

Monitoring HIV Testing and Pre-Exposure Prophylaxis Information Seeking By Combining Digital and Traditional Data

Derek Johnson (✉ dcjohnson@ucsd.edu)

University of California San Francisco

Alicia L. Nobles

University of California San Diego

Theodore L. Caputi

University of York

Michael Liu

University of Oxford

Eric C. Leas

University of California San Diego

Steffanie A. Strathdee

University of California San Diego

Davey M. Smith

University of California San Diego

John W. Ayers

University of California San Diego

Research article

Keywords: Google Trends, HIV, PrEP, Internet, HIV Testing

Posted Date: October 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-92816/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 25th, 2021. See the published version at <https://doi.org/10.1186/s12879-021-05907-0>.

Abstract

Background. Public health is increasingly turning to non-traditional digital data to inform HIV prevention and control strategies. We demonstrate a parsimonious method using both traditional survey and internet search histories to provide new insights into HIV testing and pre-exposure prophylaxis (PrEP) information seeking that can be easily extended to other settings.

Method. We modeled how US internet search volumes from 2019 for HIV testing and PrEP compared against expected search volumes for HIV testing and PrEP using state HIV prevalence and socioeconomic characteristics as predictors. States with search volumes outside the upper and lower bound confidence interval were labeled as either over or under performing. State performance was evaluated by (a) Centers for Disease Control and Prevention designation as a hotspot for new HIV diagnoses (b) expanding Medicaid coverage.

Results. Ten states over-performed in models assessing information seeking for HIV testing, while thirteen states under-performed. Fourteen states over-performed in models assessing internet searches for PrEP information, while 15 states under-performed. States that expanded Medicaid coverage were more likely to over perform in PrEP models than states that did not expand Medicaid coverage.

Conclusion. Our study derived a method of measuring HIV and PrEP information seeking that is comparable across states. Several states exhibited information seeking for PrEP and HIV testing that deviated from model assessments. Statewide search volume for PrEP information was affected by a state's decision to expand Medicaid coverage. Our research provides health officials with an innovative way to monitor statewide interest in PrEP and HIV testing using a metric for information-seeking that is comparable across states.

Background

Despite substantial investments in HIV prevention and control, former CDC Director Tom Frieden recently called for the development of timely and actionable HIV surveillance to better inform HIV policy [1]. Past efforts to enhance HIV surveillance relied mostly on upscaling traditional data (e.g., clinical records or surveys) that have intrinsic shortcomings, such as a limited ability to provide current information. For instance, the most recent data for HIV testing on AIDSvu.org is from 2016 and the most recent AIDSvu.org data on Pre-Exposure Prophylaxis (PrEP) usage is from 2018 [2]. These limitations have driven public health to increasingly turn to digital data, such as news, social media, and internet searches, to learn how people seek HIV information [3–6]. For example, internet search trends can be used to investigate public interests as evident by actor Charlie Sheen's HIV positive disclosure concurring with record levels of Google searches for HIV awareness, HIV testing, and condoms [7]. This finding was valid, as it was later confirmed by traditional data after a 16 month delay [8]. Internet search histories have potential utility for assessing both help-seeking behavior regarding public interest in PrEP for HIV prevention and for HIV testing. For example, one study conducted in Hong Kong found that a direct

relation between HIV news trends and online search behavior for issues regarding HIV/AIDS and men who have sex with men (MSM) [9]. Other studies have found that areas with high levels of HIV prevalence have greater internet search volumes for HIV related terms than areas of low HIV prevalence [10]. These studies show that the use of internet search histories combined with traditional surveillance data has the potential to create synergies that can yield new insights into HIV related health behavior

Our study methods use both internet search histories and traditional survey data to provide new insights on information seeking for HIV testing and PrEP information that can be easily replicated and extended to other settings and outcomes. Specifically, we predicted expected internet search volumes for HIV testing and PrEP based on statewide HIV prevalence and socioeconomic (SES) factors and compared them to observed search volumes in a model that allows us to identify if US states over or under perform against expectations. Moreover, we evaluated how state performance varied by (a) states that are designated as hotspots for new HIV diagnoses (b) states that received Medicaid expansion funding.

