

Original Paper

# Searching for HIV and AIDS Health Information in South Africa, 2004-2019: Analysis of Google and Wikipedia Search Trends

Babatunde Okunoye<sup>1,2</sup>, BA; Shaoyang Ning<sup>3</sup>, PhD; Dariusz Jemielniak<sup>4</sup>, PhD

<sup>1</sup>Berkman Klein Centre for Internet and Society, Harvard University, Cambridge, MA, United States

<sup>2</sup>Department of Journalism, Film and Television, University of Johannesburg, Johannesburg, South Africa

<sup>3</sup>Department of Mathematics and Statistics, Williams College, Massachusetts, MA, United States

<sup>4</sup>Management in Networked and Digital Societies Department, Kozminski University, Warsaw, Poland

**Corresponding Author:**

Babatunde Okunoye, BA

Berkman Klein Centre for Internet and Society

Harvard University

2nd Floor

23 Everett Street

Cambridge, MA, 02138

United States

Phone: 1 (617) 495 7547

Email: [bokunoye@cyber.harvard.edu](mailto:bokunoye@cyber.harvard.edu)

## Abstract

**Background:** AIDS, caused by HIV, is a leading cause of mortality in Africa. HIV/AIDS is among the greatest public health challenges confronting health authorities, with South Africa having the greatest prevalence of the disease in the world. There is little research into how Africans meet their health information needs on HIV/AIDS online, and this research gap impacts programming and educational responses to the HIV/AIDS pandemic.

**Objective:** This paper reports on how, in general, interest in the search terms “HIV” and “AIDS” mirrors the increase in people living with HIV and the decline in AIDS cases in South Africa.

**Methods:** Data on search trends for HIV and AIDS for South Africa were found using the search terms “HIV” and “AIDS” (categories: health, web search) on Google Trends. This was compared with data on estimated adults and children living with HIV, and AIDS-related deaths in South Africa, from the Joint United Nations Programme on HIV/AIDS, and also with search interest in the topics “HIV” and “AIDS” on Wikipedia Afrikaans, the most developed local language Wikipedia service in South Africa. Nonparametric statistical tests were conducted to support the trends and associations identified in the data.

**Results:** Google Trends shows a statistically significant decline ( $P < .001$ ) in search interest for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ( $P < .001$ ) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS.

**Conclusions:** This statistically significant ( $P < .001$ ) association between interest in the search terms “HIV” and “AIDS” in South Africa (2004-2019) and the number of people living with HIV and AIDS in the country (2004-2019) might be an indicator that multilateral efforts at combating HIV/AIDS—particularly through awareness raising and behavioral interventions in South Africa—are bearing fruit, and this is not only evident on the ground, but is also reflected in the online information seeking on the HIV/AIDS pandemic. We acknowledge the limitation that in studying the association between Google search interests on HIV/AIDS and cases/deaths, causal relationships should not be drawn due to the limitations of the data.

(*JMIR Form Res* 2022;6(3):e29819) doi: [10.2196/29819](https://doi.org/10.2196/29819)

**KEYWORDS**

HIV/AIDS; web search; big data; public health; Wikipedia; information seeking behavior; online behavior; online health information; Google Trends

## Introduction

A major obstacle to combating the impacts of disease in developing countries is the paucity of high-quality health data, including data regarding people's health information needs [1]. If they do not understand people's everyday concerns, health organizations and policy makers are less able to effectively target education and programming efforts for all genders and age groups [1]. People's information needs and their everyday concerns are often expressed via search engine queries as millions go online to meet their health information needs.

HIV/AIDS remains a major global public health concern, despite concerted efforts aimed at combating the disease. There are an estimated 37.9 (32.7-44.0) million people worldwide living with HIV, the virus that causes AIDS [2]. Sub-Saharan Africa has the greatest burden of HIV/AIDS, with South Africa bearing the greatest burden of the disease in the world. An estimated 7.7 million people were living with HIV and AIDS in South Africa in 2018, with an adult prevalence rate of 20.4% [2]. In 2018, approximately 71,000 South Africans died of AIDS-related causes.

Considerable progress has been made in the fight against HIV/AIDS globally. Globally, an estimated 37.9 (32.7-44.0) million people were living with HIV in 2018 [2], an increase from previous years (eg, 35.3 million in 2012), as more people receive life-saving antiretroviral therapy. At the same time, the number of AIDS-related deaths is also declining, with 770,000 deaths in 2018 [2], down from 1.6 (1.4-1.9) million deaths in 2012 and 2.3 (2.1-2.6) million in 2005 [3]. Gains have also been made toward many of the 2020 Sustainable Development Goals targets and elimination commitments, although significant challenges remain [4]. This is within the context of high-risk behavior among people living with HIV in low- and middle-income countries [5] and insufficient human resources in HIV care in low- and middle-income countries [6].

