AIDS Behav
DOI 10.1007/s10461-015-1252-2

ORIGINAL PAPER

Action Tweets Linked to Reduced County-Level HIV Prevalence in the United States: Online Messages and Structural Determinants

Molly E. Ireland[1,3] · Qijia Chen[2,3] · H. Andrew Schwartz[2,3,4] · Lyle H. Ungar[2,3] · Dolores Albarracin[2,3]

Abstract HIV is uncommon in most US counties but travels quickly through vulnerable communities when it strikes. Tracking behavior through social media may provide an unobtrusive, naturalistic means of predicting HIV outbreaks and understanding the behavioral and psychological factors that increase communities' risk. General action goals, or the motivation to engage in cognitive and motor activity, may support protective health behavior (e.g., using condoms) or encourage activity indiscriminately (e.g., risky sex), resulting in mixed health effects. We explored these opposing hypotheses by regressing county-level HIV prevalence on action language (e.g., *work*, *plan*) in over 150 million tweets mapped to US counties. Controlling for demographic and structural predictors of HIV, more active language was associated with lower HIV rates. By leveraging language used on social media to improve existing predictive models of geographic variation in HIV, future targeted HIV-prevention interventions may have a better chance of reaching high-risk communities before outbreaks occur.

Keywords General action goals · HIV · Health · Language · Twitter

✉ Molly E. Ireland
molly.ireland@ttu.edu

1 Department of Psychological Sciences, Texas Tech University, MS 2051, Lubbock, TX 79409, USA

2 University of Pennsylvania, Philadelphia, PA, USA

3 University of Illinois at Urbana-Champaign, Champaign, IL, USA

4 Present Address: Department of Computer Sciences, Stony Brook University, Stony Brook, NY, USA

Introduction

Contemporary social influence travels through the online media, because online interactions both engender and reflect real life norms. Online social influence is often likely to be relatively specific, as in the case of delivering messages encouraging targeted health behaviors (e.g., *Get tested*). Other times, however, this social influence is likely to be global, as in the case of general calls for action that do not specifically target health behaviors. For example, in the political domain, general action tendencies and words denoting action (e.g., *go*, *act*, *get engaged*) promote political participation, presumably by connecting to specific goals of political activity [1]. In the health domain, being primed with the same action words promotes physical activity [2, 3]. The present paper concerns the association of references to general action on Twitter with HIV (human immunodeficiency virus). That is, across US counties, does variability in tweets containing action words correlate with variability in HIV prevalence?

HIV remains a significant health concern in the United States. Between 1981 and 2001, over 1.3 million people in the United States were infected with HIV [4]. During the late 1990s, after the introduction of combination antiretroviral therapy, the numbers of new AIDS cases and deaths among adults and adolescents has declined substantially [5]. However, HIV continues to be a major public health threat. According to the CDC, between 2008 and 2010 the estimated number of persons living with diagnosed HIV infections in the United States increased to 282 people per 100,000. Moreover, also in 2010, an estimated 872,990 persons in the United States were living with diagnosed HIV infection. HIV and other forms of sexually transmitted infections pose substantial threats not only to individual health but also to the affected communities that

Published online: 09 December 2015

lose the social and financial contributions of some of those affected and are otherwise burdened by the disease [6–8]

General action refers to cognitive or motor output, irrespective of the particular objects or contexts of action [2, 9–11]. Accordingly, general action goals are motivational end states that can be satisfied by any expenditure of mental or physical energy, regardless of the specific behavior in question, and exert influence on both desirable and undesirable behaviors [3, 9, 12–14]. Action concepts tend to direct individuals towards active pursuits in a variety of domains, and therefore can produce either beneficial effects, such as increased condom use, or detrimental effects, such as sex with multiple partners. The notion that general action goals may manifest in either protective or risky actions has, however, remained largely unexplored up to this point.

Action and Protective Behavior

Action words in online communications could of course have a wide range of meanings and effects. The use of action words has been linked to activities ranging from greater political engagement, including attending a rally or voting in presidential elections, to exercise behavior in the health domain [1]. The four studies reported by Noguchi et al. [1]. revealed that general action tendencies correlated with political participation across regions and across participants in the same region. Moreover, experimentally presented action concepts influenced political participation in a laboratory setting. In the context of the present study, action words may have reciprocal associations with the level of political activity in a community, and such mobilization has been shown to be associated with reduced HIV rates [15]. For example, political action may lead to and stem from greater levels of collective efficacy [16, 17], which could lead to active dissemination of safety norms as well as increased prevention, testing, and treatment in a community. The spread of HIV can in turn decrease through early diagnosis and treatment to reduce viral load and risk of transmission [18].

Beyond the influence of action goals on specific behaviors, such as political participation and exercise, a rich body of research drawing on theory from clinical, social, and personality psychology suggests that being chronically active and holding positive attitudes towards activity are associated with a diverse array of positive mental and physical health outcomes. People who anticipate and actively cope with challenges (by, for example, planning ahead for potential difficulties and seeking social support when distressed) tend to be more resilient and both physically and mentally healthier than those who avoid their problems [19]. One of the main barriers to the treatment of depression is the fact that depressed individuals typically avoid motor action, such as exercise, and social interactions despite knowing that those activities facilitate recovery [20]. Likewise, research on Big Five personality traits and mental health has shown that people who are more emotionally stable have more positive attitudes towards action goals than those who are more neurotic, an association that is partially explained by neurotic individuals' greater anxiety [21].

