jones & Bartlett Publishers; 2014.

101. Hong Y, Li X, Fang X, Lin X, Zhang C. Internet use among female sex workers in china: Implications for HIV/STI prevention. *AIDS and behavior*. 2011;15(2):273-282.

CITED LITERATURE (continued)

102. Borzekowski DL, Rickert VI. Adolescent cybersurfing for health information: A new resource that crosses barriers. *Arch Pediatr Adolesc Med*. 2001;155(7):813-817.

103. Gorbach PM, Mensch BS, Husnik M, et al. Effect of computer-assisted interviewing on self-reported sexual behavior data in a microbicide clinical trial. *AIDS and Behavior*. 2013;17(2):790-800.

104. Gribble JN, Miller HG, Rogers SM, Turner CF. Interview mode and measurement of sexual behaviors: Methodological issues. *J Sex Res*. 1999;36(1):16-24.

105. Pool R, Montgomery CM, Morar NS, et al. A mixed methods and triangulation model for increasing the accuracy of adherence and sexual behavior data: The microbicides development programme. *PloS one*. 2010;5(7):e11600.

106. SocioCultural Research Consultants. Dedoose web application for managing, analyzing and presenting qualitative and mixed method research data (6.1 ed). 2015.

107. Glaser BG. *Theoretical sensitivity: Advances in the methodology of grounded theory.* Sociology Pr.; 1978.

108. Centers for Disease Control and Prevention. 2013 sexually transmitted diseases surveillance report. 2015.

109. Lampos, V., Miller, A.C., Miller, S.C., and Stefansen, C. Advances in Nowcasting Influenza-Like Illness Rates using Search Query Logs. *Scientific Reports* 2015:5.

110. StataCorp. *Stata statistical software: Release 12*. 2012.

111. Barbour, R. S. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?. *Bmj*, 2001: *322*(7294), 1115-1117.

112. Varian, H. R. Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 2014: 3-27.

113. Kamvar, M., Shumeet B. A large scale study of wireless search behavior: Google mobile search. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006.

114. Wright, KB. Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication* 2005: 10(3).

CITED LITERATURE (continued)

115. Cook, C., Heath, F., & Thompson, R. L. A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, 2000: *60*(6): 821-836.

VITA

NAME: AMY KRISTEN JOHNSON

EDUCATION: UNIVERSITY OF ILLINOIS AT CHICAGO
SCHOOL OF PUBLIC HEALTH, CHICAGO, IL
PhD Candidate in Public Health Sciences May 2016
Dissertation title: *Surveillance of Sexually Transmitted Infections: Exploring the Utility of Internet Search Engine Data*

JANE ADDAMS SCHOOL OF SOCIAL WORK
MSW, May 2006

DEPAUL UNIVERSITY, CHICAGO, IL
BA in English and Sociology, June 2003

HONORS: E. Hill De Loney Award 2010
American Public Health Association Scholarship recipient 2010 "30 Under 30" Recipient; Windy City Times, Chicago 2009 Dean's List: UIC (2004-2006)
Dean's List: DePaul University (1999-2003)
Presidential Scholar: DePaul University (1999-2003)

PROFESSIONAL EXPERIENCE: AIDS FOUNDATION OF CHICAGO
Director of Research, Evaluation and Data Services
2013 – present

ANN AND ROBERT H. LURIE CHILDREN'S HOSPITAL
Behavioral Research Scientist 2012- 2013
Research coordinator 2007- 2012

UNIVERSITY OF ILLINOIS AT CHICAGO
Data analyst
2005- 2007
Research assistant
2003- 2005

TEACHING EXPERIENCE: UNIVERSITY OF ILLINOIS CHICAGO SCHOOL OF PUBLIC HEALTH DIVISION OF EPIDEMIOLOGY & BIOSTATISTICS
Teaching Assistant
Epidemiology of HIV/AIDS, Spring 2013, 2014

VITA (continued)

PUBLICATIONS:

Johnson, A.K., Mehta, S.D. (2014). A Comparison of Internet Search Trends and Sexually Transmitted Infection Rates Using Google Trends. *Sexually Transmitted Diseases*, 41 (1):61-63.

Lyons, T., **Johnson, A.K.,** Garofalo, R. (2013). “What Could Have Been Different”: A Qualitative Study of Syndemic Theory and HIV Prevention Among Young Men Who Have Sex With Men. *Journal of HIV/AIDS & Social Services*, 12 (3-4):368-383.

Hotton, A.L., Garofalo, R., Kuhns, L.M., **Johnson, A.K.** (2013). Substance Use as a Mediator of the Relationship between Life Stress and Sexual Risk Among Young Transgender Women. *AIDS Education and Behavior*, 25 (1): 62-71.

Hidalgo, M., Cotten, C., **Johnson, A.K**., Kuhns, L., Garofalo, R. (2013). “Yes, I am more than just that”: Gay/Bisexual Young Men Residing in the United States Discuss the Influence of Minority Stress on their Sexual Risk Behavior Prior to HIV Infection. *International Journal of Sexual Health*, 25 (4):291-304.

Mustanski, B., **Johnson, A.K.**, Ryan, D., Garofalo, R. (2012). Perceived likelihood of using HIV Pre-Exposure Prophylaxis medications among young men who have sex with men. *AIDS & Behavior*, 17 (6): 2173- 2179.

Dowshen, N., Kuhns, L., **Johnson, A.K.**, Holoyda, B.J., Garofalo, R. (2012). Improving Adherence to Antiretroviral Therapy for Youth Living with HIV/AIDS: A Pilot Study Using Personalized, Interactive, Daily Text Message Reminders. *Journal of Medical Internet Research*, 14 (2): e51.

Garofalo, R., **Johnson, A.K.,** Kuhns, L., Cotton, C. (2012) Lifeskills: Evaluation of a theory-driven behavioral HIV prevention for young transgender women. *Journal of Urban Health,* 89 (3):419-31.

VITA (continued)

Brennan, J., Kuhns, L., **Johnson, A.K.**, Blezer, M., Wilson, E., Garofalo,R. (2012). Syndemic Theory and HIV-related Risk among Young Transgender Women: The Role of Multiple, Co-Occurring Health Problems. *The American Journal of Public Health,*102(9).

Dowshen, N., Forke, C.M., **Johnson, A.K.**, Kuhns, L.M., Rubin, D., Garofalo, R. (2011). Religiosity as a protective factor against HIV risk among young transgender women. Journal of Adolescent Health, 48 (4)

SELECETED ORAL PRESENTATIONS:

Johnson, A.K., Nowicki, K., Smeleka, S., Terlikowski, J. (2015). "Assessing the landscape: What reproductive health and pregnancy planning services do Midwest medical providers offer HIV impacted women?" Oral presentation at the National Sexual Health Conference (Denver, CO).

Johnson, A.K., Buenrostro, R. (2015). "Salud y Orgullo Mexicano" Oral presentation at the Minority Health Conference of the Illinois Public Health Association (Lisle, IL).

Johnson, A.K., Buenrostro, R., Rivers, M. (2015). "Salud y Orgullo Mexicano: Removing barriers to HIV testing and care for men of Mexican descent" Oral presentation at the Health Disparities Conference (Chicago, IL).

Mehta, S.D., **Johnson, A.K.** (2014). ""What Can Internet Search Engines Reveal About Population STD Trends and Risk?" Mini plenary at the STD Prevention Conference (Atlanta, Georgia).