A key component of the fight against HIV/AIDS has been behavioral interventions [3]. A global meta-analysis of studies determined that behavioral interventions reduce sexual risk behavior and reduce sexually transmitted infections and HIV [3]. Behavioral interventions typically involve harmonizing messages and the dissemination of information about HIV transmission and various prevention approaches using television, radio, outdoor advertising, and information and communications technology. The outcome of all these behavioral interventions is that in sub-Saharan Africa, the percentage of young people (15-24 years) demonstrating a comprehensive and accurate understanding of HIV has generally increased over the years. For example, this understanding rose by 5 percentage points for men and by 3 for women from 2002 to 2011, although knowledge levels remain low (36% for young men and 28% for young women) [3].

In South Africa, there have been a number of nationwide communication campaigns related to raising awareness of HIV and AIDS. Soul City and MTV Shuga are two examples of such multimedia campaigns that promote good sexual health and well-being and have effective outcomes [7]. Research into the impact of the Soul City and other communication campaigns

found that they had a positive effect on the sexual behavior of adults that had been exposed to the campaign message, for instance by bringing about positive changes in condom use and HIV testing [8].

The combined effects of these awareness campaigns have been an increase in knowledge and understanding of HIV/AIDS. One likely outcome of an increase in awareness of HIV/AIDS in South Africa is that a percentage of the population will turn to the internet via search engines for information on HIV/AIDS. Internet penetration in South Africa is 56% (32.9 million) [9].

An increasing number of people are using the internet to support their health care needs [10,11] and search engines such as Google, Yahoo! and Bing are often their first port of call. One of the early studies on search engine use found that 1 in every 28 (3.5%) pages viewed on the web is a search results page, making the use of a search engine the second most popular internet task next to email [12]. Another early study found that there are over 6.75 million health-related internet searches carried out every day worldwide, representing approximately 5% of all internet searches [13]. Search engine use is the most common approach to online information seeking [14], and a study by the Pew Research Center found that half of all internet users now use search engines on a typical day [15]. It has also been estimated that about 1 billion Google searches daily are health-related, which is about 7% of all daily searches [16].

This paper examines trends in online search on HIV and AIDS on Google, which is the dominant search engine worldwide [17] and is the top search engine in most countries with a few exceptions, such as Russia (Yandex) and China (Baidu). In South Africa, Google is also the dominant search engine [17]. Google search data have been used for understanding health information needs, health surveillance, and health forecasting [10,18-23]. Although the use of Google search engine data for health and development in general has been criticized [24,25] in recent years due to an unbalanced and opaque approach that did not satisfy scientific transparency, in recent years, organizations such as the World Bank [26,27] and several government departments around the world have begun revisiting the behavioral big data methods of search engine data in aid of international development. Drawing on feedback from the scientific community on the best practices of big data research [24,25,28,29], particularly paying attention to integrating both big data and "small data" (eg, survey data and other traditional data collection mechanisms) to create more complete and accurate data sets, they have made significant advances in this field. This paper advances the hypothesis that the online search behavior of people looking for health information can be used to understand their health information needs, and also for health surveillance and monitoring [10,18-23].

## Methods

The methods employed in this study were derived from the considerable literature regarding using search data to understand the health information needs of the public and for health surveillance [10,18-23]. The first step involved search term selection. The search terms chosen as proxies to gauge public interest in HIV and AIDS in South Africa were "HIV" and

“AIDS.” These were the only search terms selected given the specificity of the topics concerned. Data [30] on online searches for the terms “HIV” and “AIDS” were thus obtained on Google Trends (categories: health, web search) to visualize how searches were conducted on both search terms. The data were obtained for the years 2004-2019. Some of the most important search queries (top and rising) searched for by South Africans on “HIV” and “AIDS” were downloaded from the Google Trends data set [30] and recorded. Data on regional interest in both search terms for South Africa were downloaded. In addition, the top and rising search queries related to “HIV” and “AIDS” were also accessed via Google Trends.

The nonparametric Mann-Kendall test [31] was conducted to evaluate the significance of trends. The Mann-Kendall test accounts for the autocorrelation of time series (for the alternative hypothesis of monotonic trends) by the design of its test statistic on the time series increments. The Mann-Kendall test was adapted based on the log ratio of search volumes between search terms to evaluate their associations (between “HIV” and “AIDS”) and to investigate the search trend of the term “AIDS” with adjustment (log ratio between “flu” and “AIDS”).