In behavior change research specifically, interventions that take an ambitious and active approach, recommending a moderate number of behavioral changes, are more successful at effecting lasting change than interventions that target only one or a small number of behaviors [22]. Specific to HIV prevention, interventions that actively involve participants are more successful at increasing preventive health behaviors (such as condom use) than passive interventions that present protective health messages to participants in a less interactive format [23]. Thus, being active and valuing action for its own sake are theoretically linked with protective health behaviors, such as active coping, and resulting mental and physical health benefits. Given the powerful influence of social norms and social networks on health behavior [24, 25], it follows that individuals living in communities that are more active, and that value and reward action, will have lower risk of contracting HIV than individuals living in less active communities.

Action and Risky Behavior

On the other hand, separate research on health behavior and self-control may suggest a positive association between action words and disease prevalence. Previous studies have demonstrated that general action goals are capable of influencing a remarkably varied array of psychological and behavioral outcomes [2, 9]. For example, Albarracín et al. [26] demonstrated that action words embedded in exercise messages (e.g., *active*, *go*) produced a generalized desire for motor output, which in turn led to greater food consumption. General action goals have also been linked to decreased inhibitory control [27]. During a go/no-go task, participants primed with action goals had a significantly smaller P3, an event-related potential component that reflects engagement of brain-level inhibitory control processes. Given this evidence, it is not obvious based on the existing literature whether the activity that action words reflect and promote would correlate with increased or decreased HIV prevalence.

In this research, we performed county-level analyses of the correlations between action words in Twitter and HIV prevalence rates in the US. The dataset included over 150 million tweets from users in the United States. Tweets were randomly sampled from those sent between November

2008 and January 2010 and comprised 10 percent of all tweets sent during that period. As verbal behavior is the gold standard for the study of communication [28, 29], a computerized analysis of the words that users posted on Twitter should adequately capture the contents and implications of the analyzed tweets. Twitter is an ideal data source for analyzing online normative processes and public health because it provides timely and localized information on various events, such as the spread of communicable diseases. Indeed, Twitter has proven to be especially useful in tracking and predicting patterns of transmission for infectious diseases such as influenza [30–34].

Research Objectives

In sum, we explored associations between action tweets and HIV prevalence across US counties, testing the alternate hypotheses that action goals will predict higher or lower HIV rates. Although there is arguably more evidence supporting the hypothesis that action goals are protective rather than risky, the lack of specific prior research on action goals and behavior related to HIV transmission prevents us from making a single directional prediction.

In testing the association between action language and HIV prevalence, it was important to rule out potential confounds and show that action language correlates with HIV rates above and beyond standard demographic and structural predictors of the virus. First, although rural counties also experience HIV outbreaks, particularly in the southeastern United States [35], denser, more urban, and more ethnically diverse counties tend to have much higher HIV rates [36]. Denser and more populated counties also have much higher rates of Twitter usage [37] and a faster pace of life, a behavioral indicator of the general action goals in a community [1]. Therefore, our analyses controlled for population density and county population.

In addition, the percentage of residents who identified as Black was entered as a control variable in the model due to the widely recognized ethnic disparities in HIV prevalence [38–40] We also controlled for foreign-born population, because previous research demonstrated that immigrants have higher prevalence of HIV due to linguistic isolation, lack of health insurance, and limited regular medical care [41, 42], as well as varying cultural practices, relationship disruption, and lack of knowledge about HIV [43]. Lastly, we used the Gini index [44, 45] as a proxy for poverty and income inequality to control for regional economic differences. Past research has shown that poverty in general, and poverty relative to others in a community in particular, is a powerful risk factor for HIV infection [46–50] and should positively correlate with county-level HIV prevalence.

Methods

Our final collection of tweets mapped to US counties included over 150 million messages. Tweets were randomly sampled from those posted between June 2009 and March 2010 and comprised 10 % of all tweets sent in that period. Although Twitter randomly sampled the Tweets themselves and their sampling algorithm is not public, their method is computer-based and adequately represents the Twitter population [51]. We excluded counties from which we had fewer than 10,000 commonly used words ($n = 1046$), resulting in a dataset of 2079 counties. Words in Tweets were identified using an emoticon-aware tokenizer prior to text analysis.

Note that our sample contains over 150 million messages, not users. The number of users who posted those messages is smaller and not known. Individuals who tweet more often will naturally be sampled more often than others. In this analysis, we also included retweets under the assumption that the content that one decides to retweet is also informative.

