Original Paper

Internet Search Patterns of Human Immunodeficiency Virus and the Digital Divide in the Russian Federation: Infoveillance Study

Andrey Zheluk[1], BAppSc, GradDip(History), MBA, GradDip(IT); Casey Quinn[2], BCom, MPhil, PhD; Daniel Hercz[3], BSc, MSc Biostat; James A Gillespie[1], BA(Hons), PhD

[1]Menzies Centre for Health Policy, The University of Sydney, University of Sydney NSW, Australia
[2]PRMA Consulting Ltd, New York, NY, United States
[3]University of Sydney, Sydney, Australia

Corresponding Author:
Andrey Zheluk, BAppSc, GradDip(History), MBA, GradDip(IT)
Menzies Centre for Health Policy
The University of Sydney
D02 Victor Coppleson Building
University of Sydney NSW, 2006
Australia
Phone: 61 2 9351 2818
Fax: 61 2 9351 5204
Email: andreyzheluk@gmail.com

Related Article:
This is a corrected version. See correction statement in: http://www.jmir.org/2013/11/e267/

Abstract

Background: Human immunodeficiency virus (HIV) is a serious health problem in the Russian Federation. However, the true scale of HIV in Russia has long been the subject of considerable debate. Using digital surveillance to monitor diseases has become increasingly popular in high income countries. But Internet users may not be representative of overall populations, and the characteristics of the Internet-using population cannot be directly ascertained from search pattern data. This exploratory infoveillance study examined if Internet search patterns can be used for disease surveillance in a large middle-income country with a dispersed population.

Objective: This study had two main objectives: (1) to validate Internet search patterns against national HIV prevalence data, and (2) to investigate the relationship between search patterns and the determinants of Internet access.

Methods: We first assessed whether online surveillance is a valid and reliable method for monitoring HIV in the Russian Federation. Yandex and Google both provided tools to study search patterns in the Russian Federation. We evaluated the relationship between both Yandex and Google aggregated search patterns and HIV prevalence in 2011 at national and regional tiers. Second, we analyzed the determinants of Internet access to determine the extent to which they explained regional variations in searches for the Russian terms for "HIV" and "AIDS". We sought to extend understanding of the characteristics of Internet searching populations by data matching the determinants of Internet access (age, education, income, broadband access price, and urbanization ratios) and searches for the term "HIV" using principal component analysis (PCA).

Results: We found generally strong correlations between HIV prevalence and searches for the terms "HIV" and "AIDS". National correlations for Yandex searches for "HIV" were very strongly correlated with HIV prevalence (Spearman rank-order coefficient [rs]=.881, $P\leq.001$) and strongly correlated for "AIDS" ($rs=.714$, $P\leq.001$). The strength of correlations varied across Russian regions. National correlations in Google for the term "HIV" ($rs=.672$, $P=.004$) and "AIDS" ($rs=.584$, $P\leq.001$) were weaker than for Yandex. Second, we examined the relationship between the determinants of Internet access and search patterns for the term "HIV" across Russia using PCA. At the national level, we found Principal Component 1 loadings, including age (-0.56), HIV search (-0.533), and education (-0.479) contributed 32% of the variance. Principal Component 2 contributed 22% of national variance (income, -0.652 and broadband price, -0.460).

Conclusions: This study contributes to the methodological literature on search patterns in public health. Based on our preliminary research, we suggest that PCA may be used to evaluate the relationship between the determinants of Internet access and searches

XSL•FO
RenderX

for health problems beyond high-income countries. We believe it is in middle-income countries that search methods can make the greatest contribution to public health.

Introduction

Search Patterns in Health

Internet search patterns provide a low-cost, rapidly accessible data source for a range of health problems. Search patterns have been described as behavioral measures of an issue's importance to individuals [1]. If individual Internet users are concerned or interested in an issue, they are more likely to search for information related to that issue. The relative importance of an issue among populations of Internet users can thus be inferred from the volume of searches for a term or terms representing that issue. Since 2006, researchers have used search patterns to study a wide range of health problems, notably influenza [2-5], as well as undocumented adverse drug interactions [6,7], suicide-related information [8], and HIV (human immunodeficiency virus) [9]. Despite the widespread use of search patterns, researchers commonly suggest Internet users may not be representative of the entire population. Specific concerns include differences in access based on age [10], income and education [11], and gender [12]. This means the social, economic, and demographic status of Internet users may not fully reflect those of the population as a whole.

Determinants of Internet Access

The propensity to access the Internet varies between socioeconomic and demographic cohorts. The strongest determinants of Internet access are income and education. This finding is consistent in studies from the United States [13] and the European Union [14] and across middle-income countries [15]. Additionally, gender [16], English-language ability [17], broadband access price [18], urban location [19], ethnicity [20], and age [21] have also been reported as determinants of Internet use in both high- and middle-income countries. In summary, Internet users in both high- and middle-income countries are more likely to have higher incomes and higher levels of education.

The Digital Divide—Access and Use

Access to the Internet is an economic development policy issue. Telecommunications networks, including the Internet, are regarded as a catalyst for economic growth [22]. Since the early 2000s, the term "digital divide" has been widely used to describe differences in Internet access and use across socioeconomic gradients within and between countries [23]. In 2011, Hilbert reviewed international policy responses to the digital divide [24]. In his review, Hilbert proposes four classes of variables with which to analyze the digital divide. These classes are the unit of analysis (eg, individual, country), determinants of access (eg, income, education), the kind of technology (eg, cell phones, fixed broadband), and how individuals connect (ie, access vs effective use). Others have similarly argued that access to

infrastructure inadequately describes the digital divide [25]. Basing their arguments on Roger's theory of diffusion of innovations, these authors suggest analysis of the digital divide should focus on effective use, incorporating technical competence, and individuals' adaptation of technology to meet their personal needs rather than access alone.

Use of the Internet for Health Information Seeking Online

The Internet is widely used for health information seeking in high-income countries. A 2013 study found 59% of all US adults searched for health information online, with 77% of these starting at search engines such as Google [26]. Equally, there is a general scholarly consensus that a digital divide applies to online health seeking behavior [27]. In 2006, Rice described the limited research into health-related Internet use across economic and demographic gradients in the United States [28]. More recent studies European [29] and US [30] studies suggest that income and education are the most important determinants of seeking health information online.

Search Patterns and Effective Use

Although a digital divide may exist, determining the sociodemographic profile of Internet users from search results is not straightforward. Aggregated Google search queries are the most commonly used data source for search studies but carry no demographic or economic information. In the case of disease surveillance, this means that groups with a significant disease burden, such as older or economically disadvantaged people without Internet access may be excluded from search results [31]. By contrast, health information seeking research is generally based on qualitative research and statistical surveys. This research generally includes demographic characteristics and covers issues such as health literacy [32] and behaviors following access to health information [33]. In summary, researchers have widely investigated the effective use of online health information in high-income countries. It is this research that provides the empirical foundation for a rich analysis of the relationship between health information seeking across economic and demographic gradients and patterns of online search.

