

## SOCIAL SCIENCES

# Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties

Christopher A. Bail<sup>1,2\*</sup>, Friedolin Merhout<sup>1</sup>, Peng Ding<sup>3</sup>

Recent terrorist attacks by first- and second-generation immigrants in the United States and Europe indicate that radicalization may result from the failure of ethnic integration—or the rise of intergroup prejudice in communities where “home-grown” extremists are raised. Yet, these community-level drivers are notoriously difficult to study because public opinion surveys provide biased measures of both prejudice and radicalization. We examine the relationship between anti-Muslim and pro-ISIS (Islamic State of Iraq and Syria) Internet searches in 3099 U.S. counties between 2014 and 2016 using instrumental variable models that control for various community-level factors associated with radicalization. We find that anti-Muslim searches are strongly associated with pro-ISIS searches—particularly in communities with high levels of poverty and ethnic homogeneity. Although more research is needed to verify the causal nature of this relationship, this finding suggests that minority groups may be more susceptible to radicalization if they experience discrimination in settings where they are isolated and therefore highly visible—or in communities where they compete with majority groups for limited financial resources. We evaluate the validity of our findings using several other data sources and discuss the implications of our findings for the study of terrorism and intergroup relations, as well as immigration and counterterrorism policies.

## INTRODUCTION

Until recently, the vast majority of terrorist attacks by ISIS (Islamic State of Iraq and Syria) and other violent extremist organizations occurred in the Middle East, South Asia, and North Africa (1). Yet, recent years have witnessed significant terrorist attacks by first- and second-generation immigrants in the United States and Europe (2–4). The rise of “home-grown” terrorism coincides with a steady increase in anti-Muslim sentiment in both the United States and Europe (5, 6). There is thus growing concern that anti-Muslim animus may feed the narrative of extremist organizations, such as ISIS, who argue that the West is at war with Islam.

The notion that ethnic discrimination may shape radicalization has precedent in group threat theory (7–9). This theory posits that prejudice results from a perceived threat between majority and minority groups. Majority groups often develop stereotypes about minority groups based on observation of a small group of deviants among them—particularly in settings where positive interpersonal contact between majority and minority groups is rare. Minority groups that experience discrimination from majority groups often feel threatened in turn because they view this prejudice as irrational or unjustified.

Although group threat theory emphasizes symbolic or perceived differences between groups, previous studies indicate that socioeconomic and demographic factors also shape the prevalence of intergroup prejudice. For example, intergroup prejudice is more prevalent when majority and minority groups are in competition for scarce social resources such as education, employment, or wealth (7, 9). Similarly, research indicates that prejudice is more prevalent in the wake of demographic shifts that bring majority and minority groups into abrupt contact with each other, such as large-scale, labor-based migration or refugee resettlement programs (10). Finally, prejudice is particularly common

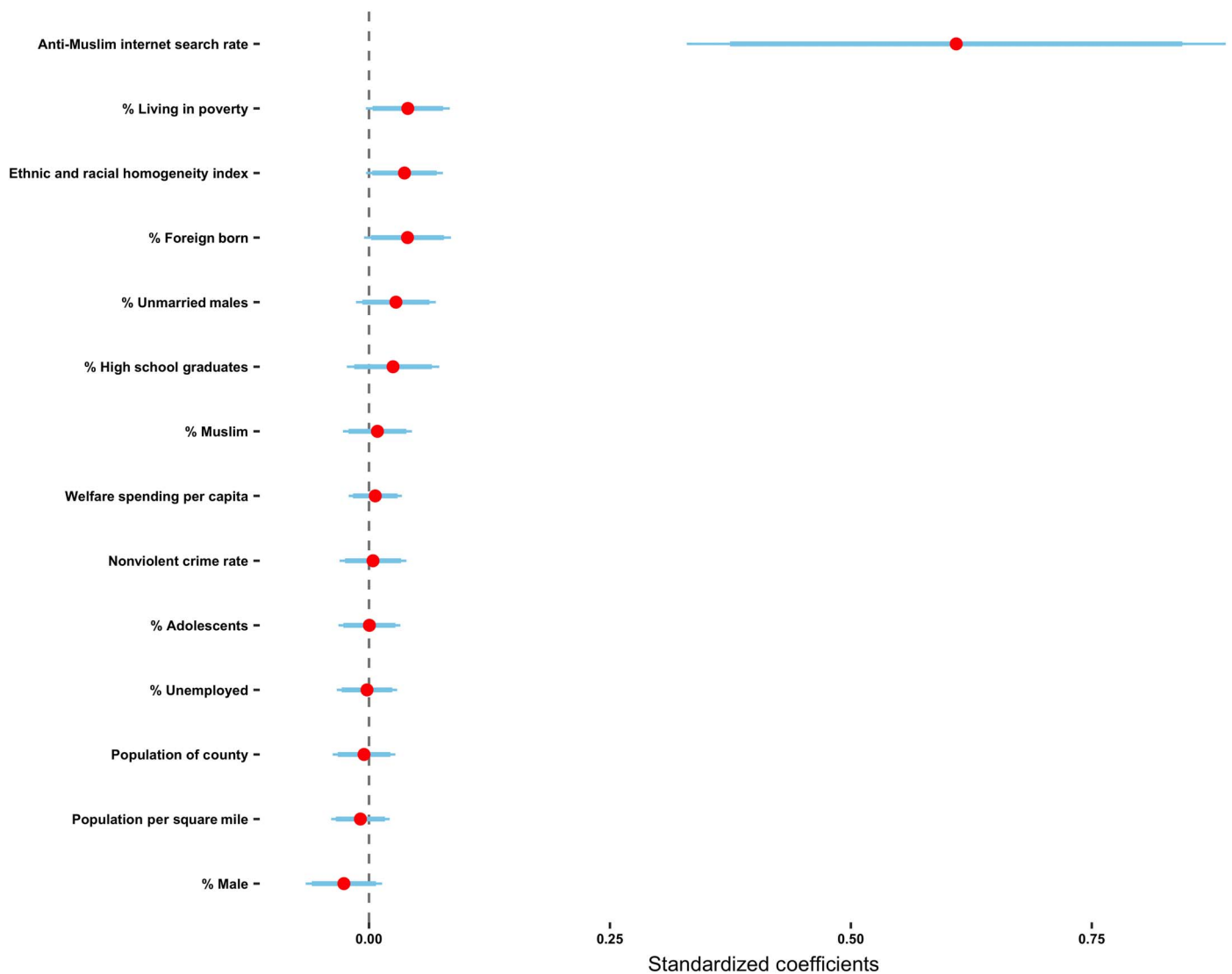
in communities with high levels of ethnic homogeneity, where distinctions between majority and minority groups are highly visible—sharpening perceptions of “us” and “them” (11, 12).

