

# COVID-19 Hospitalizations Forecasts Using Internet Search Data

**Tao Wang**

Georgia Institute of Technology

**Simin Ma**

Georgia Institute of Technology

**Soobin Baek**

Georgia Institute of Technology

**Shihao Yang** (✉ [shihao.yang@isye.gatech.edu](mailto:shihao.yang@isye.gatech.edu))

Georgia Institute of Technology

---

## Research Article

### Keywords:

**Posted Date:** February 23rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1336225/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# COVID-19 Hospitalizations Forecasts Using Internet Search Data

Tao Wang<sup>1</sup>, Simin Ma<sup>1</sup>, Soobin Baek<sup>1</sup>, and Shihao Yang<sup>1,\*</sup>

<sup>1</sup>H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30309, USA

\*Corresponding Author. shihao.yang@isye.gatech.edu

## ABSTRACT

As the COVID-19 spread over the globe and new variants of COVID-19 keep occurring, reliable real-time forecasts of COVID-19 hospitalizations are critical for public health decision on medical resources allocations such as ICU beds, ventilators, and personnel to prepare for the surge of COVID-19 pandemics. Inspired by the strong association between public search behavior and hospitalization admission, we extended previously-proposed influenza tracking model, ARGO (AutoRegression with Google search data), to predict future 2-week national and state-level COVID-19 new hospital admissions. Leveraging the COVID-19 related time series information and Google search data, our method is able to robustly capture new COVID-19 variants' surges, and self-correct at both national and state level. Based on our retrospective out-of-sample evaluation over 12-month comparison period, our method achieves on average 15% error reduction over the best alternative models collected from COVID-19 forecast hub. Overall, we showed that our method is flexible, self-correcting, robust, accurate, and interpretable, making it a potentially powerful tool to assist health-care officials and decision making for the current and future infectious disease outbreak.

## Introduction

COVID-19 (SARS-CoV-19), an acute respiratory syndrome disease caused by a coronavirus, has spread worldwide causing over 120,695,785 reported cases and 4,987,755 reported deaths<sup>1</sup>. During the continuous spread of COVID-19, many variants (alpha, delta, omicron, etc.) of COVID-19 emerges, leading to drastic surges in hospital admissions and shortages in health care resources<sup>2</sup>. Therefore, an accurate hospital admissions forecasting model is crucial to assist hospitals and policy makers with the possibilities and the timings of rapid changes, so as to further respond to and prepare for future COVID-19 spread.

The Centers for Disease Control and Prevention (CDC)<sup>3</sup> has been collecting hospitalizations predictions from various research teams and making ensemble and baseline predictions since May 2020. According to the weekly COVID-19 forecasts submissions compiled by CDC<sup>3</sup>, machine learning<sup>4,5</sup> and compartmental models<sup>6,7</sup> are the most popular forecasting approaches<sup>4,5</sup>. For example, Rodríguez, et al.<sup>4</sup> use a neural network architecture incorporating COVID-19 time series, and mobility information as inputs, whereas Jin, et al.<sup>5</sup> utilize attention and transformer models to compare and combine past COVID-19 trends for future forecasts. On the other hand, Vespignani, et al.<sup>6</sup> and Kinsey, et al.<sup>7</sup> adapt SEIR (compartmental model) as the baseline structure, and combine different exogenous variables including spatial-temporal and mobility information to build more sophisticated models to capture COVID-19 disease dynamics and forecast hospitalizations. Meanwhile, statistical and data-driven models, taking advantage of COVID-19 public search information for hospitalizations predictions, have not drawn much attention.

For the last decade, numerous studies have suggested online search data could be a valuable component to monitor and forecast infectious diseases, such as influenza<sup>8-13</sup>, HIV/AIDS<sup>14</sup>, dengue<sup>15</sup>, etc. For instance, Google Flu Trend (GFT)<sup>8</sup>, a digital disease detection system that uses the volume of selected Google search terms to estimate current influenza-like illnesses (ILI) activity, was the first among many studies to demonstrate how big data and public search behavior compliment traditional statistical predictive analysis. Later, many studies proposed different methodologies to improve upon GFT and provide more robust and accurate real-time forecast estimates of influenza activity in the U.S., including machine learning models<sup>10</sup>, statistical models<sup>9</sup>, ensemble models<sup>11</sup>. Online search information from other sources, such as Yahoo<sup>12</sup> and Baidu<sup>13</sup>, have demonstrated their predictive power for influenza tracking as well.