Methods

Our study used data from multiple sources. 1) We obtained the state-level prevalence of HIV (for 2017) from the 2018 Center for Disease Control and Prevention HIV surveillance report [11]. HIV prevalence was chosen over HIV incidence because we were interested in look at the association between the total number of HIV cases in a state and internet searches for HIV testing and PrEP 2) The following state-level socioeconomic attributes we obtained from the 2018 Center for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS): proportion of males, white non-Hispanics, people aged 45 years or older, and people with household income over \$50,000 [12]. 3) We obtained 2019 state-level annual internet search volumes for HIV testing and PrEP from using the Google Trends API. Information available through this API includes the volume of searches for each term, the number of searches per unit of time, and the geographic location of the searches (country, region, state, city, metropolitan area). Search volume data was calculated as a query fraction of the proportion of searches of a specific search term relative to all searches measured per 10 million searches. Standardizing search volumes was done in order to account for population sizes. We defined HIV testing searches as any query that included the terms "HIV" and "test," "tests," or "testing", "AIDS test", or "oraquick". We defined PrEP searches as any query that included the terms "PrEP" and "HIV" or "pre-exposure prophylaxis HIV" or "Truvada" or "Descovy".

Internet search volumes are withheld by Google for states where searches do not achieve a minimum threshold of searches. As a result, we could not obtain search data for HIV testing for five states (Alaska, Montana, South Dakota, Vermont, and Wyoming). PrEP search data could not be obtained for two states (Vermont and Wyoming). 4) We obtained data on states that expanded Medicaid coverage from the Kaiser Family Foundation [13].

Our analysis followed a four-step process. First, we fit Poisson regression models with state-level HIV prevalence data and state-level socioeconomic attributes to predict the expected internet search volumes

of HIV testing and PrEP for each state. Second, we fit a centered least squares regression line of expected search volumes from our models versus observed search volumes from Google Trends. Third, we compared the expected search volumes from our models in step one with the observed search volumes from Google Trends for each state in order to assess the level of information seeking for HIV testing and PrEP by calculating the percent difference between the observed and expected values of the centered least squares regression (i.e., $(\text{observed-fitted})/\text{fitted} * 100\%$).

States with observed information seeking (measured by observed internet search volumes for HIV testing and PrEP) greater than their expected information seeking (predicted internet search volumes by the Poisson regression model) were considered to be over performing and exhibit greater information seeking for HIV testing and PrEP than expected given their prevalence of HIV. States that were over performing above the 95% confidence interval were highlighted in our results (see example of plotting observed vs. expected observations in Fig. 1). Similarly, states with observed information seeking less than their expected information seeking were considered to be underperforming and exhibit less information seeking for HIV testing and/or PrEP than expected given their prevalence of HIV. States that were under performing below the 95% confidence interval were highlighted in our results. To describe statistical uncertainty between expected and observed search volumes, we used bootstrap sampling to calculate the 95% confidence interval (CI) for the regression line and labeled observations outside the confidence band as states that over or under performed.

Fourth, to understand which states typically under or over performed we contrasted the deviations in expected searches against (a) What states were designated as hotspots for new HIV diagnoses by the Centers for Disease Control and Prevention (CDC) [11], (b) What states received Medicaid expansion funding that covered HIV testing and PrEP [13].

Results

We observed different levels of information seeking for HIV testing and PrEP across states (Fig. 1). Ten states over-performed for HIV testing searches. North Dakota exhibited the greatest difference with 39.9% more searches than expected followed closely by Georgia (36.8%), then Rhode Island (34.2%), Indiana (26.7%), Pennsylvania (20.8%), Florida (19.9%), Iowa (16.0%), Nevada (15.5%), Louisiana (15.3%), and Washington (12.4%) (Table 1). Conversely, thirteen states under-performed for HIV testing. New Hampshire exhibited the greatest difference with -31.3% less searches than expected, followed by Maine (-30.1%), Idaho (-24.7%), Nebraska (-22.2%), Oregon (-21.9%), Arkansas (-16.7%), New Mexico (-16.0%), Mississippi (-15.0%), Massachusetts (-14.5%), Alabama (-12.5%), Arizona (-12.3%), California (-10.5%), Illinois (-7.3%).

