Identifying Spatial Disparities in Online Discussions About Clinical Trials for COVID-19 Vaccines

Raphael E. Cuomo, Ph.D., M.P.H.; Zhuoran Li, M.S.; Qing Zu, M.A.S.; Tim K. Mackey, Ph.D., M.A.S.

Introduction

- Trust and participation in clinical trials is necessary to prevent and adequately respond to disease outbreaks. Analysis of online engagement with clinical trial topics can characterize how communities perceive clinical trials.

Objectives

- This study examined spatial disparities in online discussions about COVID-19-related clinical trials in the United States using geocoded data from Twitter.

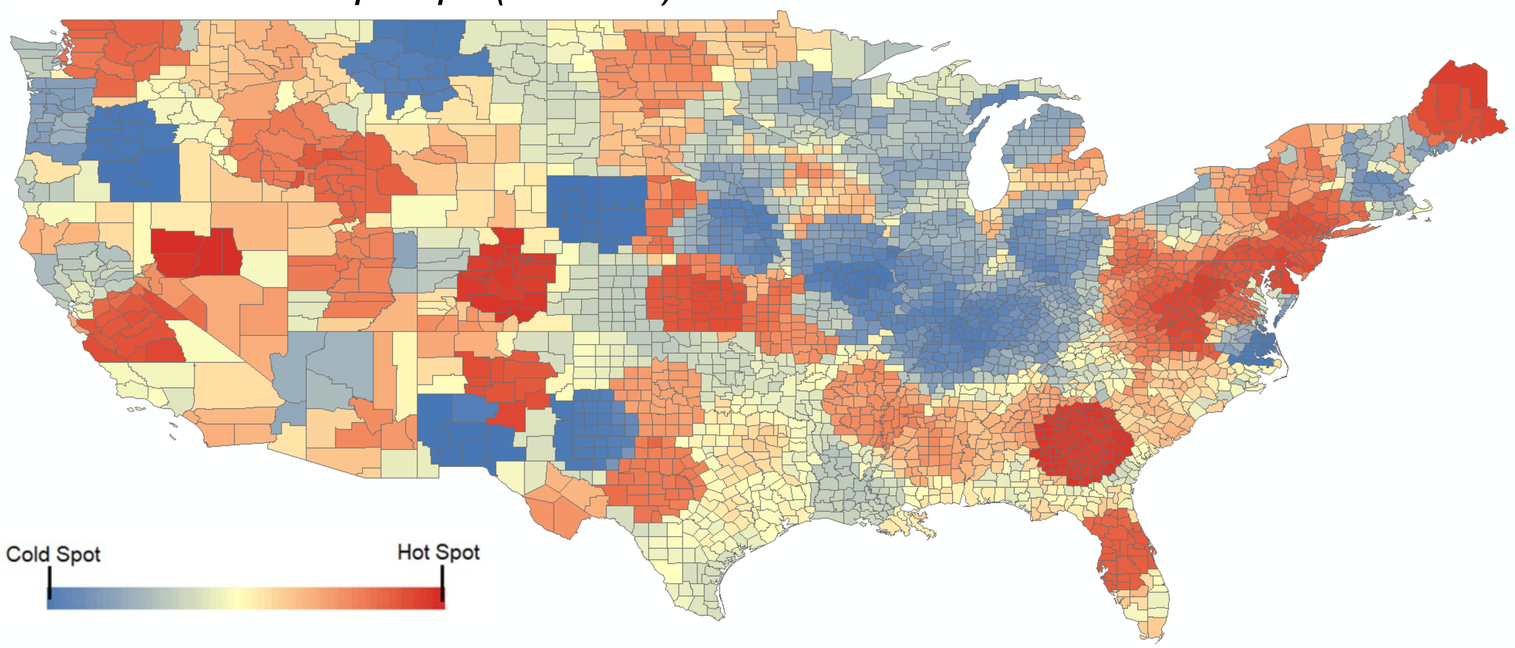
Methods

- Unsupervised topic modeling identified topics related to COVID-19 and clinical trials. A topic cluster output with a high volume of posts related to clinical trials, as well as a comparison cluster which was not about trials, was selected for geospatial analysis. After population-normalization and feature scaling, the difference between the spatial distribution of trials discussions, after adjusting for non-trials discussion, was used to compute the Getis Ord G statistic for clustering, and Pearson's correlation coefficients was used to assess correlations with demographics.

Table 1 – Linear regression model of Δ(trials-nontrials) tweets using a backward selection algorithm for covariate selection (Wald χ2)

Covariate	β	SE	*t*	*p*
Intercept	0.007	0.006	1.177	0.239
Age <5 /capita	-0.093	0.054	-1.711	0.087
Age 10-14 /capita	0.187	0.058	3.207	0.001
Age 15-19 /capita	-0.089	0.035	-2.513	0.012
Age 45-54 /capita	-0.095	0.032	-3.007	0.003
Age 55-64 /capita	0.131	0.034	3.886	<0.001
Age 65-74 /capita	-0.135	0.028	-4.785	<0.001

Figure 1 – Hot spot analysis for clustering of the difference between trials-related tweets per capita and non-trials-related tweets per capita (Getis Ord G)



Results

- County-level discussion about trials was significantly positively associated with concentration of population age 5-9, and was negatively associated with concentrations of population age 65-74, 75-84, and 85+. Statistically significant hot spots were detected in New Mexico, Nebraska, and Texas.

Table 2 – Bivariate correlations between Δ(trials-nontrials) tweets and normalized age group composition at the county level

Demographic Covariates	Pearson's rho	*p*
Age <5 /capita	0.016	0.382
Age 5-9 /capita	0.036	0.045
Age 10-14 /capita	0.035	0.053
Age 15-19 /capita	-0.02	0.265
Age 20-24 /capita	-0.022	0.222
Age 25-34 /capita	0.018	0.309
Age 35-44 /capita	0.029	0.107
Age 45-54 /capita	-0.002	0.893
Age 55-64 /capita	0.003	0.864
Age 65-74 /capita	-0.039	0.030
Age 75-84 /capita	-0.045	0.011
Age 85+ /capita	-0.053	0.003
Black / capita	0.018	0.310
Hispanic / capita	-0.009	0.607
White / capita	-0.021	0.230
Asian / capita	-0.012	0.509
Native American / capita	0.009	0.620

Conclusions

- Clusters with statistically significant departures from the overall distribution were evident, indicating uneven interest in clinical trials. This may be partly explained by differences in age group concentrations. These results suggest consequences for clinical trial participation and potential for lacking applicability of findings to elderly populations. Further study should identify the influence of sub-themes, such as whether the topic of pediatric vaccination may drive observed positive association with young age groups.