On the other hand, to best of our knowledge, the only hospitalizations prediction model that is based on the Internet search data is a vector error correction model (VECM) proposed by Philip J. Turk, et al.<sup>16</sup>. VECM combines Google search data and healthcare chatbot scores and produces COVID-19 hospitalizations predictions in Greater Charlotte market area, where the Google search queries are selected based on existing literature, and filtered through hand-craft procedures. Yet, their selection approach highly depends on literature considered and healthcare chatbot scores can be limited in-access across different

geographical areas. So far, none of the existing Internet-search-based methods provide robust weeks-ahead hospitalisation forecasts for different geographical areas in the United States that fully utilize COVID-19 related online search queries and account for their sparsity and noisiness.

In this paper, we propose the ARGO inspired model (ARGO), which combines online search data and lagged COVID-19 related information in a L1-norm penalized linear regression model to produce real-time 1-2 weeks ahead national and state-level hospitalizations predictions over 12-month out-of-sample evaluation period in the United States. Our results show that ARGO, leveraging predictive information in the selected interpretable Google search queries, is able to largely outperform baseline models and all other public available models from CDC’s COVID-19 forecast hub, which implies potential in ARGO to help with urgent public health decisions on health resources allocations.

## Results

We conducted retrospective evaluation of weekly hospital admissions for the period between January 4, 2021 and December 27, 2021, on both national and state level. To evaluate prediction performance, we calculated the root mean square error (RMSE), the mean absolute error (MAE) and the Pearson correlation coefficient (Cor) of one-week-ahead and two-week-ahead predictions (detailed in the Methods section). All comparisons are based on original scale of the ground truth of new hospital admissions released by U.S. Department of Health and Human Services (HHS)<sup>17</sup>.

### Comparisons of National COVID-19 Hospitalizations Predictions

National one-week-ahead and two-week-ahead predictions of new hospitalizations were generated using (i) ARGO inspired model, (ii) persistence (naive) model and (iii) AR7 model. The naive method simply uses past week’s hospitalizations as current week’s forecasts, without any modeling effort. AR7 is an autoregressive model of lag 7 (implemented in R package `forecast`<sup>18</sup>). For fair comparisons, all models were trained on a 56-day rolling window basis. Retrospective out-of-sample predictions of daily national hospitalizations were made every week from January 4, 2021 to December 27, 2021 by the three models and were then aggregated into weekly new hospitalizations. To further demonstrate the prediction accuracy and robustness of ARGO, we collected hospitalizations predictions of the two benchmark methods (COVIDhub-baseline and COVIDhub-ensemble) from COVID-19 forecast hub<sup>3</sup>. The COVIDhub-baseline is a persistent method that uses latest daily observation as future daily predictions<sup>3</sup>. The COVIDhub-ensemble uses medians of hospitalizations predictions submitted to COVID-19 forecast hub as its “ensemble” forecasts. We also provide a full comparison of the top 5 teams from CDC’s COVID-19 forecast hub in Supplementary Materials (Table S4).

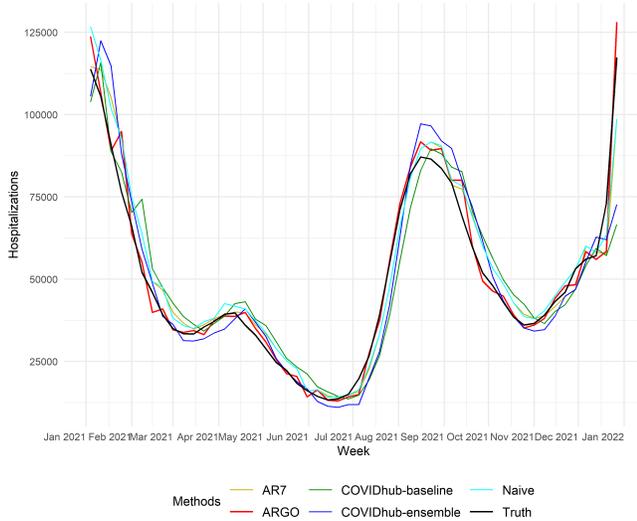
**Table 1.** National Level Comparison Error Metrics