Nonparametric Mann-Kendall tests were conducted to indicate the significant decline in AIDS-related deaths in South Africa and the increasing trend for the number of estimated adults and children living with HIV. Permutation test based on nonparametric Spearman rho [32] was conducted to verify the association between Google search interest for the term “AIDS” and AIDS-related deaths in Africa. Due to the difference in time frequency, monthly Google search data for the term “AIDS” were aggregated through moving average to estimate yearly search interest. All statistical analysis was conducted with R (version 4.1.2; R Foundation for Statistical Computing); Mann-Kendall tests were done using the *trend* package and the Spearman test was performed using the *stats* package.

Pageviews for the topics “HIV” and “AIDS” from 2015-2020 in Wikipedia Afrikaans, the most developed indigenous

language Wikipedia service in South Africa, were obtained from Wikipedia [33]. The year with the earliest data on HIV and AIDS on Wikipedia Afrikaans is 2015.

Review studies on the use of search engine data to understand public interests and trends have noted the “traps in big data” [24,25,28,29], emphasizing that data sources such as search engines are best used to supplement traditional data sources like surveys and that, in general, confidence in such data is strengthened when their findings can be replicated using other data sources and platforms. Thus, this paper also obtained and plotted data [34] on the estimated number of adults and children living with HIV, as well as AIDS-related deaths in South Africa, from the United Nations Joint Programme on HIV/AIDS for the years 2004-2019.

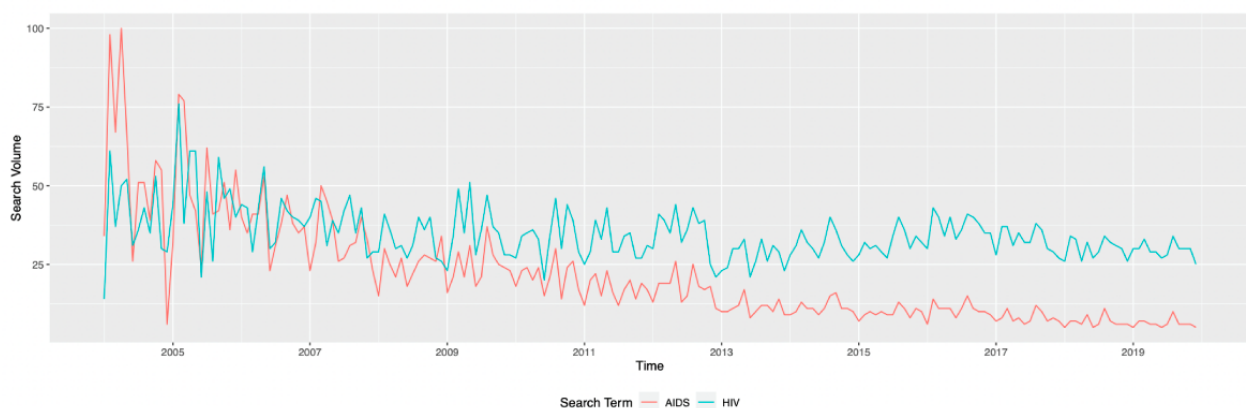
## Results

Search term trends for “HIV” and “AIDS” from 2004-2019 in South Africa showed a decline in “AIDS” searches relative to “HIV” ( $P < .001$ ; Figure 1). This observed online behavior mirrors progress on the ground for both HIV and AIDS ( $P < .001$ , association between “AIDS” searches and deaths), as there was a decrease in AIDS-related deaths ( $P < .001$ ) and an increase in the number of people living with HIV ( $P < .001$ ) in the country from 2004-2019 (Figure 2).