Group threat theory—or contemporary extensions of this theory—has been applied to explain the emergence of intergroup prejudice in many different settings. Yet, to our knowledge, this theory has not been invoked to study radicalization, or the process through which violent extremist organizations gain support. Although these organizations span many different ethnic and religious groups, inattention to the relationship between ethnic discrimination and radicalization is particularly surprising in the case of ISIS. This is because ISIS brands itself as a Muslim organization, and many immigrant communities that originated within Muslim-majority countries are currently experiencing significant prejudice amidst socioeconomic and demographic shifts that might exacerbate intergroup tensions—although there is far less evidence of either radicalization or socioeconomic deprivation among Muslim populations in the United States than many Western European countries (2, 5).

Drawing from group threat theory, we hypothesize that pro-ISIS sentiment will be most prevalent in communities with high levels of anti-Muslim sentiment—especially when these communities are ethnically homogeneous or poor. Evaluating this hypothesis presents numerous methodological obstacles. Would-be extremists are unlikely to identify as such within public opinion surveys. Nearly every study of violent extremism only examines those who successfully executed terrorist attacks or were arrested before doing so. These studies not only suffer from selection bias but also limit our understanding of the initial stages of the radicalization process and overlook the potentially sizable group of people who develop views that are sympathetic to violent extremist organizations but do not engage in violent behavior, or otherwise support these groups (13, 14). Similarly, many people who hold prejudiced views against minority groups, such as Muslims, are unlikely to express these attitudes in public opinion surveys because of social desirability bias, or their desire not to appear bigoted, or violate norms of political correctness (15).

<sup>1</sup>Department of Sociology, Duke University, 254 Sociology-Psychology Building, Durham, NC 27708, USA. <sup>2</sup>Sanford School of Public Policy, Duke University, Box 90239, Durham, NC 27708, USA. <sup>3</sup>Department of Statistics, University of California, Berkeley, 425 Evans Hall, Berkeley, CA 94720, USA.

\*Corresponding author. Email: christopher.bail@duke.edu



**Fig. 1. Standardized coefficients from two-stage least-squares regression model predicting pro-ISIS Internet search rate normalized by overall search activity in 3099 U.S. counties, 2014–2016.** Red circles describe standardized coefficients, and blue bars describe 90 and 95% confidence intervals (adjusted  $R^2 = 0.43$ ). The coefficient for the normalized anti-Muslim search rate is the strongest and only statistically significant predictor of pro-ISIS search volume.

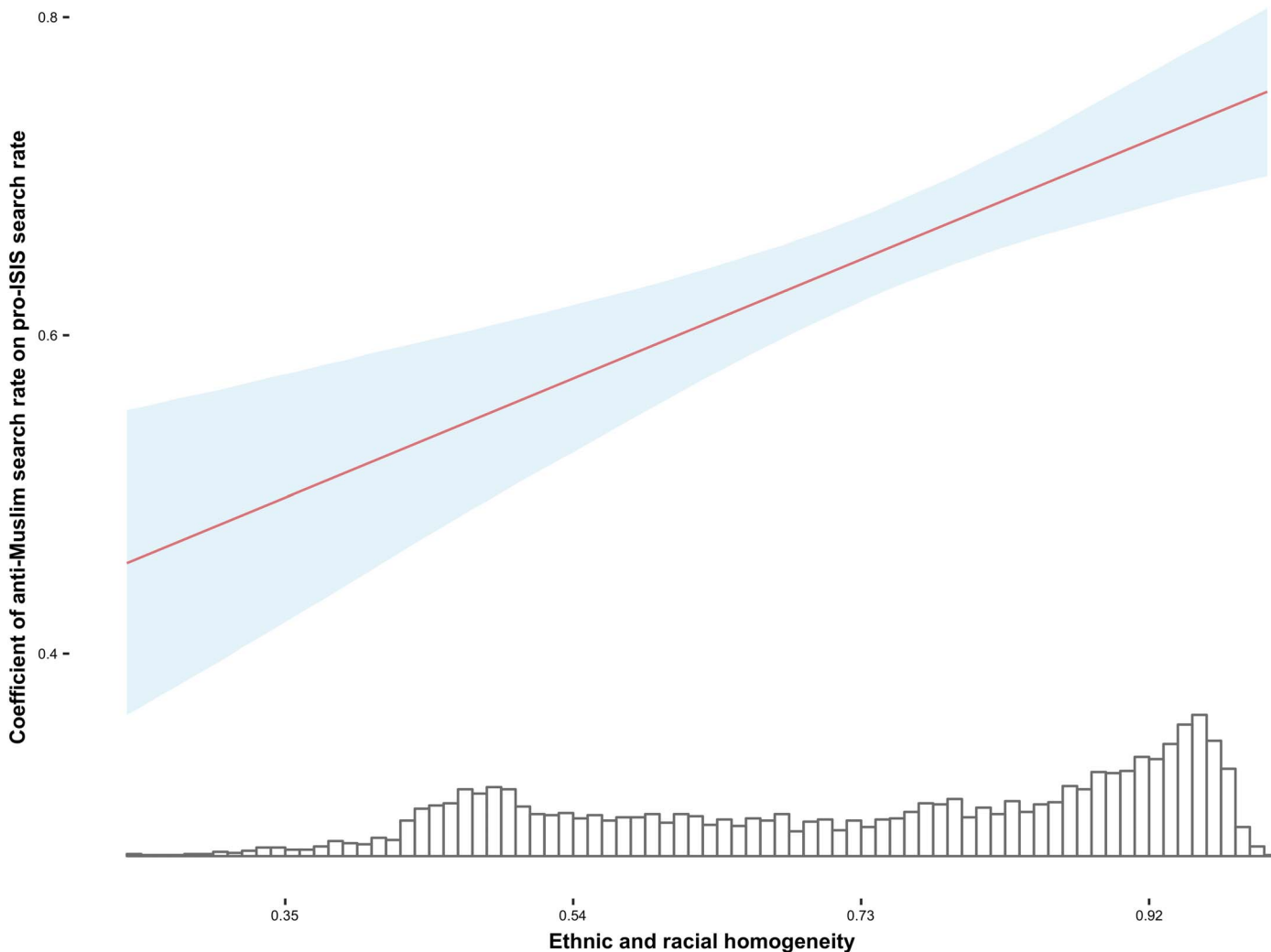
## RESULTS

We argue that Internet search data provide a valuable opportunity to examine the relationship between ethnic discrimination and radicalization at the community level—although these digital trace data are not without their own significant limitations, as we discuss below. We collected data that describe the average monthly search volume of pro-ISIS and anti-Muslim search phrases on Google and other leading search engines from August 2014 to July 2016 in 3099 U.S. counties, which we adjusted for the overall volume of Internet searches therein. These searches include phrases such as “How to join ISIS” and “Muslims are evil.” Internet search data are particularly valuable for the study of ethnic discrimination and radicalization since previous studies indicate that the relative anonymity of the Internet provides a refuge to those who hold prejudiced or extreme views (16–19). Because people typically perform Internet searches about sensitive subjects outside the observation of others, these data hold considerable promise to mitigate social desirability bias—and by extension selection bias—in

the study of ethnic discrimination and radicalization at the community level.