Table 1

Statewide Proportional Differences Between Observed and Expected Internet Search Volumes for Information Seeking About HIV testing and PrEP.

U.S State	HIV prevalence per 100,000 individuals	HIV Search Differences	PrEP Search Differences	Expanded Medicaid Coverage@	Designated Hotspot for New HIV Infections
Alabama	350	-12.5*	-4.3	Y	N
Alaska	131.4	NA	-5.6	Y	N
Arizona	314.7	-12.3*	12.8 [#]	Y	N
Arkansas	249.4	-16.7*	-0.7	Y	N
California	431.9	-10.5*	4.3	Y	Y
Colorado	299.3	2.4	4.6	Y	N
Connecticut	366.1	-3.7	4.3	Y	N
Delaware	336.4	-1.3	0.6	Y	N
District of Columbia	2,819.8	-8.0	0.7	Y	Y
Florida	705.9	19.9 [#]	-10.3*	N	Y
Georgia	694.4	36.8 [#]	-3.7	N	Y
Hawaii	240.7	-9.5	6.2	Y	N
Idaho	84.6	-24.7*	-28.5*	N	N
Illinois	367.3	-7.3*	11.1 [#]	Y	Y
Indiana	224.5	26.7 [#]	-5.2	Y	Y
Iowa	110.3	16.0 [#]	-15.1*	Y	N
Kansas	136.2	-1.0	-11.1*	N	N
Kentucky	207.2	0.3	10.8 [#]	Y	Y
Louisiana	603	15.3 [#]	-1.2	Y	Y
Maine	147.6	-30.1*	14.1 [#]	N	N

[#]Over performed in search models

*Underperformed in search models

@State decisions on the Affordable Care Act's Medicaid expansion

U.S State	HIV prevalence per 100,000 individuals	HIV Search Differences	PrEP Search Differences	Expanded Medicaid Coverage@	Designated Hotspot for New HIV Infections
Maryland	716	7.3	-6.6	Y	Y
Massachusetts	384.1	-14.5*	21.8 [#]	Y	Y
Michigan	210.2	4.9	-7.9*	Y	Y
Minnesota	197.1	-3.8	-6.4	Y	N
Mississippi	424.2	-15.0*	0.1	N	N
Missouri	264.1	-0.2	-7.1*	N	N
Montana	78.5	NA	-26.8*	Y	N
Nebraska	150.8	-22.2*	17.7 [#]	Y	N
Nevada	440.5	15.5 [#]	15.2 [#]	Y	N
New Hampshire	113	-31.3*	-22.2*	Y	N
New Jersey	505.6	-2.7	-16.1*	Y	Y
New Mexico	216	-16.0*	6.7	Y	N
New York	825.7	-0.6	5.3 [#]	Y	Y
North Carolina	404	-6.6	-9.7*	N	Y
North Dakota	67.8	39.9 [#]	13.6	Y	N
Ohio	240.7	4.7	1.2	Y	Y
Oklahoma	210	0.2	-12.3*	N	N
Oregon	224.5	-21.9*	10.9 [#]	Y	N
Pennsylvania	348.8	20.8 [#]	14.8 [#]	Y	Y
Rhode Island	294.6	34.2 [#]	31.1 [#]	Y	N
South Carolina	465.3	6.8	-24.8*	Y	N
South Dakota	78.3	NA	11.4 [#]	N	N
#Over performed in search models					
*Underperformed in search models					
@State decisions on the Affordable Care Act's Medicaid expansion					