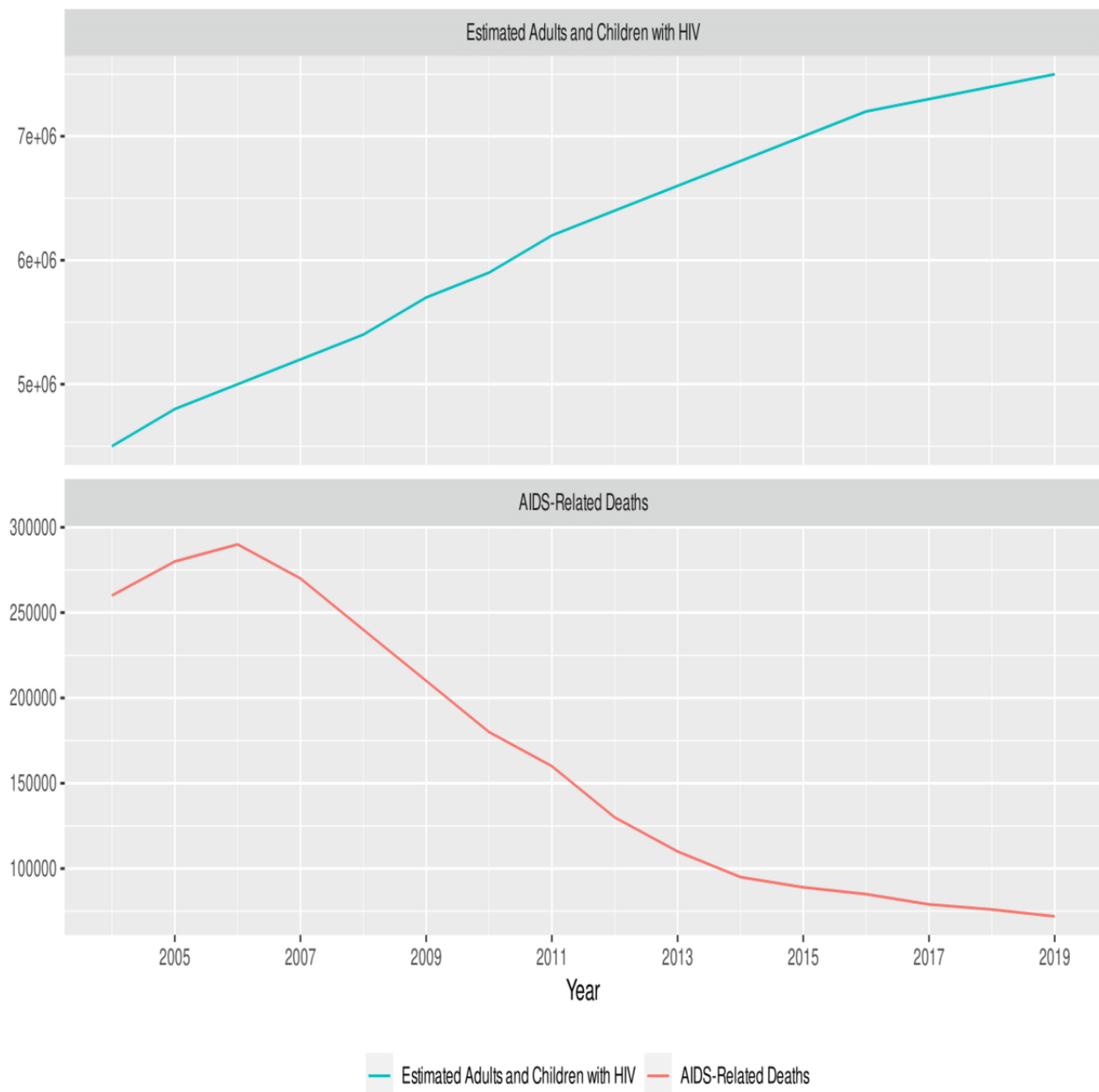
Some of the most important search queries (top and rising) searched for by South Africans on “HIV” and “AIDS” can be seen in Multimedia Appendix 1 and Multimedia Appendix 2, respectively. A complete visualization of the search data, regional interest in South Africa for the search terms “HIV” and “AIDS,” and rising and top search queries on “HIV” and “AIDS” can be accessed via Google Trends [30].

Similarly, on Wikipedia Afrikaans, for the years 2015-2020, there was greater interest in “HIV” than “AIDS” (Multimedia Appendix 3).

**Figure 1.** Search interest in the search terms “HIV” and “AIDS” (category: health, web search) in South Africa, 2004-2019 (accessed December 24, 2020).



**Figure 2.** UNAIDS data on estimated adults and children living with HIV and AIDS-related deaths in South Africa [34]. UNAIDS: Joint United Nations Programme on HIV/AIDS.



## Discussion

### Principal Findings

This paper shows a statistically significant decline ( $P < .001$ ) in search interest on Google for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ( $P < .001$ ) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS. The data on the search terms “HIV” and “AIDS” in South Africa (2004–2019) reveal a decrease in interest over time in AIDS-related search queries relative to HIV-related search queries. This mirrors progress on the ground in South Africa, where AIDS-related deaths have reduced over the years (2004–2019) and people are living longer with HIV due to life-changing medications. Similarly, behavioral

interventions through messages delivered through media sources in South Africa, including the internet, have increased the number of people with an accurate understanding of the differences between HIV and AIDS [3]. In response to these behavioral interventions, a proportion of the populations in these countries will use the internet via search engines to seek information on HIV and AIDS. From the earliest data available on Wikipedia Afrikaans (2015–2020), there has been a greater interest in HIV than AIDS in South Africa.

### Comparison to Prior Work

This work contributes to the growing literature on infodemiology, which seeks to employ search engine data in the service of health information seeking, forecasting, and surveillance [35]. Previous research has covered a wide variety of infectious diseases such as influenza [36], dengue [37],

norovirus [38], COVID-19 [39], and sexually transmitted infections [40], as well as other topics like cancer [22,23], dementia [18], and mental health [20]. Among the limited literature exploring HIV/AIDS with internet search data, the methodology is often restricted to qualitative [41] or parametric methodology [42], with limited research on Africa. Our work contributes to the limited literature (eg, [1,43]) on HIV/AIDS in the African context. Moreover, the nonparametric methodology we adopted requires minimal assumptions about the data and its distribution, and is thus robust and adaptive. Our work contributes to the niche of infodemiological study on HIV/AIDS in Africa and South Africa in particular while possessing the potential to extend to other countries/regions or geographical resolutions.