We gathered data from the U.S. Census, the American Community Survey, and the U.S. Religion Census to construct 13 county-level variables that measure the socioeconomic and demographic factors discussed above that contribute to intergroup prejudice, as well as other indicators associated with radicalization at the community level by previous studies. These include measures of population size, population density, poverty, unemployment, and high school completion rates, per capita welfare spending, nonviolent crime rate, and ethnic homogeneity, as well as the percentage of people in each county who are Muslim, unmarried males, adolescents, and born outside the United States.

Although the observational nature of our study inhibits rigorous causal inference, we attempt to identify the relationship between anti-Muslim and pro-ISIS search rates using a two-stage least-squares model that leverages the number of casualties of U.S. soldiers killed in Iraq and Afghanistan from each U.S. county as an instrument for anti-Muslim



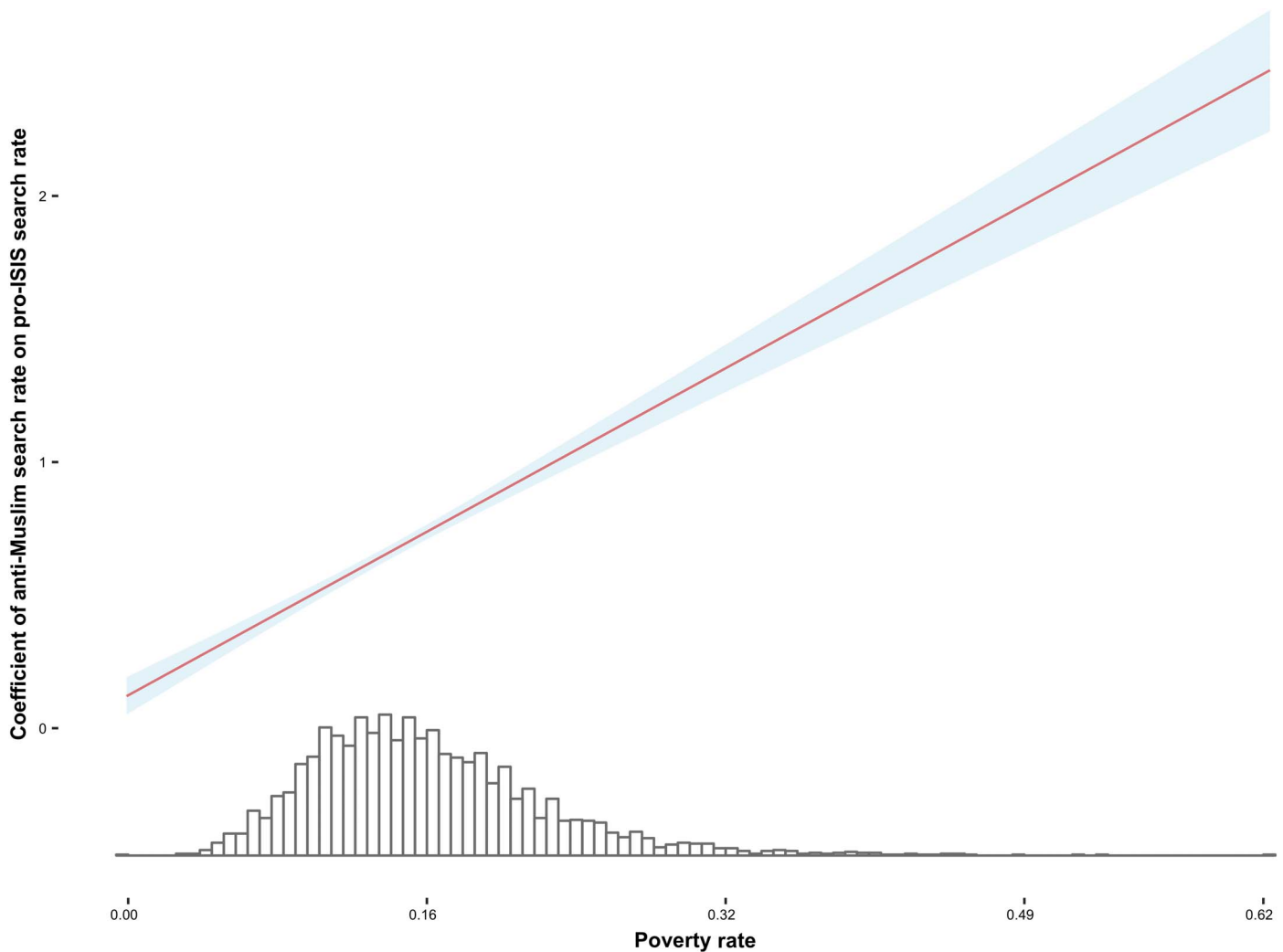
**Fig. 2. Interaction between anti-Muslim Internet search rate and ethnic homogeneity index on pro-ISIS Internet search rate in 3099 U.S. counties, 2014–2016.** Red line describes standardized coefficients from multivariate linear model, and blue band describes 95% confidence interval. Histogram on the x axis describes the frequency of counties by ethnic homogeneity. The coefficient for anti-Muslim Internet search rate increases as the ethnic homogeneity index grows in size, suggesting stronger associations between anti-Muslim Internet searches and pro-ISIS Internet searches in counties with high ethnic homogeneity compared to those with low ethnic homogeneity.

sentiment. Figure 1 shows standardized coefficients from our model that describe the association between each variable and the volume of pro-ISIS searches at the county level (adjusted for the overall volume of Internet searches). The level of anti-Muslim Internet searches is the strongest predictor of pro-ISIS searches ( $\beta = 0.61$ ,  $P < 0.001$ ). A 1-SD increase in the anti-Muslim search rate is associated with a 0.6-SD increase in the pro-ISIS search rate. Of the remaining indicators, only three approach statistical significance, the racial and ethnic homogeneity index ( $\beta = 0.04$ ,  $P = 0.07$ ), the poverty rate ( $\beta = 0.04$ ,  $P = 0.07$ ), and the proportion of foreign-born residents ( $\beta = 0.04$ ,  $P = 0.08$ ).