U.S State	HIV prevalence per 100,000 individuals	HIV Search Differences	PrEP Search Differences	Expanded Medicaid Coverage@	Designated Hotspot for New HIV Infections
Tennessee	332	-5.8	1.4	Y	Y
Texas	443	3.9	-11.2*	Y	Y
Utah	133.2	-4.0	-3.3	N	N
Vermont	128.7	NA	NA	Y	N
Virginia	339	7.7	-6.4*	N	N
Washington	240.6	12.4 [#]	17.4 [#]	Y	Y
West Virginia	123.8	11.2	34.3 [#]	Y	N
Wisconsin	140.4	-6.8	-27.0*	N	N
Wyoming	70.4	NA	NA	N	N
#Over performed in search models					
*Underperformed in search models					
@State decisions on the Affordable Care Act's Medicaid expansion					

Fourteen states over-performed for PrEP searches (Table 1). West Virginia exhibited the greatest difference with 34.3% more searches than expected, followed by Rhode Island (31.1%), Massachusetts (21.8%), Nebraska(17.7%), Washington (17.4%), Nevada (15.2%), Pennsylvania (14.8%), Maine (14.1%), Arizona (12.8%), South Dakota (11.4%), Illinois (11.1%), Oregon (10.9%), Kentucky (10.8%), and New York (5.3%). Conversely, fifteen states under-performed for PrEP searches. Idaho exhibited the greatest difference with 28.5.% less searches than expected, followed by Wisconsin (-27.0%), Montana (-26.8%), South Carolina (-24.8%), New Hampshire (-22.2%), New Jersey (-16.1%), Iowa (-15.1%), Oklahoma (-12.3%), Texas(-11.2%), Kansas (-11.1%), Florida (-10.3%), North Carolina (-9.7%), Michigan (-7.9%), Missouri (-7.1%), and Virginia (-6.4%).

States that over or under performed on HIV testing searches did not necessarily do likewise for PrEP searches ($r = 0.16$). For instance, Georgia ranked 4th for excess HIV testing searches, but then ranked 35th for PrEP searches. Three states (Nevada, Pennsylvania, and Rhode Island) over-performed for both HIV testing and PrEP searches, while only 1 state (New Hampshire) under-performed for both. Model results for HIV testing and PrEP were similar regardless of a state being designated as a hotspot for new HIV diagnoses. States that expanded Medicaid coverage were more likely to over perform on PrEP searches compared to states that did not expand Medicaid coverage ($t = -2.38$, $p < 0.025$). There was no difference

in performance on HIV testing for states that expanded Medicaid coverage compared to states that did not.

Discussion

Our study derived a method of measuring HIV testing and PrEP information seeking that is comparable across states. Several states exhibited information seeking for PrEP and HIV testing that deviated from what was expected in our models. A state's performance in our models was not affected by its designation as a hotspot for new HIV infections. However, performance for PrEP information seeking was associated with a state's decision to expand Medicaid coverage. By integrating internet search histories and traditional survey data, our results provide baseline benchmarks for monitoring statewide interest in seeking information on HIV testing and PrEP.

Our research demonstrates a need for increased access to PrEP information, particularly among states that have not expanded their Medicaid coverage. Lower interest in seeking information on PrEP for states that did not expand Medicaid coverage could be detrimental to increasing PrEP utilization given that insurance coverage affects PrEP uptake [14, 15]. Approximately 12% of PrEP users receive PrEP through Medicaid [16] and the refusal to extend coverage could deny people the ability to access PrEP. Our results, coupled with the inability to utilize PrEP due to a lack of health insurance, is a potentially disastrous combination that could result in an increase in HIV prevalence in states that underperformed in our PrEP models.

Underperformance in PrEP models could be due to the unequal distribution of PrEP across different genders, ages, and states. Five states represented 50% of PrEP prescriptions and although women represent almost 20% of new HIV infections, they represented only 7% of PrEP prescriptions [17]. These types of underlying disparities in PrEP distribution could possibly be factors influencing how people look for information on PrEP.

