RESEARCH CORRESPONDENCE

Increased Internet Search Interest for GI Symptoms May Predict COVID-19 Cases in US Hotspots

Imama Ahmad,[*] Ryan Flanagan,[‡] and Kyle Staller[§]

*Department of Medicine, North Shore Medical Center, Salem, Massachusetts; ‡Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; and §Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts

Google Trends is an online tool that allows measurement of search term popularity on Google, spatially and temporally. While not an epidemiological tool for determining incidence, it can estimate the popularity of a certain disease by search volume over time.[1,2] It has previously correlated well with infectious disease incidence and has demonstrated utility in disease forecasting, especially with influenza data.[3] We utilized Google Trends to investigate whether search interest in common gastrointestinal (GI) symptoms would correlate with coronavirus disease 2019 (COVID-19) incidence data.

Methods

We used Google Trends to measure interest in specific GI symptoms related to COVID-19 along with data from Harvard Dataverse[4] to gauge the true incidence of COVID-19. We analyzed incidence data from 15 states with top, median, and lowest COVID-19 burden for a 13-week period, subject to data availability, from January 20 to April 20, 2020.

We identified common GI symptoms attributed to COVID-19 from previous studies as search terms, which included ageusia, abdominal pain, loss of appetite, anorexia, diarrhea, and vomiting,[5,6] downloading weekly search data for each of these terms by state. The search volume of each of these terms is known as relative search volume and is assigned according to the number of searches for a term in a particular week compared to the week with the highest number of searches in the selected time period, on a scale of 0–100. The week with the highest number of searches is given a score of 100; other weeks are rated in relation to this highest search week.

We compared the search volume for each of the GI symptoms with reported incidence of COVID-19 for each of the 15 states over a period of 13 weeks, using time-lagged cross-correlations (see Supplementary Methods).

Results

We found that Google search interest in *ageusia, loss of appetite*, and *diarrhea* increased 4 weeks prior to the rise in COVID-19 cases for most states, with maximum correlation estimates of 0.998, 0.871, and 0.748, respectively (Supplementary Table 1). In general, we found that time-lag coefficients became stronger with increasing lag in weeks, highest at 4 weeks.

A lag time of 4 weeks yielded the strongest correlation between symptom search volume and COVID-19 case volume, specifically for *ageusia* (5 states), *diarrhea* (3 states), and *loss of appetite* (1 state). Plots of symptoms vs cases over time demonstrated an increase in search volume followed by an increase in COVID-19 incidence after 3–4 weeks. A representative plot for New York is shown in Figure 1. In median incidence states, ageusia was only significant in 2 of 5 states at the 4-week lag time. Likewise, ageusia was only significant in 1 of 5 low-incidence states.

Other GI symptoms did not consistently correlate with increases in COVID-19 diagnoses.

Discussion

Our results show that Google searches for specific, common GI symptoms correlated with incidence of COVID-19 in the first weeks of the pandemic in 5 states with high disease burden. Specifically, searches for *ageusia* and *loss of appetite* correlated most strongly with the rise in COVID-19 cases in high-incidence states. As the lag time between search volume and COVID incidence increased, correlation increased—with the strongest relationship at 3–4 weeks. Thus, searches for GI symptoms preceded the rise in reported COVID-19 in a predictable fashion, slightly longer than the 1- to 2-week lag time observed in prior studies on influenza.[3] The observed time difference could be related to differences in testing availability, reporting, or longer incubation period of COVID-19 compared with Influenza.