Methods	RMSE		MAE		Cor	
	1 Week Ahead	2 Weeks Ahead	1 Week Ahead	2 Weeks Ahead	1 Week Ahead	2 Weeks Ahead
ARGO	<b>4667.552</b>	<b>14010.648</b>	<b>2896.857</b>	<b>9435.759</b>	<b>0.988</b>	<b>0.906</b>
AR7	6408.431	15750.873	4726.405	10957.369	0.973	0.834
COVIDhub-ensemble <sup>3</sup>	9578.408	16327.146	6301.981	10856.115	0.942	0.851
COVIDhub-baseline <sup>3</sup>	10442.092	19210.623	7167.538	13183.231	0.916	0.738
Naive	10528.225	19831.180	7033.212	13633.750	0.918	0.732

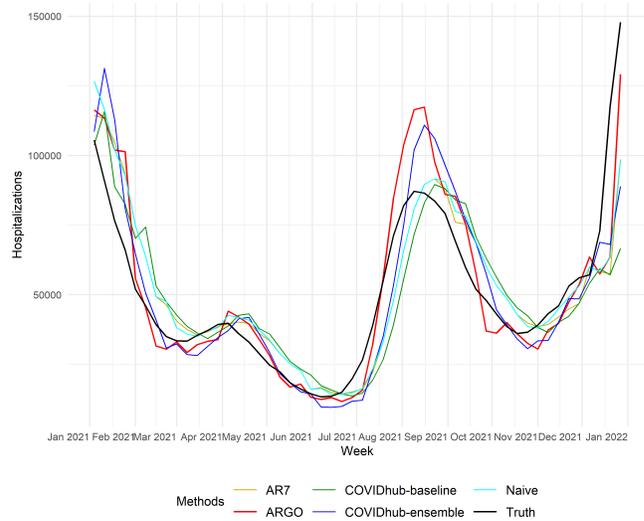
Error metrics of national one-week-ahead and two-week-ahead new hospitalizations predictions. The best scores are highlighted with boldface. All comparisons are based on the original scale of hospitalizations released by HHS. Methods are sorted by their average RMSE of one-week-ahead and two-week-ahead predictions. On average, the ARGO model outperforms the best alternative methods by approximately 18% in RMSE, 25% in MAE and 4% in Cor. Overall, ARGO has better predictions than all the benchmark methods during our comparison period.

Table 1 summarizes the national one-week-ahead and two-week-ahead predictions performance from January 4, 2021 to December 27, 2021. During this period, ARGO outperforms all the benchmark models in every error metric for both one and two weeks ahead predictions. Specifically, for the national one-week-ahead predictions, ARGO performs better than the best alternative method by around 27% in RMSE, 39% in MAE and 1.5% in Cor. The two-week-ahead ARGO forecasts have slightly lower error reduction in RMSE and MAE, and higher increase in Cor. SI Table S4 has additional national level comparison results. The results demonstrate that ARGO is able to produce accurate and robust retrospective out-of-sample national one-week-ahead and two-week-ahead hospitalizations predictions during the evaluation period.

Figure 1 displays the 5 compared methods’ estimations. During the 12-month comparison period, ARGO can accurately capture the overall trend as well as local fluctuations (the spike between April, 2021 and May, 2021) in new hospitalizations. All of the time series forecasting methods exhibit some delaying behaviors to various degree, due to the input feature of the lagged information. Fortunately, by utilizing Google search information, ARGO is able to overcome such delayed behavior and is the only method that captures the hospital admission peaks around April 2021 and September 2021 as well as the surge around December 2021 possibly caused by omicron for both 1 and 2 weeks ahead predictions. Moreover, by leveraging time



(a) One-Week-Ahead National Level Predictions



(b) Two-week-ahead National Level Predictions

**Figure 1.** One-week-ahead and two-week-ahead hospitalizations predictions of the 5 compared models. 1 and 2 weeks ahead predictions are compared weekly from 2021-01-04 to 2021-12-27. True new hospital admissions released by HHS are marked in black. Predictions of the 5 models, which consist of ARGO, AR7, naive, COVIDhub-baseline, COVIDhub-ensemble, are marked in red, gold, cyan, green, and blue, respectively.

series information and the Internet search information, ARGO is able to avoid sudden spikes and drops in the prediction. ARGO is also self-correcting and can quickly recover from the over-shooting behavior (e.g. between August to September 2021 for two weeks ahead prediction).