Chronic Illness and Internet Use

Individuals with chronic health problems and disabilities are more likely to search for health information online. Online information seeking among people with chronic and terminal diseases has been widely researched [34,35]. Cancer information seeking in particular has attracted considerable research interest due to its diversity, duration, and treatment complexity [36]. The management of HIV as a chronic illness has similarly attracted scholarly interest. Studies suggest PLHIV (people living with HIV) use the Internet extensively for health

information. A 2006 US study found that 66% of PLHIV participants searched for health information at least half the time they were online [37]. Furthermore, PLHIV Internet users were more likely to be better educated, have higher incomes, exhibit greater knowledge of HIV disease processes, and adhere to medication [38,39]. In summary, while income and education are the most important determinants of health-related Internet use, individuals with chronic diseases may have a stronger incentive to use the Internet effectively.

Online Health Information Seeking in Middle-Income Countries

While research is limited, online health information seeking also appears to be important in middle-income countries. In 2011, the international health insurer Bupa surveyed online health information seeking among Internet users in 12 high- and middle-income countries [40]. The researchers found higher rates of health information seeking in middle income countries (China 94%, Thailand 93%, and Saudi Arabia 91%) than in high-income countries (Australia 77%, United Kingdom 70%, and Spain 71%). Similarly, a 2010 Bupa study found 95% of Russian Internet users sought advice on health, medicines, or medical conditions online [41]. Bupa researchers attributed the high rates of online health information seeking in middle-income countries to the high cost of medical consultations and concerns over service quality. While not peer reviewed, these Bupa surveys point to a particularly important role for health-related searches outside of high-income countries. Conversely, these studies investigated only the propensity to access health information among Internet users, leaving aside international comparisons of how effectively online health information is used across social and economic gradients. The relationship between the need for health information and access to the Internet was not investigated.

Search Studies in Middle-Income Countries

As recently as 2009, researchers suggested that Google Trends was unsuitable for disease surveillance outside of developed countries due to insufficient Internet access [42]. However, the rapid increase of Internet use in middle-income countries suggests otherwise. Internet use is forecast to grow considerably more quickly by 2015 in middle-income than high-income countries (see Table 1; [43]). Since 2009, studies from Southeast Asia [44], Latin America [45], Russia [46], and China [47] suggest that search pattern studies are increasingly regarded as valid and reliable methods of disease surveillance in middle-income countries.

The potential of search patterns to improve public health surveillance in middle-income countries is well documented. First, online surveillance offers immediate insights into the present status of disease. That is, online surveillance may "predict the present" [48] without the reporting lags associated with complicated reporting procedures in public health bureaucracies [44]. Second, online surveillance may overcome the weaknesses of traditional surveillance systems, such as poor sensitivity to new diseases [49] and the lack of skills and equipment required for early disease detection [50]. Third, searches may overcome underreporting gaps from the private sector and from individuals who do not seek formal medical care [51]. Fourth, online surveillance may improve transparency. Central or regional governments may wish to minimize reports of disease outbreaks that could affect tourism or political reputation [52] or what sensitive issues surveys may not reveal [53,54]. In summary, online surveillance has the potential to improve disease surveillance in populations bearing the greatest burden of disease.

Consistent with the aims of infodemiology, our exploratory study examined "the science of distribution and determinants of information...(on) the Internet, or in a population, with the ultimate aim to inform public health and public policy" [55]. We examined the relationship between Internet search patterns, disease prevalence, and the determinants of Internet access using the case of HIV in a middle-income country. Through investigating these relationships, we aimed to develop methods to complement traditional HIV surveillance in Russia and contribute to the science of health-related searches.

Table 1. Changes in Internet use in selected middle- and high-income countries (values indicate penetration in %, ie, number of users divided by population).

Country	2009 actual Internet use	2015 predicted Internet use
China	28	47
India	7	19
Brazil	33	74
Russia	31	55
Indonesia	12	37
United States	70	73
Japan	74	81

Methods

Summary

This exploratory study sought to determine if search methods can be used for disease surveillance in a large middle-income country with a dispersed population. We first assessed whether online surveillance is a valid and reliable method for monitoring HIV in the Russian Federation. Second, we analyzed the determinants of Internet access to determine the extent that they explain regional variations in searches for the Russian terms for "HIV" and "AIDS".

Google and Yandex Searches in Russia

Most search pattern studies have used Google Trends (or the defunct Google Insights for Search) as the data source. Google Trends has been deployed in studies of influenza [3], dengue [11], and HIV [9]. However, the structure of the Russian-language Internet market is unique. Whereas Google provided 84% of global Internet search queries in May 2011 [56], Google's market share in Russia was only 25% in 2010/2011 [57]. The largest search provider in Russia in 2011 was Yandex, with 60% market share. In 2011, Russia overtook Germany as the European country with the highest number of unique visitors online [58]. Russian Internet users grew from 43% of the population in 2010 to 55% in 2012 [59]. In Russia, Yandex is a strong commercial competitor of Google.

The publicly available Google Trends data for Russia has several limitations. First, Google does not provide complete results, returning only subregions with the highest search volume. Google data were available for only 16 of Russia's 89 subregions for the term "HIV" and 29 for the term "AIDS" during 2011. Second, Google does not provide raw search data. This makes direct comparisons between subregions and matching with variables representing Internet access determinants complex. We used WordStat as the primary data source, as Yandex made publicly available a complete raw search dataset for all Russian regions and subregions for the full 12 months of 2011. We used Google Trends as a secondary source of aggregated search results for validation purposes.

Case Study: Why is Search-Based HIV Surveillance Important in Russia?

HIV is a serious health problem in the Russian Federation. Russia has the highest cumulative number of PLHIV of any European country, largely concentrated among people who inject drugs (PWID). On December 31, 2011, there were 650,100 PLHIV registered in Russia [60]. However, the true scale of HIV in Russia has long been the subject of considerable debate [61,62]. Feshbach and colleagues' 2005 study compiled data from official and unofficial Russian sources, as well as international agencies, to assess the quality of Russian HIV statistics [63]. The authors suggested that official Russian HIV data are frequently inconsistent, diverge markedly from alternative sources such as UNAIDS (the Joint United Nations Programme on HIV/AIDS), and present major methodological obstacles. The authors concluded that official Russian estimates of HIV prevalence were understated by a multiple of three to five times. Similar findings emerged from a 2007 UNODC (United Nations Office on Drugs and Crime) report that evaluated national data collection mechanisms related to HIV among PWIDs in nine lower income countries including Russia [64].

HIV Surveillance and Hidden Populations

HIV surveillance is further complicated by Russian drug laws, police, medical, and public attitudes. Most international observers regard Russian drug laws as punitive, unsupported by scientific evidence, and ineffective [65]. A 2010 study into police behavior found widespread reports of extrajudicial policing practices, including extortion, torture, and rape of PWIDs [66]. Attitudes among medical staff too are generally negative towards PLHIV [67,68]. Public opinion is also generally negative towards individuals acquiring HIV sexually or through drug use [69]. As a consequence of professional and social attitudes, many PLHIV avoid contact with medical organizations and avoid testing for HIV. Literature suggests there are disincentives for Russian PLHIV accessing health information directly from health professionals.