To further evaluate our hypotheses that high levels of ethnic homogeneity and poverty intensify the relationship between anti-Muslim and pro-ISIS sentiment in U.S. communities, we ran two additional models with interaction terms for each of these variables. As Figs. 2 and 3 show, the predicted association between anti-Muslim search volume and pro-ISIS search volume increases substantially alongside both the ethnic homogeneity ( $P < 0.001$ ) and poverty rate ( $P < 0.001$ ) of U.S. counties.

## DISCUSSION

Our study has several important limitations. As is true of most studies that use Internet search data, we are unable to verify the relationship between online and offline behavior (20, 21). We cannot determine how many people who type “How to join ISIS” into Internet search engines actually participate in violent extremism—yet, we present evidence below (see Materials and Methods: validation with offline measures of ethnic discrimination and radicalization) that these online measures do have significant positive associations with existing offline measures of radicalization. At the same time, it is likely that some who Google “How to join ISIS” are interested in learning about the group’s recruitment process to disrupt it. As we discuss below, these individuals might include intelligence officials, law enforcement officers, investigative journalists, or even anti-Muslim activists themselves. Another significant limitation of our study is that recent studies indicate that many ISIS recruits are approached via social media or instant messaging applications, which we do not study here. This factor, coupled with reportedly high levels of Internet



**Fig. 3. Interaction between anti-Muslim Internet search rate and poverty rate on pro-ISIS Internet search rate in 3099 U.S. counties, 2014–2016.** Red line describes standardized coefficients from multivariate linear model, and blue band describes 95% confidence interval. Histogram on the x axis describes the frequency of counties by poverty rate. The coefficient for anti-Muslim Internet search rate increases as the poverty rate increases suggesting stronger associations between anti-Muslim Internet searches and pro-ISIS Internet searches in counties with high poverty rates compared to those with low poverty rates.

surveillance by the intelligence community, suggests that many people inspired by ISIS might not express their desire to support the organization in their online searches. Our model may therefore be most useful to explain the radicalization process within communities where people cannot identify ways to connect with extremist organizations without querying an Internet search engine. This possibility may further explain why we found that the effect of anti-Muslim sentiment on radicalization is highest in counties with high levels of ethnic homogeneity, where groups of extremists are isolated from organizations they might otherwise be able to approach offline. Future studies are needed to better understand the relationship between online and offline recruitment processes and how Internet searches fit into the broader process of Internet-based recruitment.

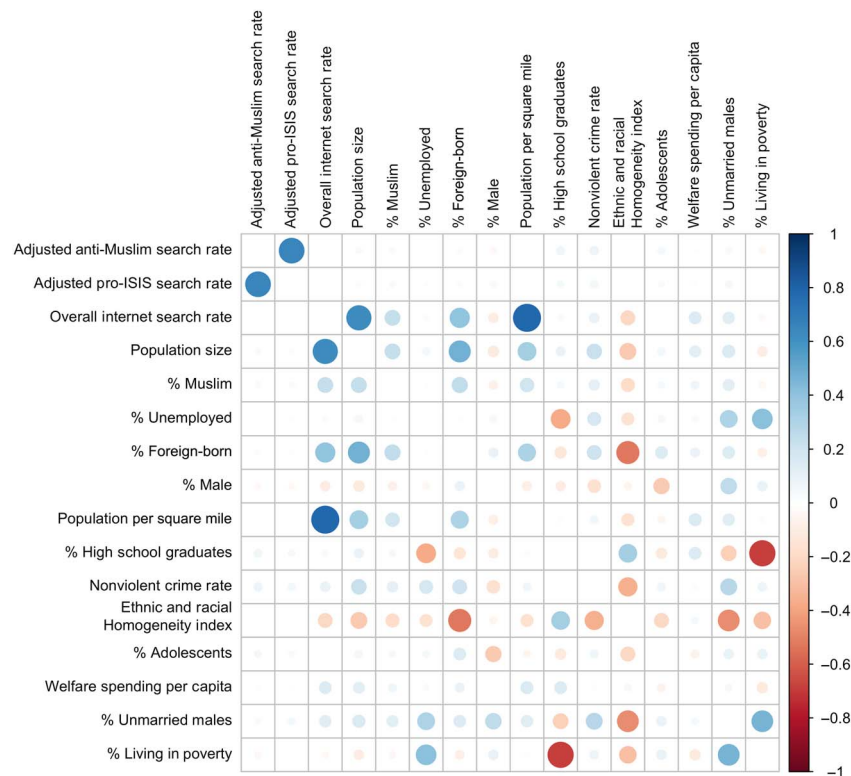
Despite these limitations, this study has important implications for the study of terrorism, ethnic discrimination, and intergroup relations. Although the process of radicalization undoubtedly involves multiple factors, ours is the first study to present empirical evidence of an association between ethnic discrimination and radicalization at the community level. Our theoretical model may provide insight into recent

domestic terrorism cases such as the arrest of Muhammad Dakhalla and Jaelyn Young—a couple of Palestinian and African-American origin, respectively—who were arrested for attempting to travel to Syria to join ISIS in 2016. These two individuals hailed from a region of Mississippi with some of the highest levels of ethnic homogeneity and anti-Muslim sentiment in the country, according to our analysis. Our findings might also help explain recent cases of terrorists who do not describe themselves as Muslims. Dylann Roof, who murdered nine African-Americans at a church in Charleston, South Carolina in June 2015, was reportedly motivated by his belief that white men were becoming an imperiled minority. Roof was also reportedly raised for much of his life in a community where African-Americans are a majority. At the same time, many convicted radicals do not describe racial or ethnic prejudice as motivating factors for their behavior, so more systematic study of other cases—and especially those in other countries—are needed to evaluate the generalizability of our findings.