Our study benefits from several strengths. We use a nationally representative survey to control for several SES covariates, ensuring that the US population is accurately represented. Because our methods adjusted for baseline state level SES characteristics, leaders in each individual state can use these methods to evaluate their state-specific progress. Our internet search volume data is measured in real-time, and while we used annual estimates, it is possible to use the same method to estimate weekly or monthly search volumes. Most importantly, our research presents a new method for surveillance and performance monitoring in HIV prevention.

Our research is not without limitations. Internet search volume data is aggregated and is susceptible to ecological confounding. Additionally, it cannot be used to determine which racial/ethnic, gender, or age groups are or are not engaged with HIV testing or PrEP. Using search data may be subject to selection bias, as not all people access the Internet equally and although some queries may reflect general curiosity rather than treatment-seeking, it is well known that internet search trends mirror many health-related behaviors [18].

Conclusions

Our results are a call-to-action for underperforming states whose populations are not engaged in searching for information on HIV testing and PrEP. Our research provides health officials with an innovative way to monitor statewide interest in PrEP and HIV testing by highlighting the states that demonstrate the least online information seeking, which is critical for the promotion of HIV testing and PrEP as a way to help end the HIV epidemic. Further research should examine why certain states are deficient, and policy makers in deficient states should make efforts to expand HIV testing and PrEP promotion, perhaps by replicating the interventions and policies of better-performing states.

Declarations

Ethics approval and consent to participate Consent for publication

Our research was exempted from an ethics review by the University of California at San Diego Human Research Protections Program.

Abbreviations

AIDS: acquired immunodeficiency syndrome

CDC: Centers for Disease Control and Prevention

CI: confidence interval

HIV: *human immunodeficiency virus*

PrEP: pre-exposure prophylaxis

SES: socioeconomic

Declarations

Ethics approval and consent to participate Consent for publication

Our research was exempted from an ethics review by the University of California at San Diego Human Research Protections Program.

Availability of data and material

All data is publically available through Google Trends and through The Behavioral Risk Factor Surveillance System (*BRFSS*)

Competing Interests

None of the authors declares any conflicts of interest.

Funding

This research was supported by funds from the California HIV/AIDS Research Program Office of the University of California (OS17-SD-001) and the National Institute of Drug Abuse (T32 DA023356, R37 DA019829).

Authors' Contributions (Last name of each author is listed under what they contributed to)

- Study Conception and Design
 - Johnson, Ayers, Nobles, Leas
- Acquisition and Preparation of Data
 - Johnson, Caputi, Liu
- Analysis and Interpretation of Data
 - Johnson, Nobles, Strathdee, Smith
- Drafting of Manuscript
 - Johnson, Nobles, Caputi, Liu, Leas, Strathdee, Smith, Ayers
- Critical Revisions/Revising
 - Johnson, Ayers, Nobles, Leas

Acknowledgements

The content of this research is solely the responsibility of the authors and does not necessarily represent the official views of the California HIV/AIDS Research Program Office or National Institute of Drug Abuse.

References

- 1) Frieden TR, Foti KE, Mermin J. Applying public health principles to the HIV epidemic-how are we doing? *N Engl J Med*. 2015. doi:10.1056/NEJMms1513641
- 2) AIDSvu.com: <https://aidsvu.org/about/>. Accessed March 10th 2020.
- 3) Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? *JAMA - J Am Med Assoc*. 2014. doi:10.1001/jama.2014.1505
- 4) Nobles AL, Leas EC, Latkin CA, Dredze M, Strathdee SA, Ayers JW. #HIV: Alignment of HIV-Related Visual Content on Instagram with Public Health Priorities in the US. *AIDS Behav*. 2020. doi:10.1007/s10461-019-02765-5