International researchers generally regard Russian HIV-positive PWIDs as a population hidden from public health surveillance. Since the early 2000s, researchers have sought to improve population estimates and document the conditions experienced by Russian PWIDs living with HIV [70,71]. Traditional surveys and sampling methods among PWIDs are unreliable, as individuals may not report accurately on stigmatized and illegal behaviors. A 2011 study in Russia found that, among 193 HIV-positive participants, only 36% were aware of their HIV status [72]. Another study of HIV-positive Russian PWIDs found persistent high-risk behaviors associated with HIV transmission [73]. Among study participants, 25% had been refused access to medical care, 18% were refused employment or fired, and 6% were forced from family homes. Researchers found 39% of participants had probable clinical depression, and 37% had anxiety levels comparable to psychiatric inpatients. In summary, there is considerable evidence that Russia has large numbers of PLHIV, many of whom are likely to be alienated from the formal health system and be absent from official statistics. The high rates of Internet searches for health information, combined with stigmatization of HIV, suggest that the Internet may be an important resource for PLHIV in Russia.

Russian injecting drug users have generally avoided contact with the formal health system. Between 2004 and 2011, much of the contact with injecting drug users and other groups at high risk of HIV was conducted by donor-funded Russian non-governmental organizations (NGOs). The behavioral surveillance data collected by these NGOs also contributed to Russian national HIV reporting to UNAIDS. However, as the result of government pressures, the number of donor-funded harm reduction NGO projects in Russia decreased from 70 in 2007 to 20 in 2011 [74]. The decrease in NGOs may also have eroded the capacity for data collection from populations at risk of HIV. In 2012, Russia did not report any HIV behavioral surveillance data associated with injecting drug use and sex work [75]. In summary, the progressive dismantling of harm reduction projects in Russia means only surveillance data from individuals formally diagnosed with HIV in government clinics are available. Injecting drug workers, sex workers, and others at risk of HIV have disappeared from Russian government reporting.

RQ1 Method: Is Search Surveillance a Valid Method for Monitoring HIV in Russia?

To answer this research question, we examined the relationship between HIV prevalence across the Russian Federation and Internet searches for the terms "HIV" and "AIDS". First, we obtained HIV prevalence data for each region and subregion from the Russian Federal AIDS Centre [60]. We chose 2011 data as this was latest complete dataset available. The Russian

Federal AIDS Centre publishes the most timely and comprehensive HIV dataset available. However, these data are limited to formally diagnosed PLHIV and likely exclude many individuals at risk of HIV, or of uncertain serostatus, who deliberately avoid contact with government health services.

Second, we selected two terms to represent HIV searches. These two search terms were "HIV" (VICh in Russian) and "AIDS" (SPID). We referred to the Google Trends related-terms feature [76] to ensure each term referred to the subject of this study. In the case of "HIV", all terms were related to HIV, whereas the term "AIDS" revealed several unrelated terms (Table 2). For example, the second most popular term associated with "AIDS" referred to the computer game "need for speed". Based on these results, we anticipated that the search term "HIV" would have a stronger positive correlation with HIV prevalence than the term "AIDS".

Third, we aggregated Yandex searches for each month of 2011 to produce a single annual search figure for the terms "HIV" and "AIDS" for each of Russia's 89 subregions (see the map in Multimedia Appendix 1; [77]) covering the data range January 1 to December 31, 2011. In Russian federal statistical compilations, several smaller subregions are routinely aggregated, producing 83 rather than 89 statistical subregions. In our calculations, we used population prevalence of HIV and searches rather than raw figures. This allowed comparison across regions and subregions. We used this single 2011 annual HIV search in our further calculations.

Fourth, we conducted Spearman correlations of per-capita Yandex monthly searches for the terms "HIV" and "AIDS" against HIV prevalence data for all Russian subregions in 2011. We repeated this process with each of eight Russian regions. This provided us with national and regional correlations between search and prevalence data for 2011.

Fifth, we obtained all available Google data for the terms "HIV" and "AIDS" for 2011 and repeated this analysis for validation purposes. Google search data for the term "HIV" were available for 16 regions and for "AIDS" for 29 regions. We then conducted Spearman correlations between Google and Yandex data for validation purposes.

Table 2. Google Trends—Related terms for HIV and AIDS in the Russian Federation in 2011.

Search related terms	Russian	Value
"HIV"		
symptoms HIV	симптомы вич	100
symptoms	симптомы	100
AIDS	спид	65
AIDS HIV	спид вич	65
HIV infection	вич инфекции	35
HIV signs	вич признаки	35
analysis for HIV	анализ на вич	35
HIV infection	вич инфекция	30
HIV dating	вич знакомства	25
HIV photo	вич фото	20
"AIDS"		
test AIDS	тест спид	100
need for speed	нид фор спид	75
AIDS HIV	спид вич	55
HIV	вич	55
AIDS info	спид инфо	50
AIDS centre	спид центр	45
AIDS symptoms	спид симптомы	25
AIDS photo	спид фото	25
AIDS test	спид тест	25
speed hack	спид хак	20

RQ2: What is the Relationship Between the Determinants of Internet Access and Searches for the Term "HIV" Across Russia?

The relationship between Internet search patterns for specific health problems and the prevalence of these problems in populations is now well established. However, Internet users may not be representative of overall populations. Further, the characteristics of the Internet using population cannot be directly ascertained from search pattern data. We sought to extend understanding of the characteristics of Internet searching populations through data matching the determinants of Internet access (ie, age, income, broadband access price, and urban to rural ratios) with search patterns through multivariate analysis.

Several studies have examined the socioeconomic factors associated with HIV prevalence and injecting drug use in Russia. Moran et al investigated the relative importance of several variables in influencing HIV prevalence in a cross-sectional study based solely on Russian federal government statistics [78]. The authors found urbanization, mobility, crime, and income growth associated with HIV prevalence. In 2011, researchers surveyed 711 PWIDs in two large provincial cities [79]. The researchers concluded PWIDs were typical Russians when compared with a random population. However, investigators drew their random sample from 2004 household survey data. While Russian per capita income grew from US $9800 in 2004 to US $17,000 in 2011 [80], the authors did not comment on this potentially important confounder. These two studies illustrate the logistic difficulties of obtaining timely, valid, and independent data in Russia.

Our Methodology

We examined the relationship between spatial patterns of online searches for the term "HIV" and the determinants of Internet access. We used data from RQ1 in our analysis. In RQ1, we demonstrated the relationship between HIV prevalence and searches for "HIV". While this relationship was generally strong, differences in search patterns across regions may reflect differences in the determinants of Internet access as well as differences in HIV rates.

We selected principal component analysis (PCA) to explore the relationship between the determinants of Internet access and searches for the term "HIV". PCA is a method of multivariate analysis for finding patterns in data rather than hypothesis testing. PCA aids in the interpretation of relationships in the original data by transforming the original variables into a new set of variables, the principal components [81,82]. PCA has been widely used in public health to study relationships of health problems to socioeconomic variables. For example, PCA has been used to investigate European tumor prevalence [83], nutritional epidemiology in Greece [84], and epidemiological analysis in low- and middle-income countries [85]. As a consequence, we considered PCA an appropriate method for this exploratory study.

First, we collated the data sources. We obtained search pattern data for the term "HIV" through PCA. We obtained Russian-language data for five determinants of Internet access (see Table 3) from the Russian federal statistics agency for 83 Russian statistical regions [86]. The determinants of Internet access comprised each a single figure for each subregion for 2011. In compiling our data, we sought to most closely align search pattern, HIV prevalence, and determinants of Internet access data.

Second, we conducted PCA on all Russian subregions to produce a national level analysis. We included the determinants of Internet access and per capita search for "HIV" for all subregions. We used a correlation matrix approach to standardize the variables, as we used different units with differing variances. Based on our review of Internet determinants, we anticipated that the variables we chose to analyze would correlate.