Despite these interesting associations, we were unable to account for potentially confounding variables related to
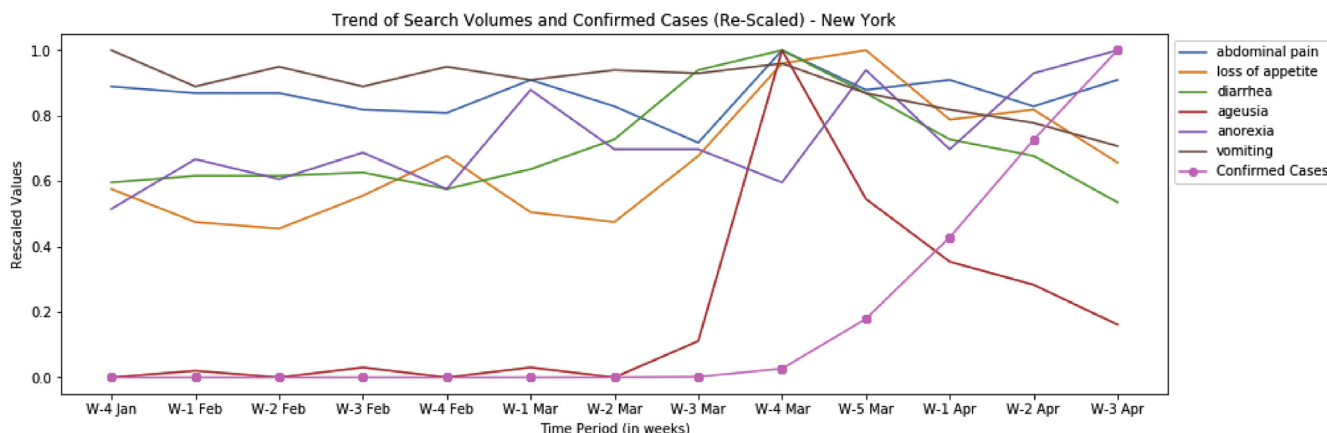
Most current article

Figure 1. Google search volumes relative to new COVID-19 cases for New York in early 2020. The y-axis values (both disease incidence and search volume data) are rescaled from 0 to 1 to ensure ease of comparison. W, week.

COVID-19 incidence and search volume, including demographics, occupational factors, or Internet use patterns. While our study provides information about popular search terms and their relationship to incidence, it is important to note that the relative nature of Google Trends data does not allow for defining specific increased interest thresholds. Further studies are needed to determine peak incidence, seasonality, and the optimal query time frame to generate predictive models for COVID-19.

Our results suggest that increased search volume for common GI symptoms may predict COVID-19 case volume, with 4 weeks as the optimal gap between increase in search volume and increased case load. Because ageusia was popular in all top-incidence states, 2 median incidence states, and 1 low-incidence state, a "dose-response" relationship may exist, with search popularity unlikely to be solely related to increasing national awareness of COVID-19. Our data underscore the importance of GI symptoms as a potential harbinger of COVID-19 infection and suggests that Google Trends may be a valuable tool for prediction of pandemics with GI manifestations.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2020.06.058.

References

1. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res 2018;20:e270.
2. Flanagan R, Kuo B, Staller K. Utilizing Google Trends to assess worlwide interest in irritable bowel syndrome and commonly associated treatments. Dig Dis Sci 2020 May 2 [E-pub ahead of print].
3. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2008;457:1012–1014.
4. US COVID-19 Daily Cases with Basemap. Harvard Dataverse. Version 24.0. Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HIDLTK. Accessed April 20, 2020.
5. Redd WD, Zhou JC, Hathorn KE, et al. Prevalence and characteristics of gastrointestinal symptoms in patients with severe acute respiratory syndrome Coronavirus 2 infection in the United States: A multicenter cohort study. Gastroenterology 2020 Apr 22;S0016-5085(20):30564–3.
6. Pan L, Mu M, Sun Y, et al. Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive cross-sectional multicenter study. Am J Gastroenterol 2020 May;115(5):766–773.

Reprint requests
Address requests for reprints to: Kyle Staller, MD, MPH, Division of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts. e-mail: kstaller@mgh.harvard.edu; fax: (627) 724-7441.