Geolocation

Tweets and Twitter accounts are not typically associated with geolocation coordinates, and mapping tweets to locations based on the city and state information that users provide themselves is a complex task [52]. For the approximately 2 % of tweets that have geolocation coordinates, it is simply a matter of finding which county the coordinates reside within. However, for the rest of the data we used the free-response location field that accompanies a tweet. This field may contain a city/state pair or an individual city, but it also often has unhelpful phrases (e.g., "behind you"). We used the set of rules described in Schwartz et al. [53] to map location fields to counties. The locations fields were broken up into sequences of words (tokenized) and then matched to country names. Out of those messages either mentioning the country as the United States or not mentioning a country, we used the words preceding the country and attempted to match city and state names. City population information was used when the user provided a city without a state in order to determine whether the city was 90 % likely to be in any state; if so, we paired the city with its most likely state. Otherwise, the tweet was discarded. For example, if Springfield, Illinois has a population of approximately 117,000 and the sum of populations across all cities named Springfield is 187,000, then we would calculate the likelihood that "Springfield" is referring to Springfield, Illinois as 117,000/187,000 = 62.6 %; thus, Springfield would not be mapped. We also excluded city names that happened to match one of the 100 largest non-US cities. Of the total sample of tweets, 78.6 % were discarded due to either a geolocation outside the U.S., lack of any geolocation data, or ambiguous geolocation (from 706 million tweets down to approximately 151 million).

Text Analysis

Our text analysis was based on the Linguistic Inquiry and Word Count (LIWC) [54], a computerized text analysis program that provides the percentage of words that fall into various grammatical, psychological, and topical categories by comparing each word in a given text to a set of internal word lists or dictionaries [55]. LIWC also allows users to create and upload custom dictionaries. We created the action dictionary by compiling LIWC's existing motion and verb categories and then supplementing those lists with synonyms of core action words (*work*, *go*, *plan*, and *think*) using the Corpus of Contemporary American English (COCA) [56]. Finally, words from the entire list that were flagged by at least one of three trained raters as being ambiguous with respect to action goals (e.g., *beat*) were culled from the dictionary if COCA showed that they were used to denote action (e.g., "I'll *beat* you at poker," "*Beat* the eggs") and inaction (e.g., "I'm *beat*," "He looks *beaten*") at similar rates. The resulting action dictionary contains 854 words and stems that are related to general motor (e.g., *fly*, *gaming*, *gym*) and cognitive activity (e.g., *plan*, *deduce, realize*).

Results

To test our hypotheses, we regressed HIV prevalence on action words while controlling for socioeconomic disease determinants, including wealth disparity (Gini index), percentage of the population that identifies as Black, percentage of the population that is foreign born, total population, and population density. We fitted all models with the nlme package [57, 58] for estimating fixed effects and variance/covariance component parameters of the linear mixed models. The nlme package is supplied in the R system for statistical computing (Ver. 3.0.1) [59] under the GNU General Public License (Version 3, June 2007). The county-level data were nested within states to control for regional interdependence between counties as a function of differences between states' public health policies, and slopes for action words were allowed to vary randomly between states [60]. All variables were z-scored to increase the interpretability of regression coefficients.

The fixed-effects coefficients indicated that more frequent action word usage on Twitter was associated with lower HIV rates ($B = -0.18$, $SE = 0.07$, $t = -2.67$, $p = .008$), even after controlling for counties' population density, population, proportion of African Americans, proportion of foreign-born populations, and the Gini index. State-level intercepts and slopes appear in Table 1 and in the map in Fig. 1. As was the case with the average fixed-effect slope, the state-level slope (i.e., the association between county-level HIV and states' average levels of action language) was also significantly negative ($B = -0.11$, $SE = 0.02$, $t = -3.19$, $p < .001$). Overall then, areas with more active tweets had lower HIV rates, although the state-level effect was slightly weaker than the county-level effect. For summary statistics for the model, see Table 2.

In line with previous findings, all factors indicating structural vulnerabilities were found to be positively associated with HIV prevalence. Specifically, in the same model, greater proportion of African-Americans ($B = 0.48$, $SE = 0.02$, $t = 27.20$, $p < .001$), population density ($B = 0.38$, $SE = 0.01$, $t = 26.95$, $p < .001$), population ($B = 0.05$, $SE = 0.01$, $t = 3.38$, $p = .001$), socioeconomic inequality (Gini index; $B = 0.06$, $SE = 0.02$, $t = 3.42$, $p = .001$), and proportion of foreign-born residents ($B = 0.15$, $SE = 0.02$, $t = 7.39$, $p < .001$) were all related to higher HIV prevalence.

Discussion

The current findings offer novel insights into preventative factors associated with reduced rates of HIV transmission. The potentially substantial preventative effects of action tendencies on HIV prevalence have important implications for public health policies and intervention implementations. We explored the possibility that action tendencies would correlate with HIV prevalence, although further examinations of causal chains are necessary to fully understand the association. Consistent with prior findings [1], general action tendencies may be responsible for greater community involvement, policy initiative, and utilization of accessible resources. When community members are more likely to take actions addressing existing problems, they are less vulnerable to the threats that are often amplified by lack of institutional support and healthcare access [61]. Therefore, active community environments may facilitate prompt diagnosis and treatment of HIV and eventually contribute to the reduction in transmission and prevalence of such diseases.