Third, we conducted separate PCAs to examine the relationship between HIV search patterns and the determinants of Internet access separately on each of the eight Russian regions. Previous research suggests that, while there is no minimum of variables and cases in PCA, a larger number is preferable [87]. In designing this study, we purposely selected a smaller number of variables. We did this to permit analysis of both national data, as well as of regions with smaller number of subregions. Through analyzing both national and regional PCA separately, we anticipated we would identify additional spatial relationships not obvious at the national level.

Table 3. Determinants of Internet access—List of variables in PCA.

Variable	Determinant of Internet access (abbreviation)
Variable 1	Higher education students per 100,000 population (age)
Variable 2	Percentage aged 25-64 with higher education (education)
Variable 3	Gross regional product per capita (income)
Variable 4	Broadband price per month (Bband price)
Variable 5	Urban / rural population (urbanization)
Variable 6	Searches for HIV per 100,000 population during 2011 (search)

Results

Correlations

We first investigated search surveillance as a valid method for monitoring HIV in Russia. We found generally strong correlations between HIV prevalence and searches for the terms "HIV" and "AIDS". Yandex searches for "HIV" were very strongly correlated with HIV prevalence (Spearman rank-order coefficient [rs]=.881, P≤.001), whereas "AIDS" was strongly correlated nationally (rs=.714, P≤.001) (see Table 4). The strength of correlations varied across Russian regions. Several regions were less strongly correlated in Yandex. For example, HIV prevalence and searches in the central and northwestern regions were moderately correlated as a result of outlier data points. Further, Google national searches for the term "HIV" were moderately correlated (rs=.672, P=.004) with HIV prevalence and weakly correlated with Yandex searches for "HIV" (rs=.584, P≤.001) (see Table 5).

Second, we examined the relationship between the determinants of Internet access and search patterns for the term "HIV" across Russia. We found considerable variation in the relationship between these determinants and search patterns. We first analyzed national PCA results (Table 6). We determined the number of components to analyze using the Kaiser, Scree, and cumulative variance methods [82]. Kaiser and scree tests suggested three principal components (PCs), and the cumulative variance method suggested four PCs should be analyzed. In PC1 (the first and most important component), HIV search, age, and educational variables were moderately correlated. In PC2, per capita income was most important. This factor was weakly correlated with searches for HIV and explained 23% of the variance. The subsequent two components, which explain less variance, are more difficult to interpret.

Table 4. HIV and AIDS correlations from Yandex—National and all federal regions of Russian Federation.

Region	HIV prevalence per 100,000 pop'n	Searches for "HIV" per 1000 pop'n	Spearman correlation for HIV prevalence & "HIV" (2-tailed P value)	Searches for "AIDS" per 1000 pop'n	Spearman correlation for HIV prevalence & term "AIDS" (2-tailed P value)
National	446.513	16.995	.881 (≤.001)	19.312	.714 (≤.001)
Central	279.2	21.832	.377 (.006)[a]	21.215	-.123 (.386)
Northwestern	586.6	23.619	.482 (≤.001)[a]	26.383	.209 (.137)
Southern	144.3	8.397	.486 (≤.001)	12.665	.486 (≤.001)
North Caucuses	58.8	2.758	-.179 (.206)	6.666	-.286 (.040)
Volga	437.8	17.322	.793 (≤.001)	20.366	.380 (.005)
Urals	805	24.037	.657 (≤.001)	19.379	.429 (≤.001)
Siberian	528.1	13.962	.804 (≤.001)	15.561	.503 (≤.001)
Far east	166.4	7.473	.017 (.907)	11.197	.083 (.557)

[a]Outliers.

Table 5. Google Trends search results.

	AIDS searches	HIV searches
Number of regions	29	16
Spearman correlation, HIV prevalence (2-tailed P value)	.584 (P≤.001)	.672 (P=.004)
Spearman correlation, Google with Yandex (2-tailed P value)	-.289 (P=.129)	.223 (P=.406)

XSL•FO
RenderX

Table 6. National PCA results.

Importance of components	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1.386	1.172	0.989	0.854	0.737	0.674
Proportion of variance	0.320	0.229	0.163	0.121	0.091	0.076
Cumulative proportion	0.320	0.549	0.712	0.834	0.924	1.000
Loadings						
Age	-0.560	–	0.105	0.269	0.762	0.121
Education	-0.479	-0.376	0.313	0.247	-0.345	-0.593
Income	0.215	-0.652	–	-0.491	0.460	-0.268
Bband price	0.329	-0.460	0.523	0.422	–	0.478
Urbanization	0.131	-0.417	-0.777	0.396	-0.110	0.190
Search	-0.533	-0.200	–	-0.539	-0.278	0.546

Biplots and Spatial Relationships

We used biplots to explain spatial relationships in our PCA results. Biplots provide a visual representation of PCA data from the first two PCs [88]. Biplots allow identification of clusters of subregions with similar characteristics. Further, the clustering of subregions along vector lines serves to highlight subregions more strongly associated with specific variables. Importantly, the clustering of subregions is subjective and requires additional analysis. On the national HIV search biplot (see Figure 1), PC1 was associated with Vector 3 (income), Vector 4 (broadband price), and Vector 5 (urbanization). PC2 was associated with Vector 1 (age), Vector 2 (education), and Vector 6 (HIV search). We obtained PCA results and biplots for all eight Russian regions. See Multimedia Appendix 2 for a list of subregions referenced in the national PCA. See Multimedia Appendices 3-5 for biplot results for each Russian federal region. Finally, we conducted a separate PCA for HIV prevalence data. We substituted the variable HIV search with HIV prevalence. The results of a PCA incorporating the variable HIV prevalence produced results with a similar form to those incorporating HIV searches at both the national and regional levels.

Table 7. Summary of national biplots.

Region	Relationship	Geographic clusters	Outliers
National	PC1: V3, V4, V5 Income, broadband fees, urbanization	Cluster 1 37. Ingushetia 41. Chechnya 79. Amursk 81. Sakhalin 83. Chukhotka	18. Moscow City 22. Nenets Autonomous Region 29. St Petersburg 61. Yamalo-Nenets
		Cluster 2 10. Moscow region 71. Kemerovo 26. Murmansk 78. Khabarovsk 59. Tyumen	
	PC2: V1, V2, V6 Age, education, and HIV prevalence/ search	Cluster 3 18. Moscow city 24. Kaliningrad 29. St Petersburg City 46. Tatarstan 54. Samara 58. Sverdlovsk 62. Chelyabinsk 70. Irkutsk 76. Kamchatka 77. Primorsk 80. Magadan	

XSL•FO
RenderX

Figure 1. National HIV search and HIV prevalence biplots.
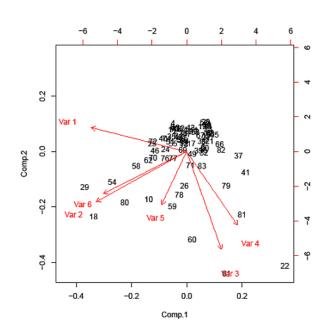

Discussion

Principal Findings

Overall, we found search patterns were a valid method of HIV surveillance in the Russian Federation. Furthermore, our research suggests that search patterns for HIV are generally not related to income or broadband price. However across Russian regions, we found considerable variation in the strength of correlations between search and disease prevalence, and the determinants of Internet access. Finally, our analysis suggested that the strong correlations between search and disease prevalence may indicate effective use of the Internet by individuals at risk of HIV and PLHIV.