Nevertheless, our study has significant implications for counterterrorism and immigration policy. Although elected officials routinely

**Table 1. Descriptive statistics.**

Indicator	N	Mean	SD	Minimum	Maximum
Pro-ISIS search rate	3146	2.41	7.65	0	180
Normalized pro-ISIS search rate	3099	19.66	264.91	-698.86	4964.53
Anti-Muslim search rate	3146	7.11	12.28	0	70
Normalized anti-Muslim search rate	3099	87.26	819.83	-1397.72	9929.06
Population size	3146	98,139.08	312,766.60	0	9,818,605
Percent Muslim	3143	0.002	0.01	0	0.29
Unemployment rate	3146	0.09	0.03	0	0.3
Percent foreign-born	3146	0.04	0.06	0	1
Percent male	3146	0.5	0.03	0.43	1
Population per square mile	3146	259.08	1723.36	0	69,467.50
High school completion rate	3146	0.82	0.08	0	0.99
Nonviolent crime rate	3143	0.04	0.03	0	0.6
Ethnic homogeneity index	3146	0.73	0.19	0	0.99
Percent adolescents	3143	0.14	0.02	0	0.24
Welfare spending per capita	3141	0.08	0.15	0	2.59
Percent unmarried males	3143	0.41	0.08	0.1	0.76
Poverty rate	3146	0.16	0.06	0	0.62



**Fig. 4. Correlation matrix for all variables from multivariate two-stage least-squares model.** Blue circles describe variables that have a positive correlation with each other, and red circles describe those that have negative correlations with each other. The size of the circles corresponds to the magnitude of the bivariate correlation.

promote counterterrorism policies that target Muslims more than other groups, our findings indicate that these policies may make communities more vulnerable to radicalization if they are interpreted as discriminatory or unfair. Moreover, our analyses indicate that restrictions on immigration and refugee resettlement may accelerate the cyclical relationship between ethnic discrimination and radicalization, since ethnic diversity appears to mitigate the association between ethnic discrimination and radicalization—at least insofar as it can be analyzed via Internet search data. Compared to many regions of Europe that have high levels of radicalization and ethnic homogeneity, it is possible that the ethnic diversity of the United States is protective against radicalization because communities are less prone to organize along binary identity categories that pit a unified “us” against an alien “them” (22, 23). It is further possible that ethnically diverse communities experience less radicalization because positive intergroup contact reduces the prevalence of ethnic discrimination in turn.

## MATERIALS AND METHODS

### Measuring ethnic discrimination and radicalization using Internet search data

We collected data that describe the average monthly search volume for pro-ISIS and anti-Muslim phrases on Google and other leading search engines via the Keyword Planner tool within the Google AdWords service, which is commonly used by advertisers to research advertising campaigns. Google AdWords features several advantages over the more well-known Google Trends website that describes relative increases in search terms over time and across geographic locations: First, Google AdWords reports the absolute ranges of search volume for terms or search phrases instead of a measure that describes search volume relative to its highest point over time. Second, Google AdWords data describe search behavior at the county and city level, whereas Google Trends only provides data for countries, states or provinces, or metropolitan statistical areas. Third, Google AdWords data describe search volume ranges for search terms with low frequency, whereas Google Trends only provides estimates for high-volume searches. Finally, unlike Google Trends, Google AdWords data describe not only the search volume for terms or phrases typed into Google.com but also other leading Internet search engines.

At the same time, Google AdWords data have two major limitations. First—and perhaps most importantly—the data generation process was highly opaque. Google did not provide detailed information about the way it produces the “average monthly search volume” metric we used in this study. Second, Google provided this metric in the form of 10-point ranges, rather than raw figures.

For each county in the United States, we queried Google’s keyword planner for the average monthly search range of the following pro-ISIS phrases: “How to join ISIS,” “How to join the Islamic State,” “How to support ISIS,” and “How to support the Islamic State.” We selected these phrases because we reasoned that very few people would type them into Google or another search engine unless they were planning to join or otherwise support ISIS. Possible exceptions include investigative journalists or intelligence personnel, who may type these searches into Google to research ISIS recruitment tactics. To account for the latter, we conducted sensitivity tests that dropped the greater Washington D.C. area from our analysis, which did not produce substantively different results. Our search phrases also included the negative keyword “security” to prevent confusion between the so-called Islamic State and the Washington D.C.–based Institute for Science

and International Security, which share the same acronym. To measure anti-Muslim animus, we queried Google AdWords for the average monthly search volume of the following search phrases in each U.S. county: “Muslims are terrorists,” “Muslims are bad,” “Muslims are dangerous,” and “Muslims are evil.” Once again, we chose these terms for their unequivocal character expressing anti-Muslim animus. Table 1 provides descriptive statistics for the raw search range data we obtained from Google for both pro-ISIS and anti-Muslim search phrases (combined).

A third major limitation of our data source is that Google recently discontinued users’ ability to obtain search range estimates for exactly worded phrases such as “How to join ISIS.” When we collected our data, Google AdWords enabled this data collection, but the AdWords service now only provides estimates for these phrases based on all combinations of keywords in each search phrase. Hence, a query for the phrase “Muslims are evil,” at present, will return the same search volume as “Are Muslims evil?” even though the order of the words in such phrases can change the meaning of the metric entirely. Unfortunately, this recent change prevents us from further validating our sample by comparing data extracted at multiple times to determine whether our results are sensitive to shifts in the procedures used to create search estimates or possible issues related to the sampling procedures Google uses to produce these estimates.

### Adjusting for Internet penetration

To account for the likelihood that both anti-Muslim and pro-ISIS searches would be more common in areas with high rates of Internet connectivity, we first weighted Internet search volume for the phrases described above by the ratio of high-speed Internet connections to households in each county using data from the U.S. Federal Communications Commission. We reported the results of our model with this measure in fig. S1. Unfortunately, data on Internet penetration were not available for 215 U.S. counties, so we used a separate measure derived from Google AdWords itself. We collected data on the number of searches for the term “weather” in each county as a gauge of overall search activity and divided the average monthly volume of anti-Muslim and pro-ISIS search terms (which Google reports in increments of 10) by the standardized weather search rate for each county. Table 1 provides descriptive statistics for these normalized pro-ISIS and anti-Muslim search rate measures.

### Algorithmic confounding

As Lazer *et al.* (24) show, the popular Google Flu Trends tool—which can be used to track the spread of influenza via Internet search behavior—consistently overestimated the spread of the disease during the first few years of its existence. One explanation for this disparity, according to Lazer *et al.*, is “blue team” dynamics or properties of Google’s own search algorithm that might distort search estimates. Google users searching for products to treat a common cold, for instance, may see “recommended search” links that discuss the flu or how to prevent getting the flu. If users who do not have the flu click on these links, estimates of influenza prevalence will be exaggerated. Although this is an obvious concern for a study of influenza prevalence, it is *prima facie* unlikely that Google’s search algorithm would recommend anti-Muslim search links to those who type “How to join ISIS” into their interface or vice versa—not only because advertisers targeting one of these audiences are presumably unlikely to target the other but also because Google’s Terms of Service prevent advertisements that include derogatory speech.