Future research should further seek to disentangle the effects of general and specific action motivation on community health behavior. That is, the current data cannot determine whether it is generally active communities or communities that are specifically proactive with regards to sexual health that promote protective behaviors and

Table 1 State-level random effects of action words on counties' HIV prevalence

State	Intercept	*B*
Alabama	−0.30	−0.15
Arizona	−0.07	0.05
Arkansas	−0.16	−0.07
California	−0.19	−0.15
Colorado	0.04	−0.35
Connecticut	−0.08	−0.41
Delaware	−0.11	−0.46
Florida	0.08	−0.60
Georgia	−0.11	−0.23
Hawaii	−0.07	−0.04
Idaho	−0.09	−0.04
Illinois	−0.13	0.12
Indiana	−0.06	−0.04
Iowa	−0.11	−0.02
Maryland	−0.33	−2.42
Massachusetts	0.00	−0.14
Michigan	−0.17	0.05
Minnesota	−0.11	−0.05
Mississippi	−0.32	−0.38
Missouri	−0.01	−0.19
Montana	−0.09	−0.02
Nebraska	−0.09	0.01
Nevada	−0.07	−0.18
New Hampshire	−0.08	−0.08
New Jersey	−0.17	−0.67
New Mexico	−0.08	−0.05
New York	0.16	0.24
North Carolina	−0.07	−0.19
Ohio	−0.11	0.03
Oklahoma	−0.13	−0.03
Oregon	−0.10	−0.15
Pennsylvania	−0.01	−0.12
Rhode Island	−0.15	−0.15
South Carolina	−0.07	−0.29
Tennessee	0.00	−0.16
Texas	−0.23	−0.06
Utah	−0.12	0.00
Vermont	−0.07	−0.01
Virginia	−0.01	−0.20
Washington	−0.14	−0.05
West Virginia	−0.07	−0.02
Wisconsin	−0.14	0.01
Wyoming	−0.07	0.00

All variables were z-scored before analyses. Random effects are from a hierarchical linear model that allowed slopes and intercepts for action language to vary randomly between states and controlled for major socioeconomic determinants of HIV

ultimately reduce HIV risk. In theory, the two will be closely related: Communities in which individuals tend to vote and volunteer are also likely to be communities that take proactive steps toward preventing HIV.

It is important to interpret our results cautiously, however. At present, we are not making the relatively strong claim that people benefit from reading or posting social media messages containing references to general cognitive or motor action. Rather, we propose that people are healthier and at lower risk of HIV when they internalize their community's tendency to be active and value action. These proactive social norms are reflected and, in part, disseminated by general action words—regardless of whether those words specifically refer to health behaviors that are relevant to HIV, such as using condoms or getting tested for STIs.

Although social media can be a powerful source of social influence, this study primarily uses Twitter as a relatively representative source of data on real-life communication and naturalistic behavior. In fact, it is reasonable to assume most of our tweets came from individuals without HIV, and thus the connection with HIV rates is merely reflective of community characteristics rather than individual differences in the likelihood of personally contracting HIV. In other words, our conclusions do not hinge on the assumption that Twitter is itself the mechanism of behavior change or even the means through which people support and disseminate social norms. At minimum, without making any claims about whether people are directly influenced by others' tweets, action language on Twitter is a face-valid indicator of a community's endorsement of general action goals. It is the normative belief that action is good and valuable for its own sake that, we argue, may lead to a constellation of protective and proactive behaviors that make individuals healthier in general and less vulnerable to HIV infection in particular.

Limitations

Our ability to offer theoretical explanations and make inferences about causal mechanisms is undoubtedly limited by the cross-sectional nature of the data. Further exploration of potential causal pathways will be necessary to explicate the associations between action words and changes in HIV transmission and prevalence over time. If we provisionally assume that action language is causally related to protective behaviors, we are still left with the question of whether personally using or merely being around more active language use leads to protective behaviors, such as wearing condoms or getting tested for