RQ1: Is Search Surveillance a Valid Method for Monitoring HIV in Russia?

We found online search patterns for HIV were correlated with HIV prevalence in both Google and Yandex at the national level. It is noteworthy that the latest official Russian HIV data available at the time of writing in mid-2013 were for the year 2011. By contrast, Yandex search data were available with a delay of 4 weeks and Google data with a 48-hour delay. This timely availability illustrates the potential contribution of search pattern data to disease surveillance.

Second, we found considerable variation in the strength of correlations among regions in Yandex data. Overall, we found Yandex searches for the term "HIV" and HIV prevalence were most strongly correlated. This suggests PLHIV are more likely to search for "HIV" than "AIDS". In the North Caucuses and far eastern regions, HIV prevalence was not positively correlated with search. We attribute this to the low HIV prevalence and low search volumes in these regions. By contrast, in the central and northwestern regions, search volumes and HIV prevalence were high, but correlations were moderate. We attributed the weaker correlations to outliers in Yandex data. Removing the central (Moscow subregion) and northwestern region (Leningrad

subregion), outliers strengthened correlations from 0.377 to 0.551 and 0.482 to 0.939 respectively. This suggests correlation analysis should routinely account for outlying subregions.

Third, we found Google data are not adequate for subnational HIV surveillance in Russia. We attribute the low correlations to the multiple zero values present in our Google dataset. Of the 15 regions for which Google data were available, many months recorded a zero search value for the term "HIV". These zero values were consistent with an earlier study [46] of the use of Google search for health policy analysis in Russia that found Google Trends requires an unknown threshold before results are displayed. While national level Google data were correlated with HIV prevalence, our analysis suggests it should not be used for regional analysis.

Finally, our results contribute to understanding of hidden populations of PLHIV in Russia. There is a general consensus that Russian HIV rates are underreported. Previous studies have reported considerable at-risk populations unaware of their HIV status in subregions with high HIV prevalence (eg, [74]). However, we found strong spatial correlations between official HIV rates and searches for HIV. This finding has several interpretations. First, the spatial variation in search results also appears in traditional surveillance. Researchers have found high populations of unknown HIV serostatus in subregions of high HIV prevalence. That is, additional searches for HIV related information by populations at risk of HIV and unknown serostatus may inflate already high search volumes in those subregions with high HIV prevalence. However, these additional searches would not change the overall spatial distribution of search patterns. A second interpretation relates to the search data used. Our analysis relied on annually aggregated search results. Our results are thus a static view of HIV prevalence over a 12-month period. This static view does not capture longitudinal anomalies in search patterns. While this 12-month snapshot was appropriate for the purposes of this study,

monitoring of weekly and monthly search patterns may produce different results and reveal spatial variations in searches.

RQ2: What is the Relationship Between the Determinants of Internet Access and Search Patterns for the Term "HIV" Across Russia?

We analyzed national and regional PCA results separately. First, we examined the two national level biplots. One biplot incorporated HIV prevalence as a variable and the other searches for "HIV". All other variables remained consistent (see Figure 1). We found these two biplots to be near isomorphic.

The separate national biplots containing both HIV prevalence and HIV search produced three logically coherent geographical clusters. National cluster 1 was characterized by low-income, non-ethnic Russian subregions with low HIV prevalence. The exception to this is the Sakhalin subregion, with high per capita income. This clustering occurred along the broadband vector (V4), suggesting high broadband prices and limited access. National cluster 2 was associated with the urbanization vector (V5). It includes urbanized non-metropolitan areas. National cluster 3 included the Russian cities with the highest prevalence of HIV, along the HIV search/prevalence (V6) and education vectors (V2). In addition, it included the Magadan subregion. The Magadan subregion is highly urbanized but has a low HIV prevalence and low population. We attribute the inclusion of Magadan in this cluster as an indicator of potential HIV risk. Conversely, the isolation of Magadan, in northeastern Russia, away from borders and drug routes, suggests a lower risk of HIV transmission through injecting drug use.

Both national PCA biplots featured several outlier subregions (see Figure 1). For example, the Yamalo Nenets subregion (61) was an outlier. This is an oil-producing subregion, with very high per capita incomes and below average HIV rates. It was strongly associated with the income vector (V3). Second, Moscow and St Petersburg were outliers. We attribute these cities' outlier positions to a statistical anomaly. Each city had a rural to urban ratio of zero and very high Internet access rates and incomes. In summary, national level PCA analysis of both the HIV prevalence and HIV search biplots suggested a stronger relationship between broadband access prices in several subregions. See Multimedia Appendix 4 for a discussion of PCA in Russian regions.

In summary, PCA is not a technique that establishes causal relationships. However, based on our preliminary analysis, we suggest that income and broadband prices do not generally appear to be associated with HIV searches, either positively or negatively, in the subregions of highest HIV prevalence. Further research, in the form of confirmatory factor analysis and regression analysis is needed to establish this relationship statistically. Contingent upon the results of this additional analysis, HIV search pattern data may be incorporated into HIV modeling.

Our findings extend beyond an examination of the digital divide in Russia as defined by access to the Internet. There is also a behavioral dimension implicit in our two research questions. Search patterns measure aggregate behavior at the population level, with important issues more frequently searched. Searches

for the term "HIV" measure the importance of this disease in a population. Consequently, the generally strong correlations between search patterns and disease prevalence lead us to infer that the Internet is being used effectively by PLHIV. That is, searchers for "HIV" demonstrate the technical competence to search for health information they consider important. However, this is a cautious conclusion, and one that merits further research.

Further Research and Limitations

Our research suggests that further exploratory analysis applying search pattern methods to HIV surveillance in Russia is warranted. First, PWID and sex worker populations may be at increased risk of HIV as the result of the Russian government's censorship of prevention, treatment, and care information [89,90] and decreased behavioral monitoring capacity among internationally funded NGOs. Further, in 2013, concerns emerged about the capacity of independent Russian social research organizations to continue unencumbered data collection [91]. Search methods may present a partial solution to these emerging information constraints. Internet search patterns provide a valid near real time measure of health behaviors in the field at population level.

Second, additional research is required to establish how effectively Russians use the Internet for HIV and health information. Qualitative and survey research among populations at risk of HIV and PLHIV will assist the further development of search surveillance methods and the planning of online interventions. Research in Russia should also examine the quality of health information available to PLHIV, both through domestic and international Russian language websites.

Third, organizations working with at-risk populations and PLHIV may consider initiating studies that establish baseline measures of search patterns for HIV and related diseases. From these baselines, longitudinal studies will be able to rapidly identify unanticipated shifts in spatial and temporal patterns of HIV-related searches and HIV prevalence, well in advance of official incidence and prevalence data.

Fourth, the method described in this paper can be extended to other communicable and non-communicable diseases in Russian-speaking countries. Broader application of this method may require initial disease-by-disease and country-by-country validation. However, even without validation, this method provides a low-cost, rapid, timely initial assessment with which to shape further planning, analysis, and decision making.