- 5) Young SD, Yu W, Wang W. Toward automating HIV identification: Machine learning for rapid identification of HIV-related social media data. *J Acquir Immune Defic Syndr*. 2017. doi:10.1097/QAI.0000000000001240
- 6) Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with web search. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.1005962107
- 7) Ayers JW, Althouse BM, Dredze M, Leas EC, Noar SM. News and internet searches about human immunodeficiency virus after Charlie Sheen's disclosure. *JAMA Intern Med*. 2016. doi:10.1001/jamainternmed.2016.0003
- 8) Allem JP, Leas EC, Caputi TL, et al. The Charlie Sheen Effect on Rapid In-home Human Immunodeficiency Virus Test Sales. *Prev Sci*. 2017. doi:10.1007/s11121-017-0792-2
- 9) Chiu, A. P. Y., Lin, Q., & He, D. (2017). News trends and web search query of HIV/AIDS in Hong Kong. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0185004>
- 10) Mahroum, N., Bragazzi, N. L., Brigo, F., Waknin, R., Sharif, K., Mahagna, H., Amital, H., & Watad, A. (2019). Capturing public interest toward new tools for controlling human immunodeficiency virus (HIV) infection exploiting data from Google Trends. *Health Informatics Journal*. <https://doi.org/10.1177/1460458218766573>
- 11) Department of Health and Human Services, Centers for Disease Control and Prevention. HIV Surveillance Report 2017. <https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>.
- 12) Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2018
- 13) Kaiser Family Foundation. "Status of State Medicaid Expansion Decisions: Interactive Map". <https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/>. Accessed April 1st, 2020
- 14) Patel RR, Mena L, Nunn A, et al. Impact of insurance coverage on utilization of pre-exposure prophylaxis for HIV prevention. *PLoS One*. 2017. doi:10.1371/journal.pone.0178737
- 15) Doblecki-Lewis S, Liu A, Feaster D, et al. Healthcare access and PrEP continuation in San Francisco and Miami after the US PrEP demo project. In: *Journal of Acquired Immune Deficiency Syndromes*. ; 2017. doi:10.1097/QAI.0000000000001236
- 16) Huang YLA, Zhu W, Smith DK, Harris N, Hoover KW. Hiv preexposure prophylaxis, by race and ethnicity – United States, 2014–2016. *Morb Mortal Wkly Rep*. 2018. doi:10.15585/MMWR.MM6741A3

17). AIDSvu.com. Mapping PrEP, First Ever Data on PrEP Users Across the U.S.: <https://aidsvu.org/prep/>. Accessed March 10th 2020.

18) Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with web search. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.1005962107

Figures

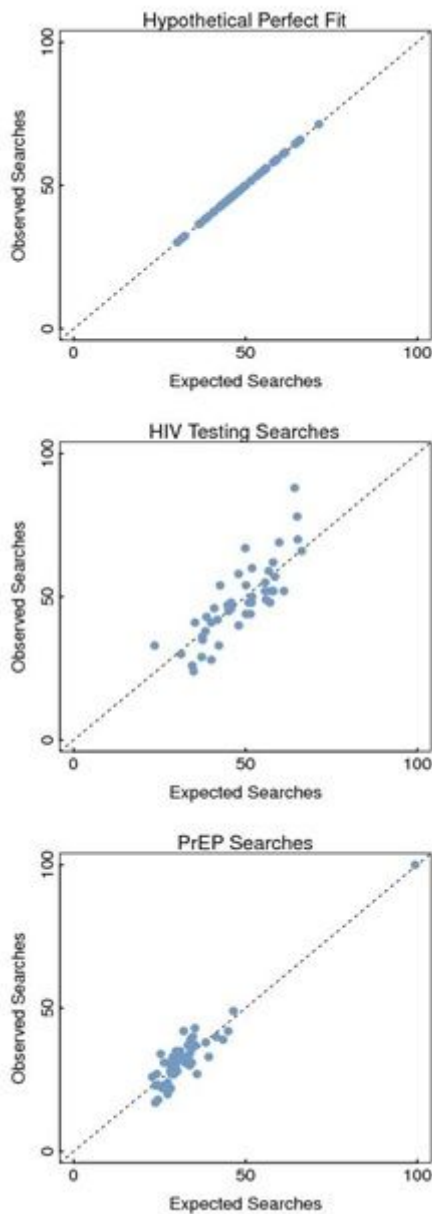


Figure 1

Observed vs. Expected HIV Testing and PrEP Internet Searches as Compared to a Hypothetical Perfectly Fitting Regression Line