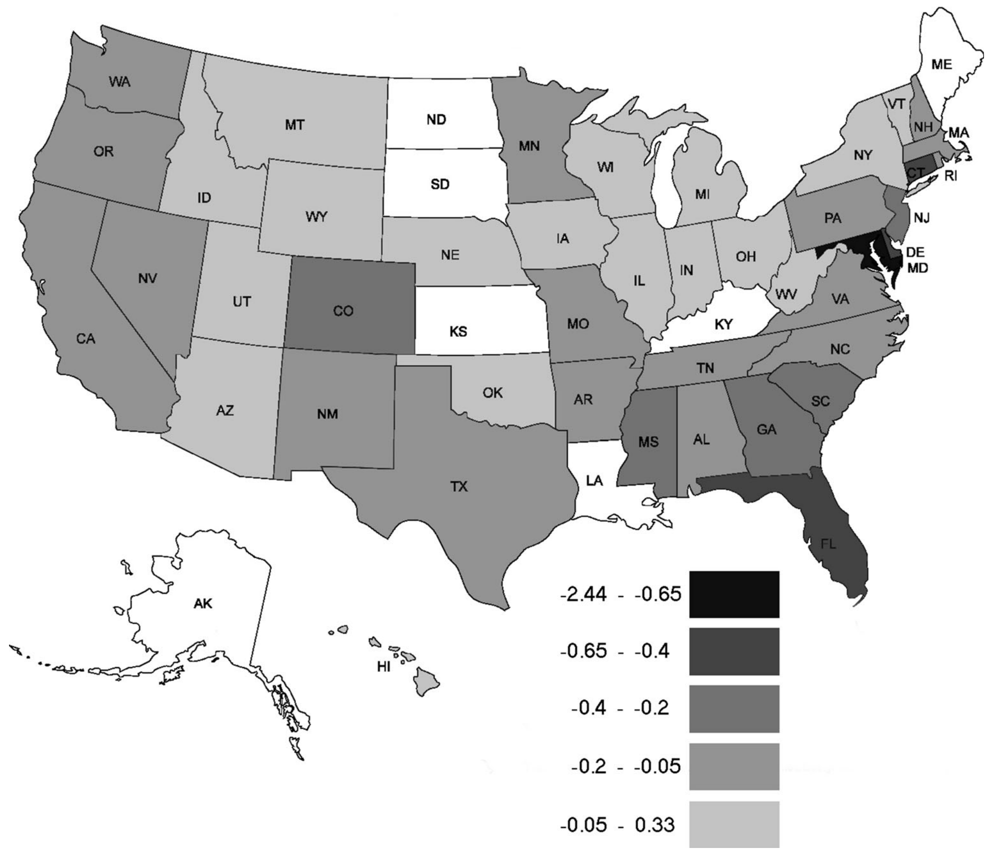

Fig. 1 Associations between HIV prevalence and action tweets by US state. *Numbers* in the legend are random effects from a hierarchical linear model that allowed slopes and intercepts for action language to vary randomly between states and controlled for major socioeconomic determinants of HIV. (For specific state-level intercepts and regression coefficients, see Table 1.)

Table 2 Results of the random coefficients model

Fixed effect	B	SE
Intercept, γ_{00}	−0.10***	0.03
Action words, γ_{01}	−0.18***	0.07
Black population, γ_{02}	0.48***	0.02
Foreign-born population, γ_{03}	0.15***	0.02
Population density, γ_{04}	0.38***	0.01
Gini index, γ_{05}	0.06***	0.02
Total population, γ_{06}	0.05***	0.01
State-level slope	*B*	*SE*
State average	−0.11***	0.02
Random effect	Variance component	
State mean, u_{0j}	0.02	
Action words slopes, u_{1j}	0.17	
Level-1 effect: e_{ij}	0.24	

Results are from a hierarchical linear model allowing slopes and intercepts for counties' action language to vary randomly between states. All tests are two-tailed

*** $p < 0.001$

HIV. In the future, these questions could be examined by comparing the effects of exposing individuals to action language versus inducing individuals to use more active language themselves. Given that language use, motivation, and emotion all tend to be contagious [62–64], we suspect that both processing and producing action language will be positively related to proactive health behaviors (e.g., using condoms) as well as broader community-level actions (e.g., voting in government elections) that have been associated with action goals in past research [1]. It may be the case, however, that the effects of action messages fall along a continuum, with self-motivated action language having the strongest relation to proactive behavior.

Another potentially important limitation of this research has to do with our text analysis method itself. Any of the words in the action dictionary can be referred to in a negative or positive context. Indeed, many common action words, such as *work* or *exercise*, are likely to engender mixed feelings. LIWC, like most text analysis tools that simply search for key words in an internal dictionary, is blind to context. Thus, whether people say that they hate or love their work, our action dictionary considers that message to be active or reflect action goals. Likewise, a message that mentions work in passing to introduce a joke or derisive comment about work ("This sums up 90 % of my *work* day") [65] will be considered as active as an explicitly pro-work message ("*Work* harder than anyone else") [66].

Text analysis research that uses the context-blind dictionary approach is typically defended with the argument that, even when a word is accompanied by a negation or a negative emotion word, mentioning a word means that you have that concept in mind [67]. A person who tweets about not wanting to go to work nevertheless is employed and is thinking about their job at the time of the tweet. The same is true of negative messages about protective health actions such as working out or going to the doctor, or political actions, such as voting. A person would be unlikely to complain about the lines at a voting booth or not wanting to go for a run if they did not live in a community where voting and exercising were valued, normative behaviors.

There are other cases, however, in which people may use action words that have little to do with action goals in general or action tendencies in their community (e.g., "Siri, delete my spam. *Move* all pizza coupons to inbox" [68]). However, in terms of accurately assessing the degree to which a community cares about action, LIWC and similar probabilistic dictionary approaches are designed to be right more often than wrong. Especially when the linguistic sample comprises billions of words, the dictionary approach that we have used in this study is a straightforward and efficient means of measuring what communities tend to talk about and do. This claim is supported by other studies that have used dictionary methods to accurately map the spread of influenza and cholera [69, 70] and uncover associations between references to HIV-risk behaviors (e.g., sex, drug use) and HIV prevalence [31].