Because investigatory journalists have recently suggested that it is still possible to purchase advertising campaigns on Google that use derogatory language for Jews and African-Americans, we took additional steps to assess algorithmic confounding using Google Correlate, a tool that allows users to examine which search terms are most frequently typed into Google's search engine alongside each other. Unfortunately, Google Correlate only provides estimates of search term co-occurrence for high-volume searches. Instead of using a search phrase such as "How to join ISIS," which has a very low search volume, we used the term "Infidel," which is frequently used by extremist groups to refer to nonextremists. Instead of a search phrase such as "Muslims are evil," we used the terms "Haji" and "Mussie," which are derogatory terms for Muslims. As table S1 shows, none of the top 50 search terms that are routinely searched for alongside these terms would indicate substantial risk of algorithmic confounding. This also indicates that few—if any—anti-Muslim activists google pro-ISIS search terms.

We thank an anonymous reviewer for identifying another type of algorithmic confounding that might occur because of Google's autocomplete search function. That is, someone who wants to use Google to search for "I hate chocolate" might see the suggestion "I hate Muslims" after typing the first two words, as one of Google's search algorithms tries to predict what the user might be interested in and then performs a search for the latter phrase out of curiosity. Upon researching the issue, we discovered an official Google website, which states that the autocomplete function "removes predictions that include language that denigrates or insults individuals or groups on the basis of race, ethnic origin, religion, disability, gender, age, nationality, veteran status, sexual orientation, or gender identity." The link further states that Google "remove(s) predictions that include graphic descriptions of violence or advocate violence generally." Concerned that Google may have adopted this policy in response to recent events, we consulted the Internet Wayback Machine to view an earlier version of this Web page that includes very similar language. Although it is presumably easier for Google to exercise oversight over its autocomplete function than ad campaign keywords—which can be suggested by anyone—we nevertheless performed searches that began with "How to join," "I hate," and "Muslims are," and did not receive autocomplete suggestions for any of the search phrases used in our analyses. Once again, however, we cannot rule out the possibility that Google has removed these suggestions after we finished our data collection, as a defensive response to recent events indicating that Google is not always effective at enforcing its anti-hate speech policies. We also cannot rule out the possibility that these autocomplete phrases might appear for people with socioeconomic or demographic characteristics that are different than our own.

### Validation with offline measures of ethnic discrimination and radicalization

Another strategy for detecting the existence of algorithmic confounding and further testing the validity of our measures is to compare the Internet search data we analyzed to other data sources (24, 25). Unfortunately, available estimates of pro-ISIS and anti-Muslim sentiment pale in comparison to the measures of influenza prevalence collected by the Centers for Disease Control and Prevention that were used by Lazer *et al.* to assess algorithmic confounding. Nevertheless, we compared our pro-ISIS search volume measure to the most comprehensive database of terrorist incidents in the United States currently available (26). This database includes a count of the number of individuals who committed or supported acts of terrorism in different geographic locations between 2001 and 2016. We used these data to count the num-

ber of terrorist incidents within each county during our study period and found a significant association between this indicator and our pro-ISIS search volume measure (bivariate,  $\beta = 0.112$ ,  $P < 0.001$ ; multivariate,  $\beta = 0.022$ ,  $P = 0.074$ ). Second, we compared our anti-Muslim search volume measure to data collected by the American Civil Liberties Union that describe the prevalence of mosque controversies between 2008 and 2016 (27). Once again, we used these data to count the number of anti-Muslim incidents by county during our study period and found a strong and highly significant association between this indicator and our anti-Muslim search volume measure (bivariate,  $\beta = 0.117$ ,  $P < 0.001$ ; multivariate,  $\beta = 0.068$ ,  $P < 0.001$ ). Unfortunately, other public opinion surveys that include measures of anti-Muslim sentiment, as well as support for violence against civilians from the Pew Research Center, do not have the geographic resolution necessary to perform a meaningful analysis.

### Description of independent variables

Table 1 shows descriptive statistics for each of the variables in our model. In addition to the socioeconomic and demographic control variables that we measured to evaluate group threat theory, we included other covariates that previous research has linked to radicalization. Figure 4 describes the correlation between the 13 control variables as well as anti-Muslim and pro-ISIS search rates without adjustment for overall search volume. Red values indicate that variables have a negative correlation to each other, and blue values indicate that they have a positive relationship. The size of each circle corresponds to the strength of this correlation.

### Demographic factors

Previous studies have identified several demographic factors associated with the risk of radicalization at the community level. First, the population size and density of a given community have been shown to create greater risk—not only because larger numbers of people increase the overall probability of radicalization but also because radical ideas spread more easily across groups of people who live in close proximity to each other—and particularly those in urban areas (14, 28, 29). We measured these factors using data that describe the overall size of the residential population of each county from the 2010 U.S. Census, as well as the population per square mile from the same source.

To measure the ethnic homogeneity of each county, we collected data that describe the population size of six racial and ethnic groups within each county from the 2009 American Community Survey. These data allowed us to create an ethnic homogeneity index ( $\alpha$ ) for each county in our sample using a Herfindahl-style index as follows

$$\alpha = \sum_{i=1}^N s_i^2$$

where  $s_i$  is the share of the population made up of each ethnic or racial group  $i$ . Higher levels of  $\alpha$  thus indicate higher levels of ethnic homogeneity.

Although recent studies indicate that women play an increasingly important role within terrorist networks, the vast majority of people indicted upon charges of domestic terrorism are male (29). Previous studies further indicate that radicalization is most prevalent among adolescents or unmarried males who turn to extremism out of sexual frustration or due to a broader sense of powerlessness (30). Our models therefore include a variable that describes the percentage of the population that is between 10 and 20 years old, following the World Health Organization's definition of adolescence, as well as a variable capturing the proportion of men 15 years and older who are either (i) never

married, (ii) married but living separately from their spouse, or (iii) divorced within each county. Finally, our models include a variable from the 2010 U.S. Census that describes the percentage of the population in each county that is born outside the United States, since reported rates of radicalization are far lower inside the United States than in other countries (31).

### Socioeconomic factors

Evidence that economic factors or human capital drive radicalization is mixed (32–35). Although some studies indicate that economic deprivation can create a frustration-aggression reaction, sociodemographic profiles of convicted violent extremists include those who are very wealthy and highly educated. Nevertheless, we included the civilian labor force unemployment rate by county from the 2010 U.S. Census, as well as the percentage of people of all ages in poverty and the percentage of the population with a high school degree in 2009 from the American Community Survey. Finally, we included a measure of local government direct general expenditures for public welfare per capita by county from the same survey from 2002, which is the most recent year for which these data are available.