Supplementary Methods

On the basis of highest incidence, New York, New Jersey, California, Massachusetts, and Illinois were used in this analysis. For median incidence states, Alabama, South Carolina, Nevada, Mississippi, and Rhode Island were used. Hawaii, Montana, North Dakota, Wyoming, and Alaska were used as the lowest-incidence states. Daily incidence data were downloaded for each state from the Harvard Dataverse website, which was then converted to weekly new cases to ensure consistency when comparing to Google Trends data. These weekly new cases, labeled "Confirmed Cases," represent the weekly incidence of the virus in a particular state.

We identified common gastrointestinal (GI) symptoms attributed to coronavirus disease 2019 (COVID-19) from previous studies as search terms, which included *ageusia*, *abdominal pain*, *loss of appetite*, *anorexia*, *diarrhea*, and *vomiting*,[1,4] downloading weekly search data for each of these terms by state. The search volume of each of these terms is known as relative search volume, and is assigned according to the number of searches of a particular term in a particular week compared with the week with the highest number of searches in the selected time period, on a scale of 0–100. The week with the highest number of searches is given a score of 100; other weeks are rated in relation to this highest search week. Hence, this does not represent the absolute number of searches, but instead shows the search volume relative to a defined time period.

We compared the search volume for each of the GI symptoms with reported incidence of COVID-19 for each of the 15 states over a period of 13 weeks. These data were used to compute time-lagged cross correlations to gauge the association between GI symptom searches and disease incidence and its direction, magnitude, and significance. We analyzed search volume with assessment of COVID-19 incidence at lag times of 1, 2, 3, and 4 weeks. Incident COVID-19 cases occurring at predictable lag times relative to increased GI symptom search volume assessed whether such search volume data can predict COVID-19 case volume. We plotted search volume data and disease incidence data against time to provide graphical evidence of the correlation and its associated lag time.

Supplementary Table 1. Correlation Between Search Interest and COVID-19 Cases by State

Incidence	State	Lag 1	Lag 2	Lag 3	Lag 4	Search term
High	New York	0.482	0.504	0.739[a]	0.871[a]	loss of appetite
		−0.856[a]	−0.753[a]	−0.489	0.257	vomiting
		0.026	0.269	0.192	0.652	abdominal pain
		0.554	0.280	0.641[a]	−0.214	anorexia
		0.270	0.360	0.523	0.998[a]	ageusia
		0.028	0.186	0.434	0.748[a]	diarrhea
	New Jersey	0.253	0.147	0.296	−0.005	loss of appetite
		−0.698[a]	−0.668[a]	−0.728[a]	−0.276	vomiting
		−0.342	−0.360	0.088	−0.148	abdominal pain
		0.216	0.006	0.055	0.271	anorexia
		0.297	0.379	0.456	0.996[a]	ageusia
		−0.096	0.245	0.276	0.405	diarrhea
	California	0.238	0.038	0.232	0.643	loss of appetite
		−0.599[a]	−0.489	0.244	−0.080	vomiting
		0.074	0.546	0.671[a]	0.556	abdominal pain
		0.225	0.163	−0.086	−0.217	anorexia
		0.381	0.395	0.589	0.985[a]	ageusia
		−0.162	0.066	0.522	0.745[a]	diarrhea
	Massachusetts	0.211	0.189	0.076	0.612	loss of appetite
		−0.468	−0.223	−0.427	−0.408	vomiting
		−0.194	−0.149	0.297	−0.415	abdominal pain
		0.047	0.057	0.503	−0.083	anorexia
		0.235	0.283	0.543	0.980[a]	ageusia
		−0.132	0.076	0.448	0.708[a]	diarrhea
	Illinois	0.732[a]	0.676[a]	0.703[a]	0.590	loss of appetite
		−0.719[a]	−0.808[a]	−0.898[a]	−0.576	vomiting
		0.254	0.227	−0.010	−0.025	abdominal pain
		0.541	0.214	0.028	−0.442	anorexia
		0.286	0.233	0.602	0.992[a]	ageusia
		−0.208	0.152	0.017	0.651	diarrhea
Median	Alabama	−0.166	0.163	0.408	0.392	loss of appetite
		−0.370	0.097	−0.546	0.137	vomiting
		−0.045	0.037	−0.590	0.186	abdominal pain
		0.353	0.314	0.012	0.809[a]	anorexia
		0.633[a]	0.657[a]	0.549	0.274	ageusia
		−0.199	0.126	0.363	0.490	diarrhea
	South Carolina	0.178	−0.109	−0.254	−0.424	loss of appetite
		−0.355	−0.411	−0.350	−0.254	vomiting
		0.166	−0.073	0.074	−0.057	abdominal pain
		0.570	0.323	0.393	0.233	anorexia
		0.048	0.069	0.141	−0.260	ageusia
		0.185	0.318	0.320	0.458	diarrhea
	Nevada	0.047	0.179	−0.041	−0.031	loss of appetite
		−0.474	−0.611[a]	0.027	0.566	vomiting
		0.341	0.250	0.144	0.380	abdominal pain
		0.299	−0.018	0.202	0.224	anorexia
		0.542	0.337	0.345	0.540	ageusia
		−0.040	−0.105	0.140	0.491	diarrhea
	Mississippi	−0.115	−0.315	−0.251	0.407	loss of appetite
		0.095	0.178	−0.190	−0.073	vomiting
		−0.358	0.050	0.432	−0.284	abdominal pain
		0.133	−0.472	0.111	0.389	anorexia
		0.558	0.767[a]	0.806[a]	0.999[a]	ageusia
		0.063	0.161	0.431	0.391	diarrhea
	Rhode Island	0.193	0.206	0.184	0.112	loss of appetite
		−0.457	−0.099	0.063	−0.546	vomiting
		−0.443	−0.176	−0.559	−0.127	abdominal pain
		0.459	−0.115	−0.606	−0.013	anorexia
		0.338	0.380	0.453	0.741[a]	ageusia
		−0.051	−0.285	−0.122	0.312	diarrhea