### Further Directions

An important follow-up to this study might be to understand how South Africans search for health information on other diseases and topics of public interest on search engines. In addition, motivated by our findings in this work, we may further investigate and develop a statistically principled and coherent framework to use Google search data for forecasting HIV/AIDS trends. In the meantime, a quantitative (survey) and qualitative (interviews) study of how South Africans access HIV/AIDS information online is planned, to provide insights for the research question by using traditional research methods. This study aims to provide data across South African provinces, disaggregated by age and gender. Additionally, broadening the scope of future studies could be achieved by studying other search engines, social media, and knowledge repositories, as research shows that, for example, Wikipedia is a good reflection of medical interests of the lay population [44], search engine queries are a good way of estimating outbreaks [45], and social media platforms are an important source of medical information today [46].

### Limitations

This paper reports on how, in general, interest in the search terms “HIV” and “AIDS” mirrors the increase in people living with HIV and the decline in AIDS cases in South Africa. Here, we report some limitations of the study.

First, there is an acknowledgement that the population of health information seekers on HIV/AIDS online may be quite different from the offline population. For instance, not everyone searching for information on HIV/AIDS may be connected online or use search engines, and the number of people connected to the internet changed over the years of the study (2004-2019). Hence, this paper is not an exact mapping of the online behavior of all the people searching for HIV/AIDS information in South Africa. This limitation is common in research using online big data methods such as search engine data, where the nature of the specific platform from which the data were obtained shapes the results obtained [25]. To mitigate this, a follow-up quantitative

and qualitative study of the HIV/AIDS health information seeking behavior of South Africans using Google is planned. This would provide data through traditional (non-big data) research methods, following best practices from digital big data research, where the best results are often obtained when traditional data collection methods like surveys are combined with big data methods to inform research inquiry [25].

Second, we also acknowledge that the observed trends in the Google search data could be explained by the general decrease in people’s interest in such diseases and public health matters. To adjust for this, we used the search term “flu” as a control and the declining trend of the search term “AIDS” remained significant after adjustment ( $P < .001$ ). In addition, this paper acknowledges the effects of media coverage on interest in specific terms, which is not accounted for in this paper.

Third, this paper focuses only on studying the association between the Google search interest in HIV/AIDS and cases/deaths. We acknowledge that causal relationships should not be drawn due to the limitations of the data. In addition, we acknowledge that although we used the Google Trends category “health, web search” to refine the search results for “AIDS,” the returned results may contain other meanings or associations of the word “AIDS,” as seen in some entries in [Multimedia Appendix 2](#).

### Conclusions

A major obstacle to combating the impacts of disease in developing countries is the paucity of high-quality health data, particularly for understanding people’s health information needs [1]. People’s information needs and their everyday concerns are often expressed via search engine queries as millions go online to meet their health information needs. This paper shows a statistically significant decline ( $P < .001$ ) in search interest on Google for AIDS relative to HIV in South Africa. This trend mirrors progress on the ground in South Africa and is significantly associated ( $P < .001$ ) with a decline in AIDS-related deaths and people living longer with HIV. This trend was also replicated on Wikipedia Afrikaans, where there was a greater interest in HIV than AIDS.

In developing country contexts where high-quality data on the health information needs of people is often lacking, understanding the population’s use of search engines to meet health information needs can provide useful data. Investigating people’s use of search engines to meet information needs can also reveal if education and programming efforts have been effective. Consequently, a natural progression of this study might be a quantitative and qualitative study of the use of search engines and other online information sources to meet HIV- and AIDS-related health information needs in South Africa, which might mitigate platform-specific limitations of findings as outlined in the Limitations section.

### Acknowledgments

BO conceived, designed, and carried out the study. SN performed statistical analysis and reporting on data, including preparation of figures. DJ provided Wikipedia data and critical feedback on the manuscript. All authors contributed to writing the manuscript and read, reviewed, and revised the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Top and rising related queries for HIV, 2004-2019.

[\[DOCX File , 14 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Top and rising related queries for AIDS, 2004-2019.

[\[DOCX File , 13 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Wikipedia Afrikaans pageview statistics on HIV and AIDS 2015-2020 (accessed November 5, 2020).