Finally, our sample is limited by missing data for counties with lower HIV rates. Although rules for data suppression vary between counties and states, many counties with lower rates of HIV (e.g., fewer than 20 cases) do not report incidence rates due to confidentiality concerns [71]. Many of these counties tend to have smaller and more rural populations. Because of the low incidence rates in these counties, they may not be accustomed to considering HIV a threat or may stigmatize HIV testing to a greater degree than counties with higher HIV rates; therefore, these less populated counties may be at particularly high risk for rapid outbreaks when a few new cases arise. Indeed, large rural populations, conservatism, and distrust of medical doctors are partially responsible for the intransigence of HIV rates in the southern United States [35]. It is also possible that some of these suppressed counties have low HIV rates despite having large or urban populations, and thus could serve as informative examples of communities with highly successful preventive health infrastructures. Future studies with a narrower focus on communities with lower HIV incidence may seek to work around data suppression rules by contacting county health departments and accessing suppressed data directly.

Conclusion

Despite all limitations, the current findings are critically informative and have the potential to be highly transformative for prevention and surveillance practices. By combining Twitter data extraction and sophisticated geocoding techniques, public health researchers can potentially estimate the current HIV risk of specific locations and implement preventive measures before the actual outbreak occurs. Furthermore, the current findings indicate that in addition to conventional intervention targets, such as changing individuals' high-risk behavioral patterns, we might also be able to reduce HIV transmission indirectly by mobilizing local communities and promoting more active and enterprising community norms.

Acknowledgments We are grateful to Justin Hepler and Melanie Tannenbaum for their help in creating the action dictionary, and to Micah Iserman for his figure-making assistance.

References

1. Noguchi K, Handley IM, Albarracín D. Participating in politics resembles physical activity: general action patterns in international archives, United States archives, and experiments. Psychol Sci. 2011;22(2):235–42. doi:10.1177/0956797610393746.
2. Albarracín D, Handley IM, Noguchi K, et al. Increasing and decreasing motor and cognitive output: a model of general action and inaction goals. J Pers Soc Psychol. 2008;95(3):510–23. doi:10.1037/a0012833.
3. Hepler J, Wang W, Albarracin D. Motivating exercise: the interactive effect of general action goals and past behavior on physical activity. Motiv Emot. 2012;36(3):365–70. doi:10.1037/e527772014-020.
4. Fleming PL. The epidemiology of HIV and AIDS. In: Wormser GP, editor. AIDS and other manifestations of HIV infection. 4th ed. London: Elsevier Academic Press; 2004. p. 3–29.
5. Karon JM, Fleming PL, Steketee RW, De Cock KM. HIV in the United States at the turn of the century: an epidemic in transition. Am J Public Health. 2001;91(7):1060–8. doi:10.2105/ajph.91.7.1060.
6. Chesson HW, Gift TL, Owusu-Edusei K Jr, Tao G, Johnson AP, Kent CK. A brief review of the estimated economic burden of sexually transmitted diseases in the United States: inflation-adjusted updates of previously published cost studies. Sex Transm Dis. 2011;38(10):889–91. doi:10.1097/olq.0b013e318223be77.
7. Owusu-Edusei K Jr, Chesson HW, Gift TL, et al. The estimated direct medical cost of selected sexually transmitted infections in the United States, 2008. Sex Transm Dis. 2013;40:197–201. doi:10.1097/olq.0b013e318285c6d2.
8. Steiner MJ, Cates W. Condoms and sexually-transmitted infections. New Engl J Med. 2006;354(25):2642–3. doi:10.1056/NEJMp068111.
9. Albarracin D, Hepler J, Tannenbaum M. General action and inaction goals: their behavioral, cognitive, and affective origins and influences. Curr Dir Psychol Sci. 2011;20(2):119–23. doi:10.1177/0963721411402666.
10. Gendolla GHE, Silvestrini N. The implicit "go": masked action cues directly mobilize mental effort. Psychol Sci. 2010;21(10):1389–93. doi:10.1177/0956797610384149.