### Size of Muslim population

Although there is ample evidence that more violent extremist acts are committed by non-Muslims than by Muslims in the United States (26, 31), we collected estimates of the size of the Muslim population in a given county because we focused on radicalization inspired by ISIS, which describes itself as an Islamic organization. The variable described in the main text of our article is an estimate of the Muslim population created by the U.S. Religion Census coordinated by the Association of Statisticians of American Religious Bodies. These data rely on the existence of formal religious organizations, such as mosques or Islamic community centers, as a sampling frame to estimate the size of the Muslim population and therefore likely underestimate the size of this population in counties where these institutions do not exist or are not formally organized.

Unfortunately, official estimates of the size of the Muslim population are not currently available because the U.S. Census and the American Community Survey do not collect information on religious affiliation of respondents. Because of the aforementioned limitations of the U.S. Religion Census data, we created our own estimate of the Muslim population of each county by collecting the overall volume of searches for the term “halal” within each county. Although it is, of course, possible that non-Muslims might search for this term to better understand the Muslim religion or out of consideration for Muslim guests, we reasoned that this measure would nevertheless add value to our study in the absence of more accurate population estimates. This reasoning is supported by further analyses, which showed that the search volume for “halal” is positively and significantly associated with the U.S. Religion Census measure of the Muslim populations. Figure S2 reproduces the regression models in the main text using the Google-based measure of the Muslim population in lieu of the U.S. Religion Census data. The inclusion of this indicator does not alter the effect of anti-Muslim searches on pro-ISIS searches, and the indicator itself is not a significant predictor of pro-ISIS searches.

### Nonviolent crime

Some studies indicate that violent extremists commit petty crimes before becoming radicalized—or become radicalized while serving time in prison for committing these crimes (28). Therefore, we included a measure of the nonviolent crime rate by county in 2008, including burglary, larceny, grand theft, automobile theft, and property crimes, from the American Community Survey.

### Detecting causal heterogeneity with machine learning

To further assess the possibility of omitted variable bias, we used machine learning models to analyze an additional 27,281 county-level variables that are currently available from the U.S. Census. We used a recursive partitioning technique to detect causal heterogeneity among subgroups of counties that combines LASSO and *k*-fold cross-validation (36). This analysis identified no additional variables that are consistently associated with pro-ISIS searches across repeated partitions of the data.

### Estimation strategy

#### Two-stage least-squares estimation

Our primary identification strategy is to leverage an instrumental variable and use two-stage least-squares regression to identify the causal effect of anti-Muslim searches on pro-ISIS searches fitting the following equation

$$Y_i = \beta_0 + \beta D_i + X_i + \varepsilon_i$$

where *Y* is the volume of pro-ISIS searches adjusted for overall search volume, *D* is the volume of anti-Muslim searches similarly adjusted and instrumented through the number of U.S. soldier casualties in Iraq and Afghanistan before 2014, *X* is the vector of control variables, and  $\varepsilon$  is the error term. We obtained the casualty data from the Defense Manpower Data Center’s Defense Casualties Analysis System for Operations Enduring Freedom, Iraqi Freedom, New Dawn, Inherent Resolve, and Freedom’s Sentinel (37). Because this variable was not normally distributed, we used the square root of U.S. soldier casualties by county to predict the volume of anti-Muslim searches in the first stage of the two-stage least-squares regression. Model diagnostics further indicated that this transformation was warranted— $R^2 = 0.58$  with the square root transformation and 0.52 without transformation.

Our estimation strategy assumes that the only causal pathway that links the number of casualties of U.S. soldiers in a given county to pro-ISIS sentiment therein operates through the anti-Muslim sentiment generated by these casualties. Although this instrumental variable greatly exceeds generally accepted values of the weak instrument test ( $F = 28.69$ ,  $P < 0.001$ ), it is nevertheless possible that pro-ISIS or anti-American sentiment among Muslim-Americans causes anti-Muslim sentiment—although there is very little evidence that these sentiments are widespread (2). Nevertheless, additional research is needed to further identify the causal influence of anti-Muslim sentiment on pro-ISIS sentiment—although we believe that the mere strength of the association between anti-Muslim and pro-ISIS search rates reported above is cause for major concern about the relationship between intergroup prejudice and radicalization, regardless of the direction of causality.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/6/eaao5948/DC1>

fig. S1. Standardized coefficients from two-stage least-squares regression model predicting pro-ISIS search rate normalized by Internet penetration in 2921 U.S. counties, 2014–2016.

fig. S2. Standardized coefficients from two-stage least-squares regression model predicting pro-ISIS Internet search rate normalized for overall search activity in 3146 U.S. counties, 2014–2016.

table S1. Results of Google Correlate searches to examine algorithmic confounding.

table S2. Individuals who were indicted for terrorist activity or completed an attack by state.

table S3. Individuals who were indicted for terrorist activity or completed an attack by year.