[\[DOCX File , 151 KB-Multimedia Appendix 3\]](#)

## References

1. Abebe R, Hill S, Vaughan J, Small P, Schwartz H. Using Search Queries to Understand Health Information Needs in Africa. In: Proceedings of the Thirteenth International AAAI Conference on Web and Social Media. Burnaby: PKP Publishing Services Network; 2019 Presented at: Thirteenth International AAAI Conference on Web and Social Media. Association for the Advancement of Artificial Intelligence; June 11-14, 2019; Munich, Germany URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/3360/3228>
2. UNAIDS data 2019. UNAIDS. URL: <https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data> [accessed 2020-10-31]
3. 2012 UNAIDS Report on the Global AIDS Epidemic. UNAIDS. URL: [https://www.unaids.org/en/resources/documents/2012/20121120\\_UNAIDS\\_Global\\_Report\\_2012](https://www.unaids.org/en/resources/documents/2012/20121120_UNAIDS_Global_Report_2012) [accessed 2020-10-31]
4. The Sustainable Development Goals and the HIV response. UNAIDS. 2017. URL: [https://www.unaids.org/sites/default/files/media\\_asset/SDGsandHIV\\_en.pdf](https://www.unaids.org/sites/default/files/media_asset/SDGsandHIV_en.pdf) [accessed 2020-10-31]
5. Nguyen PT, Gilmour S, Le PM, Nguyen TT, Tanuma J, Nguyen HV. Factors associated with high-risk behaviors of people newly diagnosed with HIV/AIDS: results from a cross-sectional study in Vietnam. *AIDS Care* 2021 May;33(5):607-615. [doi: [10.1080/09540121.2020.1761519](https://doi.org/10.1080/09540121.2020.1761519)] [Medline: [32397740](https://pubmed.ncbi.nlm.nih.gov/32397740/)]
6. Nguyen PT, Nguyen HV, Phan HT, Cao CT. Situation of human resources and staff training of four provincial HIV/AIDS control centers in Northern Vietnam and some related factors in 2013. *VPJM* 2015;7(167):75-82.
7. HIV and AIDS in South Africa. Avert. 2020. URL: <https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/south-africa> [accessed 2022-01-20]
8. Peltzer K, Parker W, Mabaso M, Makonko E, Zuma K, Ramlagan S. Impact of national HIV and AIDS communication campaigns in South Africa to reduce HIV risk behaviour. *ScientificWorldJournal* 2012;2012:384608-384606 [FREE Full text] [doi: [10.1100/2012/384608](https://doi.org/10.1100/2012/384608)] [Medline: [23213285](https://pubmed.ncbi.nlm.nih.gov/23213285/)]
9. Individuals using the Internet 2000-2018. International Telecommunications Union. 2019. URL: [https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2019/Individuals\\_Internet\\_2000-2018\\_Dec2019.xls](https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2019/Individuals_Internet_2000-2018_Dec2019.xls) [accessed 2020-10-31]
10. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are Parents Getting it Right? A Survey of Parents' Internet Use for Children's Health Care Information. *Interact J Med Res* 2015 Jun 22;4(2):e12 [FREE Full text] [doi: [10.2196/ijmr.3790](https://doi.org/10.2196/ijmr.3790)] [Medline: [26099207](https://pubmed.ncbi.nlm.nih.gov/26099207/)]
11. Massey PM. Where Do U.S. Adults Who Do Not Use the Internet Get Health Information? Examining Digital Health Information Disparities From 2008 to 2013. *J Health Commun* 2016;21(1):118-124. [doi: [10.1080/10810730.2015.1058444](https://doi.org/10.1080/10810730.2015.1058444)] [Medline: [26166484](https://pubmed.ncbi.nlm.nih.gov/26166484/)]
12. Jansen BJ, Pooch U. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology* 2001;52(3):235-246. [doi: [10.1002/1097-4571\(2000\)9999:9999::AID-ASI1607>3.0.CO;2-F](https://doi.org/10.1002/1097-4571(2000)9999:9999::AID-ASI1607>3.0.CO;2-F)]
13. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. *AMIA Annu Symp Proc* 2003:225-229 [FREE Full text] [Medline: [14728167](https://pubmed.ncbi.nlm.nih.gov/14728167/)]
14. Morris M, Teevan J, Panovich K. A comparison of Information seeking on search engines and social networks. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Menlo Park, California: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26; Washington DC, USA p. 1-4 URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/icwsm10.pdf>