11. Laran J. The influence of information processing goal pursuit on postdecision affect and behavioral intentions. J Pers Soc Psychol. 2010;98(1):16–28. doi:10.1037/a0017422.
12. Albarracín D, Handley IM. The time for doing is not the time for change: effects of general action and inaction goals on attitude retrieval and attitude change. J Pers Soc Psychol. 2011;100(6): 983–98. doi:10.1037/a0023245.
13. Hepler J, Albarracin D, McCulloch KC, Noguchi K. Being active and impulsive: the role of goals for action and inaction in self-control. Motiv Emot. 2012;36(4):416–24. doi:10.1007/s11031-011-9263-4.
14. Albarracin D, Hart W. Positive mood + action = negative mood + inaction: effects of general action and inaction concepts on decisions and performance as a function of affect. Emotion. 2011;11(4):951–7. doi:10.1037/a0024130.
15. Mino M, Deren S, Kang S-Y, Guarino H. Associations between political/civic participation and HIV drug injection risk. Am J Drug Alcohol Abuse. 2011;37(6):520–4. doi:10.3109/00952990.2011.600384.
16. Fernández-Ballesteros R, Díez-Nicolás J, Caprara GV, Barbaranelli C, Bandura A. Determinants and structural relation of personal efficacy to collective efficacy. Appl Psychol-Int Rev. 2002;51(1):107. doi:10.1111/1464-0597.00081.
17. Galavotti C, Wheeler T, Sebert Kuhlmann A, et al. Navigating the swampy lowland: A framework for evaluating the effect of community mobilisation in female sex workers in Avahan, the India AIDS initiative. J Epidemiol Commun Health. 2012;66: ii9–15. doi:10.1136/jech-2011-200465.
18. Quinn TC, Wawer MJ, Sewankambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. New Engl J Med. 2000;342(13):921–9. doi:10.1093/oxfordjournals.aje.a009331.
19. Tugade MM, Fredrickson BL, Feldman-Barrett L. Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. J Pers. 2004;72(6):1161–90. doi:10.1111/j.1467-6494.2004.00294.x.
20. Mazzucchelli T, Kane R, Rees C. Behavioral activation treatments for depression in adults: a meta-analysis and review. Clin Psychol. 2009;16(4):383–411. doi:10.1111/j.1468-2850.2009.01178.x.
21. Ireland ME, Hepler J, Li H, Albarracín D. Neuroticism and attitudes toward action in 19 countries. J Pers. 2015;83(3): 243–50.
22. Wilson K, Senay I, Durantini M, Sanchez F, Hennessy M, Spring B, Albarracin D. When it comes to lifestyle recommendations, more is sometimes less: a meta-analysis of theoretical assumptions underlying the effectiveness of interventions promoting multiple behavior domain change. Psychol Bull. 2015;141(2): 474–509. doi:10.1037/a0038295.
23. Albarracin D, Gillette JC, Earl A, Glasman LR, Durantini MR, Ho M-H. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull. 2005;131(6):856–97. doi:10.1037/0033-2909.131.6.856.
24. Latkin C, Donnell D, Liu TY, Davey-Rothwell M, Celentano D, Metzger D. The dynamic relationship between social norms and behaviors: the results of an HIV prevention network intervention for injection drug users. Addiction. 2013;108:934–43.
25. Latkin CA, Knowlton AR. Micro-social structural approaches to HIV prevention: a social ecological perspective. AIDS Care. 2005;17(sup1):102–13. doi:10.1080/09540120500121185.
26. Albarracin D, Wang W, Leeper J. Immediate increase in food intake following exercise messages. Obesity. 2009;17(7):1451–2. doi:10.1038/oby.2009.16.
27. Hepler J, Albarracin D. Complete unconscious control: using (in)action primes to demonstrate completely unconscious activation of inhibitory control mechanisms. Cognition. 2013; 128(3):271–9. doi:10.1016/j.cognition.2013.04.012.
28. Marmarosh CL, Kivlighan DM Jr. Relationships among client and counselor agreement about the working alliance, session evaluations, and change in client symptoms using response surface analysis. J Couns Psychol. 2012;59(3):352–67. doi:10.1037/a0028907.
29. Pennebaker JW. Theories, therapies, and taxpayers: on the complexities of the expressive writing paradigm. Clin Psychol. 2004;11(2):138–42. doi:10.1093/clipsy.bph063.
30. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza a H1N1 pandemic. PLoS One. 2011;6(5):e19467. doi:10.1371/journal.pone.0019467.
31. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med. 2014;63:112–5. doi:10.1016/j.ypmed.2014.01.024.
32. Thompson WW, Comanor L, Shay DK. Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease. J Infect Dis. 2006;194:S82–91. doi:10.1086/507558.
33. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012–4. doi:10.1038/nature07634.
34. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. Ann Emerg Med. 2003;41(4):447–52. doi:10.1067/mem.2003.104.
35. Adimora AA, Ramirez C, Schoenbach VJ, Cohen MS. Policies and politics that promote HIV infection in the southern United States. AIDS. 2014;500:13–01357. doi:10.1097/QAD.0000000000000225.
36. Holmberg SD. The estimated prevalence and incidence of HIV in 96 large US metropolitan areas. Am J Public Health. 1996;86(5): 642–54. doi:10.2105/ajph.86.5.642.
37. Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in twitter. Proceedings of the 2nd international workshop on Search and mining user-generated contents—SMUC'10 [Internet]. ACM Press; 2010. doi:10.1145/1871985.1871993.
38. Adimora AA, Schoenbach VJ. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. J Infect Dis. 2005;191:S115–22. doi:10.1086/425280.
39. Prejean J, Song R, Hernandez A, et al. Estimated HIV incidence in the United States, 2006-2009. PLoS One. 2011;6(8):e17502. doi:10.1371/journal.pone.0017502.
40. Song R, Hall HI, Harrison KM, Sharpe TT, Lin LS, Dean HD. Identifying the impact of social determinants of health on disease rates using correlation analysis of area-based summary information. Public Health Rep. 2011;126(Suppl 3):70–80.
41. Shedlin MG, Deren S. Cultural factors influencing HIV risk behavior among dominicans in new york city. J Ethn Subst Abuse. 2002;1(1):71. doi:10.1300/j233v01n01_05.
42. Levy V, Page-Shafer K, Evans J, et al. HIV-related risk behavior among Hispanic immigrant men in a population-based household survey in low-income neighborhoods of northern California. Sex Transm Dis. 2005;32(8):487–90. doi:10.1097/01.olq.0000161185.06387.94.
43. Kandula NR, Kersey M, Lurie N. Assuring the health of immigrants: what the leading health indicators tell us. Ann Rev Publ Health. 2004;25:357–76. doi:10.1146/annurev.publhealth.25.101802.123107.