Supplementary Table 1. Continued

Incidence	State	Lag 1	Lag 2	Lag 3	Lag 4	Search term
Low	Hawaii	−0.324	−0.594	−0.537	−0.062	loss of appetite
		0.009	−0.252	0.155	−0.537	vomiting
		0.231	−0.061	−0.144	−0.160	abdominal pain
		−0.106	−0.726[a]	−0.237	0.245	anorexia
		0.187	−0.146	0.265	0.490	ageusia
		−0.213	0.287	−0.040	0.307	diarrhea
	Montana	0.370	0.240	−0.250	−0.193	loss of appetite
		−0.012	−0.589	−0.480	0.376	vomiting
		−0.191	−0.153	−0.201	−0.407	abdominal pain
		0.295	0.146	0.136	0.212	anorexia
		0.583[a]	0.570	0.543	0.588	ageusia
		0.054	−0.421	−0.513	−0.309	diarrhea
	North Dakota	−0.332	−0.287	−0.179	0.320	loss of appetite
		−0.262	−0.480	−0.679[a]	−0.184	vomiting
		−0.352	−0.431	−0.584	−0.346	abdominal pain
		−0.104	0.278	0.170	−0.301	anorexia
		−0.176	−0.119	0.045	0.773[a]	ageusia
		−0.196	−0.223	−0.201	0.124	diarrhea
	Wyoming	—	—	—	—	loss of appetite
		0.000	0.335	−0.627	−0.036	vomiting
		−0.519	−0.482	0.016	0.227	abdominal pain
		−0.283	−0.135	0.356	0.313	anorexia
		—	—	—	—	ageusia
		0.109	0.254	0.311	0.574	diarrhea
	Alaska	—	—	—	—	loss of appetite
		−0.650[a]	−0.556	−0.574	−0.407	vomiting
		—	—	—	—	abdominal pain
		−0.364	−0.218	0.209	0.211	anorexia
		0.242	0.645[a]	0.548	−0.131	ageusia
		−0.051	0.184	−0.123	0.466	diarrhea

COVID-19, coronavirus disease 2019.

[a]$P < .05$.