15. Fallows D. Search Engine Use. Pew Research Center. 2008. URL: <https://www.pewresearch.org/internet/2008/08/06/search-engine-use/> [accessed 2020-10-31]
16. Murphy M. Dr Google will see you now: Search giant wants to cash in on your medical queries. The Telegraph. 2019. URL: <https://www.telegraph.co.uk/technology/2019/03/10/google-sifting-one-billion-health-questions-day/> [accessed 2021-05-27]
17. Search Engine Market Share. StatCounter Global Stats. URL: <https://gs.statcounter.com/search-engine-market-share> [accessed 2020-10-30]
18. Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. *J Med Internet Res* 2015;17(11):e264 [FREE Full text] [doi: [10.2196/jmir.4516](https://doi.org/10.2196/jmir.4516)] [Medline: [26586281](https://pubmed.ncbi.nlm.nih.gov/26586281/)]
19. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. *JMIR Public Health Surveill* 2016 Oct 12;2(2):e156 [FREE Full text] [doi: [10.2196/publichealth.6504](https://doi.org/10.2196/publichealth.6504)] [Medline: [27733330](https://pubmed.ncbi.nlm.nih.gov/27733330/)]
20. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. *Am J Prev Med* 2013 May;44(5):520-525. [doi: [10.1016/j.amepre.2013.01.012](https://doi.org/10.1016/j.amepre.2013.01.012)] [Medline: [23597817](https://pubmed.ncbi.nlm.nih.gov/23597817/)]
21. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. *Australas Med J* 2014;7(1):24-28 [FREE Full text] [doi: [10.4066/AMJ.2014.1900](https://doi.org/10.4066/AMJ.2014.1900)] [Medline: [24567763](https://pubmed.ncbi.nlm.nih.gov/24567763/)]
22. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. *J Med Internet Res* 2018 Jan 08;20(1):e6 [FREE Full text] [doi: [10.2196/jmir.8870](https://doi.org/10.2196/jmir.8870)] [Medline: [29311051](https://pubmed.ncbi.nlm.nih.gov/29311051/)]
23. Xu C, Yang H, Sun L, Cao X, Hou Y, Cai Q, et al. Detecting Lung Cancer Trends by Leveraging Real-World and Internet-Based Data: Infodemiology Study. *J Med Internet Res* 2020 Mar 12;22(3):e16184 [FREE Full text] [doi: [10.2196/16184](https://doi.org/10.2196/16184)] [Medline: [32163035](https://pubmed.ncbi.nlm.nih.gov/32163035/)]
24. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014 Mar 14;343(6176):1203-1205. [doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506)] [Medline: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)]
25. Ruths D, Pfeffer J. Social sciences. Social media for large studies of behavior. *Science* 2014 Nov 28;346(6213):1063-1064. [doi: [10.1126/science.346.6213.1063](https://doi.org/10.1126/science.346.6213.1063)] [Medline: [25430759](https://pubmed.ncbi.nlm.nih.gov/25430759/)]
26. Reyes G. Leveraging behavioral insights in the age of big data. *World Bank Blogs*. 2017. URL: [https://blogs.worldbank.org/latinamerica/leveraging-behavioral-insights-age-big-data?cid=SHR\\_BlogSiteShare\\_EN\\_EXT](https://blogs.worldbank.org/latinamerica/leveraging-behavioral-insights-age-big-data?cid=SHR_BlogSiteShare_EN_EXT) [accessed 2020-10-31]
27. Afif Z, Islan W, Calvo-Gonzalez O. Behavioral science around the world: Profiles of 10 countries. World Bank. 2018. URL: <https://documents1.worldbank.org/curated/en/710771543609067500/pdf/132610-REVISED-00-COUNTRY-PROFILES-dig.pdf> [accessed 2020-10-31]
28. Butler D. When Google got flu wrong. *Nature* 2013 Feb 14;494(7436):155-156. [doi: [10.1038/494155a](https://doi.org/10.1038/494155a)] [Medline: [23407515](https://pubmed.ncbi.nlm.nih.gov/23407515/)]
29. Mills KA. What are the threats and potentials of big data for qualitative research? *Qualitative Research* 2017 Nov 30;18(6):591-603. [doi: [10.1177/1468794117743465](https://doi.org/10.1177/1468794117743465)]
30. Compare 'aids' and 'hiv'. *Google Trends*. URL: <https://trends.google.com/trends/explore?date=2004-01-01%202019-12-31&geo=ZA&q=aids,hiv> [accessed 2020-10-31]
31. Hipel K, McLeod A. *Time Series Modelling of Water Resources and Environmental Systems*, 1st edition. New York: Elsevier Science; 1994.
32. Hollander M, Wolfe D, Chicken E. *Nonparametric statistical methods*, 2nd edition. Hoboken, New Jersey: John Wiley & Sons; 2013.
33. Wikipedia Afrikaans Pageviews Analysis on HIV and AIDS 2015-2020. *Wikipedia*. URL: <https://pageviews.toolforge.org/?project=af.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2015-07&end=2020-10&pages=HIV|AIDS> [accessed 2020-11-05]
34. HIV estimates with uncertainty bounds 1990-Present. UNAIDS. URL: [https://www.unaids.org/en/resources/documents/2020/HIV\\_estimates\\_with\\_uncertainty\\_bounds\\_1990-present](https://www.unaids.org/en/resources/documents/2020/HIV_estimates_with_uncertainty_bounds_1990-present) [accessed 2020-10-31]
35. Mavragani A, Ochoa G, Tzagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *J Med Internet Res* 2018 Nov 06;20(11):e270 [FREE Full text] [doi: [10.2196/jmir.9366](https://doi.org/10.2196/jmir.9366)] [Medline: [30401664](https://pubmed.ncbi.nlm.nih.gov/30401664/)]
36. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
37. Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E. Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study. *J Med Internet Res* 2020 Jul 24;22(7):e17633 [FREE Full text] [doi: [10.2196/17633](https://doi.org/10.2196/17633)] [Medline: [32706682](https://pubmed.ncbi.nlm.nih.gov/32706682/)]
38. Yuan K, Huang G, Wang L, Wang T, Liu W, Jiang H, et al. Predicting Norovirus in the United States Using Google Trends: Infodemiology Study. *J Med Internet Res* 2021 Sep 29;23(9):e24554 [FREE Full text] [doi: [10.2196/24554](https://doi.org/10.2196/24554)] [Medline: [34586079](https://pubmed.ncbi.nlm.nih.gov/34586079/)]