44. Ceriani L, Verme P. The origins of the GINI index: extracts from variabilità e mutabilità (1912) by Corrado Gini. J Econ Inequal. 2011;10(3):421–43. doi:10.1007/s10888-011-9188-x.
45. Gini C. Variabilità e mutabilità. Bologna: C. Cuppini; 1912.
46. Joy R, Druyts EF, Brandson EK, et al. Impact of neighborhood-level socioeconomic status on HIV disease progression in a universal health care setting. JAIDS. 2008;47(4):500–5. doi:10.1097/qai.0b013e3181648dfd.
47. McIntosh WA, Thomas JK. Economic and other societal determinants of the prevalence of HIV: a test of competing hypotheses. Soc Psychol Q. 2004;45(2):303–24. doi:10.1111/j.1533-8525.2004.tb00014.x.
48. McMahon J, Wanke C, Terrin N, Skinner S, Knox T. Poverty, hunger, education, and residential status impact survival in HIV. AIDS Behav. 2011;15:1503–11. doi:10.1007/s10461-010-9759-z.
49. Krueger LE, Wood RW, Diehr PH, Maxwell CL. Poverty and HIV seropositivity: the poor are more likely to be infected. AIDS. 1990;4(8):811–4. doi:10.1097/00002030-199008000-00015.
50. Harrison KM, Ling Q, Song R, Hall HI. County-level socioeconomic status and survival after HIV diagnosis, United States. Ann Epidemiol. 2008;18(12):919–27. doi:10.1016/j.annepidem.2008.09.003.
51. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough?. AAAI: Comparing data from Twitter's streaming API with Twitter's firehose; 2013.
52. Hecht B, Hong L, Suh B, Chi EH. Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. *SIGCHI*. 2011. p. 237–46.
53. Schwartz HA, Eichstaedt JC, Kern ML et al. Characterizing geographic variation in well-being using Tweets. *ICWSM*. 2013. p. 583–91.
54. Pennebaker JW, Booth RE, Francis ME. Linguistic inquiry and word count: LIWC. LIWC.Net. [computer program]. 2007.
55. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. The development and psychometric properties of LIWC2007 [Software manual]. *LIWC.net*. 2007. http://goo.gl/CPK5gt.
56. Davies M. The corpus of contemporary american english: 450 million words, 1990-present [Online linguistic corpus]. 2008. http://corpus.byu.edu/coca/.
57. *The lme4 package: Linear mixed-effects models using s4 classes* [computer program]. Version 0.999375-282009.
58. Pinheiro JC, Bates DM. Mixed effects models in s and s-plus. New York: Springer; 2000.
59. Team RDC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2011.
60. Bryk AS, Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Newbury Park: Sage; 1992.
61. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health. 2010;100(Suppl 1):S40–6. doi:10.2105/ajph.2009.184036.
62. Chartrand TL, van Baaren R. Human mimicry. In: Zanna MP, editor. Exp Soc Psychol, vol. 41. San Diego: Elsevier Academic Press; 2009. p. 219–74. doi:10.1016/s0065-2601(08)00405-x.
63. Dik G, Aarts H. Behavioral cues to others' motivation and goal pursuits: the perception of effort facilitates goal inference and contagion. J Exp Soc Psychol. 2007;43(5):727–37. doi:10.1016/j.jesp.2006.09.002.
64. Ireland ME, Pennebaker JW. Language style matching in writing: synchrony in essays, correspondence, and poetry. J Pers Soc Psychol. 2010;99(3):549–71. doi:10.1037/a0020386.
65. @IrisClasson. This sums up 90% of my work day. https://twitter.com/irisclasson/status/587710479444877312. Accessed 13 April 2015.
66. @RealMichaelKay. Work harder than anyone else and treat people the way you want to be treated. https://twitter.com/RealMichaelKay. Accessed 13 April 2015.
67. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol. 2010;29(1):24–54. doi:10.1177/0261927x09351676.
68. @madpoetry. "Siri, delete my spam." "Move all pizza coupons to inbox." https://twitter.com/madpoetry. Accessed 13 April 2015.
69. Radinsky K, Horvitz E. Mining the web to predict future events. ACM-WSDM. 2013. doi:10.1145/2433396.2433431.
70. Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. *AAAI*. 2012. p. 136–42.
71. Delcher PC, Edwards KT, Stover JA, Newman LM, Groseclose SL, Rajnik DM. Data suppression strategies used during surveillance data release by sexually transmitted disease prevention programs. J Public Health Man. 2008;14(2):E1–8. doi:10.1097/01.phh.0000311902.95948.f5.