Finally, our research had several limitations. First, we were constrained by the absence of time series data. To conduct data matching for the PCA, we used a single aggregate figure to represent total searches nationally and within each subregion. We believe this analysis would be strengthened by a month-to-month comparison of HIV prevalence data in each Russian region. Such data are not publicly available. Second, Google is the only data source in most middle income countries. This limits the application of this method. An important exception is China, where Baidu is increasingly being used alongside Google for disease surveillance.

Conclusions

The use of data for disease surveillance has been widely promoted in popular literature. Under the rubric of "big data", journalists have popularized the novel application of Internet search patterns in medicine [92]. Scholars too, have speculated that data availability will lead to the evolution of new models of disease surveillance [93-95]. While the potential application of large scale data analysis in health care has generated considerable popular and scholarly interest, most research has focused on high-income countries with well-functioning public health information systems. We believe it is in middle-income countries that search methods can make the greatest contribution to public health. It is in these countries that traditional surveillance systems are least developed and health data least available.

Clearly, a digital divide between rich and poor countries persists. However, Internet access in middle-income countries is growing rapidly, and online health information is in demonstrably high demand. Based on our preliminary research, we are cautiously optimistic in suggesting that access to the Internet should therefore not be considered a constraint to conducting search studies beyond high-income countries. It is in lower income countries that search pattern surveillance may move beyond a statistical novelty and be incorporated into local health data collection and decision making.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Map of Russian Federation.

[PNG File, 546KB-Multimedia Appendix 1]

Multimedia Appendix 2

Reference list of Russian regions and subregions for PCA biplots.

[PDF File (Adobe PDF File), 63KB-Multimedia Appendix 2]

Multimedia Appendix 3

Narrative analysis of PCA in Russian regions.

[PDF File (Adobe PDF File), 22KB-Multimedia Appendix 3]

Multimedia Appendix 4

PCA biplots - Russian regional HIV search and HIV prevalence.

[PDF File (Adobe PDF File), 477KB-Multimedia Appendix 4]

Multimedia Appendix 5

Table summarizing PCA biplot results in Russian regions.

[PDF File (Adobe PDF File), 32KB-Multimedia Appendix 5]

References

1. Granka L. Inferring the Public Agenda from Implicit Query Data. Boston: ACM SIGIR Conference; 2009. URL: http://ceur-ws.org/Vol-512/paper07.pdf [accessed 2013-11-05] [WebCite Cache ID 6KtuAa2j2]
2. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
3. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
4. 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] [Medline: 19020500]
5. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012;14(5):e125 [FREE Full text] [doi: 10.2196/jmir.2102] [Medline: 23037553]
6. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013 May 1;20(3):404-408. [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]

XSL•FO
RenderX

7.   Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]

8.   Wong PW, Fu KW, Yau RS, Ma HH, Law YW, Chang SS, et al. Accessing suicide-related information on the internet: a retrospective observational study of search behavior. J Med Internet Res 2013;15(1):e3 [FREE Full text] [doi: 10.2196/jmir.2181] [Medline: 23305632]

9.   Jena AB, Karaca-Mandic P, Weaver L, Seabury SA. Predicting new diagnoses of HIV infection using internet search engine data. Clin Infect Dis 2013 May;56(9):1352-1353. [doi: 10.1093/cid/cit022] [Medline: 23334812]

10. Fox S, Duggan M. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2013. Health Online 2013 URL: http://www.pewinternet.org/~/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2013-10-17] [WebCite Cache ID 6KQX6Gwdl]

11. Ayers JW, Althouse BM, Allem JP, 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] [Medline: 23597817]

12. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. BMC Cancer 2011;11:442 [FREE Full text] [doi: 10.1186/1471-2407-11-442] [Medline: 21993136]

13. Peslak A. First Monday. 2004 Mar 01. An analysis of regional and demographic differences in United States Internet usage URL: http://firstmonday.org/ojs/index.php/fm/article/view/1124 [accessed 2013-11-05] [WebCite Cache ID 6KtlONfCd]

14. Vicente MR, López AJ. Some empirical evidence on Internet diffusion in the New Member States and Candidate Countries of the European Union. Applied Economics Letters 2008 Oct 24;15(13):1015-1018. [doi: 10.1080/13504850600972352]

15. Chinn MD, Fairlie RW. ICT use in the developing world: an analysis of differences in computer and internet penetration. Review of International Economics 2010;18(1):153-167. [doi: 10.1111/j.1467-9396.2009.00861.x]

16. Li N, Kirkup G. Gender and cultural differences in Internet use: A study of China and the UK. Computers & Education 2007 Feb;48(2):301-317. [doi: 10.1016/j.compedu.2005.01.007]

17. Gil-Garcia JR, Helbig NC, Ferro E. Is it only about Internet access? An empirical test of a multi-dimensional Digital Divide. In: EGOV 2006 proceedings. Berlin: Springer; 2006 Presented at: Electronic Government 5th International Conference; Sept. 4-8, 2006; Krakow, Poland. [doi: 10.1007/11823100_13]

18. Chaudhuri A, Flamm KS, Horrigan J. An analysis of the determinants of internet access. Telecommunications Policy 2005 Oct;29(9-10):731-755. [doi: 10.1016/j.telpol.2005.07.001]

19. Demoussis M, Giannakopoulos N. Facets of the digital divide in Europe: Determination and extent of internet use. Economics of Innovation and New Technology 2006 Apr;15(3):235-246. [doi: 10.1080/10438590500216016]

20. Horrigan JB. Federal Communications Commission. 2010. Broadband adoption and use in America URL: http://transition.fcc.gov/DiversityFAC/032410/consumer-survey-horrigan.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGvdN15]

21. Lera-López F, Billon M, Gil M. Determinants of Internet use in Spain. Economics of Innovation and New Technology 2011 Mar;20(2):127-152. [doi: 10.1080/10438590903378017]

22. Qiang CZW, Rossotto CM, Kimura K. Economic impacts of broadband. In: Information and Communications for Development. Washington, DC: World Bank; 2009:35-50.

23. OECD. 2001. Understanding the Digital Divide URL: http://www.oecd.org/internet/ieconomy/1888451.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGpjtkV]

24. Hilbert M. The end justifies the definition: The manifold outlooks on the digital divide and their practical usefulness for policy-making. Telecommunications Policy 2011 Sep;35(8):715-736. [doi: 10.1016/j.telpol.2011.06.012]

25. Hamel JY. UNDP Human Development Research Paper. New York: UNDP; 2010. ICT4D and the human development and capabilities approach: the potentials of information and communication technology URL: http://hdr.undp.org/es/informes/mundial/idh2010/trabajos/HDRP_2010_37.pdf [accessed 2013-08-30] [WebCite Cache ID 6JFiv5EWO]

26. Fox S, Duggan M. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2013. Health Online 2013 URL: http://www.pewinternet.org/Reports/2013/Health-online.aspx [accessed 2013-08-14] [WebCite Cache ID 6IrGmeWvy]

27. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]

28. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]

29. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]

30. Fox S, Purcell K. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2010. Chronic disease and the Internet URL: http://web.pewinternet.org/~/media/Files/Reports/2010/PIP_Chronic_Disease_with_topline.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGj0jbz]

31. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. Urology 2013 Jan;81(1):37-42. [doi: 10.1016/j.urology.2011.04.024] [Medline: 21676450]

XSL•FO
RenderX

32. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]

33. Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information seeking. J Health Commun 2008 Mar;13(2):181-199 [FREE Full text] [doi: 10.1080/10810730701854110] [Medline: 18300068]

34. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]

35. Fox S. Pew Internet and American Life Project. Washington, DC: Pew Research Center; 2008. The engaged e-patient population URL: http://nicolaziady.com/wp-content/uploads/2011/07/PIP_Health_Aug08.pdf.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGdiDcm]

36. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]

37. Kalichman SC, Cherry C, Cain D, Weinhardt LS, Benotsch E, Pope H, et al. Health information on the Internet and people living with HIV/AIDS: information evaluation and coping styles. Health Psychol 2006 Mar;25(2):205-210. [doi: 10.1037/0278-6133.25.2.205] [Medline: 16569112]

38. Kalichman SC, Benotsch EG, Weinhardt LS, Austin J, Luke W. Internet use among people living with HIV/AIDS: association of health information, health behaviors, and health status. AIDS Educ Prev 2002 Feb;14(1):51-61. [Medline: 11900110]

39. Samal L, Saha S, Chander G, Korthuis PT, Sharma RK, Sharp V, et al. Internet health information seeking behavior and antiretroviral adherence in persons living with HIV/AIDS. AIDS Patient Care STDS 2011 Jul;25(7):445-449 [FREE Full text] [doi: 10.1089/apc.2011.0027] [Medline: 21682586]

40. BUPA. Bupa health pulse 2011: Global trends, attitudes and influences. London: BUPA; 2011. URL: http://www.bupa.com/media/288798/bupa_health_pulse_report_2011.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGWUag4]

41. McDaid D, Park AL. Online health: untangling the web. London: BUPA; 2010. URL: http://www.bupa.com/media/44806/online_20health_20-_20untangling_20the_20web.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGTnXXS]

42. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]

43. Aguiar M, Boutenko V, Michael D, Rastogi V, Subramanian A, Zhou Y. The Internet's New Billion: Digital Consumers in Brazil, Russia, India, China, and Indonesia. Boston: Boston Consulting Group; 2010. URL: http://www.bcg.com/documents/file58645.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrGR8Yom]

44. 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 Negl Trop Dis 2011 May;5(5):e1206 [FREE Full text] [doi: 10.1371/journal.pntd.0001206] [Medline: 21647308]

45. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]

46. Zheluk A, Gillespie JA, Quinn C. Searching for truth: internet search patterns as a method of investigating online responses to a Russian illicit drug policy debate. J Med Internet Res 2012;14(6):e165 [FREE Full text] [doi: 10.2196/jmir.2270] [Medline: 23238600]

47. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]

48. Choi H, Varian H. Predicting the present with google trends. Economic Record 2012 Jun;88(s1):2-9. [doi: 10.1111/j.1475-4932.2012.00809.x]

49. Chunara R, Freifeld CC, Brownstein JS. New technologies for reporting real-time emergent infections. Parasitology 2012 Dec;139(14):1843-1851 [FREE Full text] [doi: 10.1017/S0031182012000923] [Medline: 22894823]

50. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]

51. Khan K, McNabb SJ, Memish ZA, Eckhardt R, Hu W, Kossowsky D, et al. Infectious disease surveillance and modelling across geographic frontiers and scientific specialties. Lancet Infect Dis 2012 Mar;12(3):222-230. [doi: 10.1016/S1473-3099(11)70313-9] [Medline: 22252149]

52. Madoff LC, Fisman DN, Kass-Hout T. A new approach to monitoring dengue activity. PLoS Negl Trop Dis 2011 May;5(5):e1215 [FREE Full text] [doi: 10.1371/journal.pntd.0001215] [Medline: 21647309]

53. Curioso WH, Kurth AE. Access, use and perceptions regarding Internet, cell phones and PDAs as a means for health promotion for people living with HIV in Peru. BMC Med Inform Decis Mak 2007;7:24 [FREE Full text] [doi: 10.1186/1472-6947-7-24] [Medline: 17850656]

54. Stephens-Davidowitz S. The cost of racial animus on a black presidential candidate: using Google search data to find what surveys miss. SSRN Journal 2012:1-55. [doi: 10.2139/ssrn.2238851]

55. 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. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

56. Netmarketshare.com. Aliso Viejo, CA: Net Applications; 2012 Jun 15. Search Engine Market Share URL: http://www. netmarketshare.com/report.aspx?qprid=4&qptimeframe=M&qpsp=148 [accessed 2012-06-16]

57. Crisp T. ComScore Voices. Reston, VA: comscore.com; 2011. Yandex: From Russia with Love URL: http://blog. comscore.com/2011/06/yandex_russia_with_love.html [accessed 2013-08-14] [WebCite Cache ID 6IrGCavjr]

58. Comscore Press Releases. London, UK; 2011 Nov 14. comScore Releases Overview of European Internet Usage in September 2011 URL: http://www.comscore.com/Insights/Press_Releases/2011/11/ comScore_Releases_Overview_of_European_Internet_Usage_in_September_2011 [accessed 2013-11-05] [WebCite Cache ID 6KtrskSAe]

59. WCIOM. Russ Public Opin Res Cent WCIOM. Moscow; 2012 Feb 13. Russians networking: rating of social media URL: http://wciom.com/index.php?id=61&uid=63 [accessed 2012-04-05] [WebCite Cache ID 6BTiwJdNa]

60. Pokrovsky V, Ladnaya N, Sokolova E, Buravtsova E. HIV infection - Information Bulletin No 36. Moscow: Federal Scientific-methodological center for the prevention of AIDS; 2012. URL: http://www.hivrussia.ru/files/bul_36.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrOXIJ7V]

61. Naskhoev M, Sergeyev B. AIDS in the Commonwealth of Independent States. Chisinau: UNAIDS; 2008. URL: http://aids. md/aids/files/1012/CIS%20MAP%20Report_English%20Final%204%20print.doc [accessed 2013-11-06] [WebCite Cache ID 6KtneBHLq]

62. Maier CB, Martin-Moreno JM. Quo vadis SANEPID? A cross-country analysis of public health reforms in 10 post-Soviet states. Health Policy 2011 Sep;102(1):18-25. [doi: 10.1016/j.healthpol.2010.08.025] [Medline: 20864203]

63. Feshbach M, Galvin CM. Wash DC Woodrow Wilson Int Cent Sch. 2005. HIV/AIDS in Russia an analysis of statistics URL: https://docs.google.com/file/d/ 0ByO7HajOC-FJOGM4NTRlM2MtNWMwOC00ZmY1LWJhOGQtOGEyMzU5OWRiYTk5/edit?usp=sharing [accessed 2013-09-14] [WebCite Cache ID 6JcUjo7lr]

64. Burrows D, Birgin R, Burns K, Zheluk A. Measuring Coverage of HIV Prevention and Care Services for Injecting Drug Users. Vienna: UNODC; 2009.