39. An L, Russell DM, Mihalcea R, Bacon E, Huffman S, Resnicow K. Online Search Behavior Related to COVID-19 Vaccines: Infodemiology Study. *JMIR Infodemiology* 2021 Nov 12;1(1):e32127 [[FREE Full text](#)] [doi: [10.2196/32127](https://doi.org/10.2196/32127)] [Medline: [34841200](https://pubmed.ncbi.nlm.nih.gov/34841200/)]
40. Johnson AK, Bhaumik R, Tabidze I, Mehta SD. Nowcasting Sexually Transmitted Infections in Chicago: Predictive Modeling and Evaluation Study Using Google Trends. *JMIR Public Health Surveill* 2020 Nov 05;6(4):e20588 [[FREE Full text](#)] [doi: [10.2196/20588](https://doi.org/10.2196/20588)] [Medline: [33151162](https://pubmed.ncbi.nlm.nih.gov/33151162/)]
41. Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. *PLoS One* 2017 Sep 18;12(9):e0185004 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0185004](https://doi.org/10.1371/journal.pone.0185004)] [Medline: [28922376](https://pubmed.ncbi.nlm.nih.gov/28922376/)]
42. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. *J Big Data* 2018 May 19;5(1):1-21. [doi: [10.1186/s40537-018-0126-7](https://doi.org/10.1186/s40537-018-0126-7)]
43. Al-Rawi AK. Exploring HIV/AIDS interests in the MENA region using Internet based searches. *First Monday*. URL: <http://hdl.handle.net/1765/102387> [accessed 2020-02-23]
44. Chrzanowski J, Sotek J, Fendler W, Jemielniak D. Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis. *J Med Internet Res* 2021 Apr 12;23(4):e26331 [[FREE Full text](#)] [doi: [10.2196/26331](https://doi.org/10.2196/26331)] [Medline: [33667176](https://pubmed.ncbi.nlm.nih.gov/33667176/)]
45. Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. *J Med Internet Res* 2014 Dec 16;16(12):e289 [[FREE Full text](#)] [doi: [10.2196/jmir.3680](https://doi.org/10.2196/jmir.3680)] [Medline: [25517353](https://pubmed.ncbi.nlm.nih.gov/25517353/)]
46. Zhang D, Shi Z, Hu H, Han GK. Classification of the Use of Online Health Information Channels and Variation in Motivations for Channel Selection: Cross-sectional Survey. *J Med Internet Res* 2021 Mar 09;23(3):e24945 [[FREE Full text](#)] [doi: [10.2196/24945](https://doi.org/10.2196/24945)] [Medline: [33687342](https://pubmed.ncbi.nlm.nih.gov/33687342/)]

*Edited by T Leung; submitted 21.04.21; peer-reviewed by S Yang, P Nguyen; comments to author 17.05.21; revised version received 29.05.21; accepted 04.02.22; published 11.03.22*

*Please cite as:*

Okunoye B, Ning S, Jemielniak D

*Searching for HIV and AIDS Health Information in South Africa, 2004-2019: Analysis of Google and Wikipedia Search Trends*

*JMIR Form Res* 2022;6(3):e29819

URL: <https://formative.jmir.org/2022/3/e29819>

doi: [10.2196/29819](https://doi.org/10.2196/29819)

PMID:

©Babatunde Okunoye, Shaoyang Ning, Dariusz Jemielniak. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 11.03.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.