## REFERENCES AND NOTES

1. E. Miller, *Overview: Terrorism in 2016* (National Consortium for the Study of Terrorism and Responses to Terrorism, 2015).
2. C. Kurzman, *The Missing Martyrs: Why There are so Few Muslim Terrorists* (Oxford Univ. Press, 2011).
3. T. Hegghammer, Should I stay or should I go? Explaining variation in Western jihadists' choice between domestic and foreign fighting. *Am. Polit. Sci. Rev.* **107**, 1–15 (2013).
4. M. Sageman, *Leaderless Jihad: Terror Networks in the Twenty-First Century* (University of Pennsylvania Press, 2008).
5. C. Bail, *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream* (Princeton Univ. Press, 2015).
6. K. Bansak, J. Hainmuller, D. Hangartner, How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science* **22**, aag2147 (2016).
7. L. Blumer, Race prejudice as a sense of group position. *Pac. Sociol. Rev.* **1**, 3–7 (1958).
8. H. Tajfel, *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations* (Academic Press in cooperation with European Association of Experimental Social Psychology, 1978).
9. L. D. Bobo, Prejudice as group position: Microfoundations of a sociological approach to racism and race relations. *J. Soc. Issues* **55**, 445–472 (1999).
10. H. M. Blalock Jr., Per cent non-white and discrimination in the South. *Am. Sociol. Rev.* **22**, 677–682 (1957).
11. T. F. Pettigrew, Reactions toward the new minorities of Western Europe. *Annu. Rev. Sociol.* **24**, 77–103 (1998).
12. G. W. Allport, *The Nature of Prejudice* (Addison-Wesley Pub. Co., 1955).
13. M. B. Altier, J. Horgan, C. Thoroughgood, In their own words? Methodological considerations in the analysis of terrorist autobiographies. *J. Strategic Secur.* **5**, 85–98 (2012).
14. J. Horgan, N. Shortland, S. Abbasciano, S. Walsh, Actions speak louder than words: A behavioral analysis of 183 individuals convicted for terrorist offenses in the United States from 1995 to 2012. *J. Forensic Sci.* **61**, 1228–1237 (2016).
15. T. F. Pettigrew, R. W. Meertens, Subtle and blatant prejudice in western Europe. *Eur. J. Soc. Psychol.* **25**, 57–75 (1995).
16. S. Stephens-Davidowitz, The cost of racial animus on a black candidate: Evidence using Google search data. *J. Public Econ.* **118**, 26–40 (2014).
17. A. D. Simmons, L. D. Bobo, Can non-full-probability internet surveys yield useful data? A comparison with full-probability face-to-face surveys in the domain of race and social inequality attitudes. *Sociol. Methodol.* **45**, 357–387 (2015).
18. N. F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, H. Qi, N. Velasquez, P. Manrique, D. Johnson, E. Restrepo, C. Song, S. Wuchty, New online ecology of adversarial aggregates: ISIS and beyond. *Science* **352**, 1459–1463 (2016).
19. J. Adams, V. J. Roscigno, White supremacists, oppositional culture and the World Wide Web. *Soc. Forces* **84**, 759–778 (2005).
20. S. A. Golder, M. W. Macy, Digital footprints: Opportunities and challenges for online social research. *Annu. Rev. Sociol.* **40**, 129–152 (2014).
21. C. A. Bail, The cultural environment: Measuring culture with big data. *Theory Soc.* **43**, 465–482 (2014).
22. L. Bobo, V. L. Hutchings, Perceptions of racial group competition: Extending Blumer's theory of group position to a multiracial social context. *Am. Sociol. Rev.* **61**, 951–972 (1996).
23. C. A. Bail, The configuration of symbolic boundaries against immigrants in Europe. *Am. Sociol. Rev.* **73**, 37–59 (2008).
24. D. Lazer, R. Kennedy, G. King, A. Vespignani, The parable of Google Flu: Traps in big data analysis. *Science* **343**, 1203–1205 (2014).
25. D. Ruths, J. Pfeffer, Social media for large studies of behavior. *Science* **346**, 1063–1064 (2014).
26. C. Kurzman, *Muslim-American Involvement with Violent Extremism, 2016* (Triangle Center on Terrorism and Homeland Security, 2017).
27. American Civil Liberties Union, "Nationwide Anti-Mosque Activity," March 2018; [www.aclu.org/issues/national-security/discriminatory-profiling/nationwide-anti-mosque-activity?redirect=map/nationwide-anti-mosque-activity](http://www.aclu.org/issues/national-security/discriminatory-profiling/nationwide-anti-mosque-activity?redirect=map/nationwide-anti-mosque-activity).
28. P. Gill, J. Horgan, P. Deckert, Bombing alone: Tracing the motivations and antecedent behaviors of lone-actor terrorists. *J. Forensic Sci.* **59**, 425–435 (2014).
29. P. Manrique, Z. Cao, A. Gabriel, J. Horgan, P. Gill, H. Qi, E. M. Restrepo, D. Johnson, S. Wuchty, C. Song, N. Johnson, Women's connectivity in extreme networks. *Sci. Adv.* **2**, e1501742 (2016).
30. M. Juergensmeyer, *Terror in the Mind of God: The Global Rise of Religious Violence* (University of California Press, 2000).
31. D. Schanzer, C. Kurzman, J. Toliver, E. Miller, "The challenge and promise of using community policing strategies to prevent violent extremism" (no. 249674, Triangle Center on Terrorism and Homeland Security, 2016).
32. A. B. Krueger, *What Makes a Terrorist: Economics and the Roots of Terrorism* (Princeton Univ. Press, 2008).
33. A. Abadie, Poverty, political freedom, and the roots of terrorism. *Am. Econ. Rev.* **96**, 50–56 (2006).
34. A. B. Krueger, D. Laitin, Kto Kogo?: A cross-country study of the origins and targets of terrorism, terrorism economic development, and political openness, in *Terrorism, Economic Development, and Political Openness*, P. Keefer, N. Loayza, Eds. (Cambridge Univ. Press, 2007).
35. A. B. Krueger, J. Malečková, Education, poverty, and terrorism: Is there a causal connection. *J. Econ. Perspect.* **17**, 119–144 (2003).
36. S. Athey, G. Imbens, Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7353–7360 (2016).
37. Defense Casualty Analysis System, "Conflict casualties," 2016; <https://dcas.dmdc.osd.mil/dcass/pages/casualties.xhtml>.

**Acknowledgments:** We thank A. Kadivar, C. Kurzman, A. Wimmer, and three anonymous reviewers who provided comments on this paper. **Funding:** This publication was made possible, in part, by a grant from the Carnegie Corporation of New York to the first author (award no. G-F-17-54413). The statements made and views expressed are solely the responsibility of the authors. **Author contributions:** C.A.B. and F.M. collected the data, analyzed the data, and wrote the manuscript. P.D. contributed to data analysis. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All replication materials necessary to evaluate the conclusions reported in this paper and the Supplementary Materials are available at the following link: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GZSYCO>

Submitted 7 August 2017

Accepted 24 April 2018

Published 6 June 2018

10.1126/sciadv.aao5948

**Citation:** C. A. Bail, F. Merhout, P. Ding, Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties. *Sci. Adv.* **4**, eao5948 (2018).

## Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties

Christopher A. Bail, Friedolin Merhout and Peng Ding

*Sci Adv* 4 (6), eaao5948.  
DOI: 10.1126/sciadv.aao5948

ARTICLE TOOLS	<a href="http://advances.sciencemag.org/content/4/6/eaao5948">http://advances.sciencemag.org/content/4/6/eaao5948</a>
SUPPLEMENTARY MATERIALS	<a href="http://advances.sciencemag.org/content/suppl/2018/06/04/4.6.eaao5948.DC1">http://advances.sciencemag.org/content/suppl/2018/06/04/4.6.eaao5948.DC1</a>
REFERENCES	This article cites 24 articles, 5 of which you can access for free <a href="http://advances.sciencemag.org/content/4/6/eaao5948#BIBL">http://advances.sciencemag.org/content/4/6/eaao5948#BIBL</a>
PERMISSIONS	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2018 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).