65. Elovich R, Drucker E. On drug treatment and social control: Russian narcology's great leap backwards. Harm Reduct J 2008;5:23 [FREE Full text] [doi: 10.1186/1477-7517-5-23] [Medline: 18577225]

66. Sarang A, Rhodes T, Sheon N, Page K. Policing drug users in Russia: risk, fear, and structural violence. Subst Use Misuse 2010 May;45(6):813-864 [FREE Full text] [doi: 10.3109/10826081003590938] [Medline: 20397872]

67. Gerber T, Mendelson S. csis.org. Madison: University of Wisconsin-Madison; 2006 Jan. A Survey of Russian Doctors on HIV/AIDS URL: http://csis.org/files/media/csis/pubs/060707_ruseura_mendelsondocrepjan06.pdf [accessed 2013-08-29] [WebCite Cache ID 6JDyt6pqu]

68. Bikmukhametov DA, Anokhin VA, Vinogradova AN, Triner WR, McNutt LA. Bias in medicine: a survey of medical student attitudes towards HIV-positive and marginalized patients in Russia, 2010. J Int AIDS Soc 2012;15(2):17372 [FREE Full text] [Medline: 23031336]

69. Balabanova Y, Coker R, Atun RA, Drobniewski F. Stigma and HIV infection in Russia. AIDS Care 2006 Oct;18(7):846-852. [doi: 10.1080/09540120600643641] [Medline: 16971297]

70. Kozlov AP, Shaboltas AV, Toussova OV, Verevochkin SV, Masse BR, Perdue T, et al. HIV incidence and factors associated with HIV acquisition among injection drug users in St Petersburg, Russia. AIDS 2006 Apr 4;20(6):901-906. [doi: 10.1097/01.aids.0000218555.36661.9c] [Medline: 16549975]

71. Heimer R, White E. Estimation of the number of injection drug users in St. Petersburg, Russia. Drug Alcohol Depend 2010 Jun 1;109(1-3):79-83 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.12.010] [Medline: 20060238]

72. Niccolai LM, Verevochkin SV, Toussova OV, White E, Barbour R, Kozlov AP, et al. Estimates of HIV incidence among drug users in St. Petersburg, Russia: continued growth of a rapidly expanding epidemic. Eur J Public Health 2011 Oct;21(5):613-619 [FREE Full text] [doi: 10.1093/eurpub/ckq115] [Medline: 20798184]

73. Amirkhanian YA, Kelly JA, Kuznetsova AV, DiFranceisco WJ, Musatov VB, Pirogov DG. People with HIV in HAART-era Russia: transmission risk behavior prevalence, antiretroviral medication-taking, and psychosocial distress. AIDS Behav 2011 May;15(4):767-777. [doi: 10.1007/s10461-010-9793-x] [Medline: 20803063]

74. esvero.ru. Global Fund Transitional Funding Mechanism single country application: Sections 1-2. Geneva: theglobalfund.org; 2012. URL: http://www.theglobalfund.org/grantdocuments/RUS(ESVERO)-TFM-HA_Proposal_0_en/ [accessed 2013-08-14] [WebCite Cache ID 6IrFRcmKU]

75. UNAIDS. Report on the Global AIDS Epidemic. Geneva; 2012. URL: http://www.unaids.org/en/media/unaids/contentassets/ documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf [accessed 2013-08-14] [WebCite Cache ID 6IrFNuEBS]

76. Google.com. What are top searches?. Mountain View, CA; 2013. URL: https://support.google.com/trends/answer/99742?hl=en [accessed 2013-08-30] [WebCite Cache ID 6JFi8oHwv]

77. Russian Regions.: Wikipedia.org; 2010 Mar 09. URL: http://en.wikipedia.org/wiki/File:Russian_Regions-EN.svg [accessed 2013-09-03] [WebCite Cache ID 6JLYxbbIc]

78. Moran D, Jordaan JA. HIV/AIDS in Russia: determinants of regional prevalence. Int J Health Geogr 2007;6:22 [FREE Full text] [doi: 10.1186/1476-072X-6-22] [Medline: 17553136]

XSL·FO
RenderX

79. Wall M, Schmidt E, Sarang A, Atun R, Renton A. Sex, drugs and economic behaviour in Russia: a study of socio-economic characteristics of high risk populations. Int J Drug Policy 2011 Mar;22(2):133-139. [doi: 10.1016/j.drugpo.2010.10.001] [Medline: 21055913]

80. CIA World Fact Book - Russia. Washington, DC: Central Intelligence Agency; 2013. URL: https://www.cia.gov/library/publications/the-world-factbook/geos/rs.html [accessed 2013-08-14] [WebCite Cache ID 6IrG6HL0c]

81. Joliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. Stat Methods Med Res 1992;1(1):69-95. [Medline: 1341653]

82. Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pr Assess Res Eval 2005;10(7):1-9 [FREE Full text]

83. Decarli A, La Vecchia C. Environmental factors and cancer mortality in Italy: correlational exercise. Oncology 1986;43(2):116-126. [Medline: 3951786]

84. Kourlaba G, Panagiotakos DB, Mihas K, Alevizos A, Marayiannis K, Mariolis A, et al. Dietary patterns in relation to socio-economic and lifestyle characteristics among Greek adolescents: a multivariate analysis. Public Health Nutr 2009 Sep;12(9):1366-1372. [doi: 10.1017/S1368980008004060] [Medline: 19063765]

85. Howe LD, Galobardes B, Matijasevich A, Gordon D, Johnston D, Onwujekwe O, et al. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: a methods of measurement in epidemiology paper. Int J Epidemiol 2012 Jun;41(3):871-886 [FREE Full text] [doi: 10.1093/ije/dys037] [Medline: 22438428]

86. Russian Federal State Statistics Service. Moscow: Rosstat; 2013. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/ [accessed 2013-08-14] [WebCite Cache ID 6IrF5tUEs]

87. Jolliffe IT. Principal component analysis. New York: Springer; 2002.

88. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. Biometrika 1971 Dec;58(3):453. [doi: 10.2307/2334381]

89. HRW. Russia: Government Shuts HIV-Prevention Group's Website. 2012. URL: http://www.hrw.org/news/2012/02/08/russia-government-shuts-hiv-prevention-group-s-website[WebCite Cache ID 6KxrHaruJ]

90. Golichenko M, Sarang A. Atmospheric pressure - Russian Drug Policy as a Driver for Violations of the UN Convention against Torture. Moscow: Andrey Rylkov Foundation; 2011. URL: http://www.aidslaw.ca/publications/interfaces/downloadFile.php?ref=1949 [accessed 2013-08-14] [WebCite Cache ID 6IrFBV2ZQ]

91. Shlapentokh V. ODRussia. London, UK: opendemocracy.net; 2013 Jun 05. Sticks and stones URL: http://www.opendemocracy.net/od-russia/vladimir-shlapentokh/sticks-and-stones [accessed 2013-08-14] [WebCite Cache ID 6IrFE80lT]

92. Mayer-Schönberger V, Cukier K. Big Data: A Revolution that Will Transform how We Live, Work, and Think. London, UK: Eamon Dolan/Houghton Mifflin Harcourt; 2013.

93. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]

94. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]

95. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013;10(4):e1001413 [FREE Full text] [doi: 10.1371/journal.pmed.1001413] [Medline: 23565065]

Abbreviations

AIDS: Acquired Immune Deficiency Syndrome
HIV: Human Immunodeficiency Virus
NGO: non-governmental organization
PC: principal component
PCA: principal components analysis
PLHIV: people living with HIV
PWID: people who inject drugs
UNAIDS: Joint United Nations Programme on HIV/AIDS
UNODC: United Nations Office on Drugs and Crime

XSL•FO
RenderX

XSL•FO

RenderX