### Comparisons of State-level COVID-19 Hospitalizations

We also conducted retrospective out-of-sample one and two weeks ahead predictions for the 51 states in the U.S. (including Washington D.C.) during the same period of January 4, 2021 to December 27, 2021.

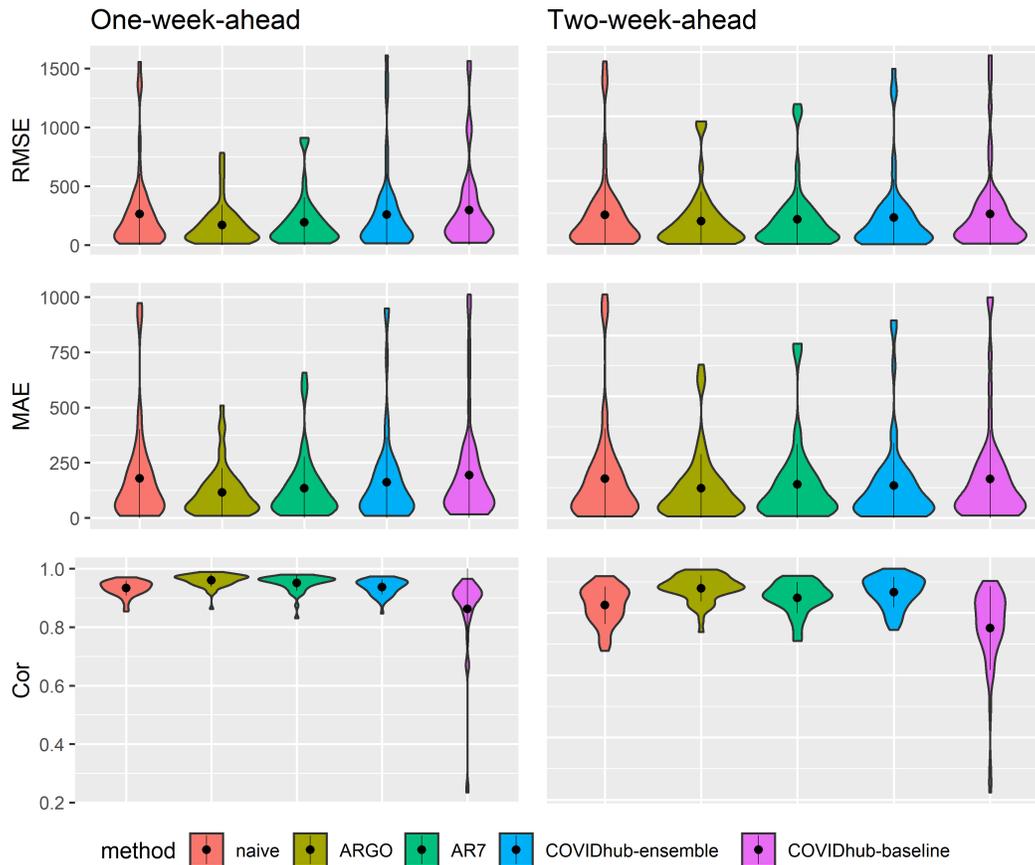
**Table 2.** State-Level Comparison Error Metrics

Methods	RMSE		MAE		Cor	
	1 Week Ahead	2 Weeks Ahead	1 Week Ahead	2 Weeks Ahead	1 Week Ahead	2 Weeks Ahead
ARGO	<b>170.960</b>	<b>374.330</b>	<b>114.98</b>	<b>243.765</b>	<b>0.960</b>	<b>0.879</b>
AR7	193.424	399.253	134.454	276.255	0.951	0.849
COVIDhub-ensemble <sup>3</sup>	259.348	427.926	161.523	265.042	0.937	0.867
Naive	265.789	469.206	179.084	322.021	0.934	0.825
COVIDhub-baseline <sup>3</sup>	296.963	482.160	193.150	320.048	0.862	0.751

Error metrics of state-level one-week-ahead and two-week-ahead new hospitalizations predictions, averaging across the states. The best scores are highlighted with boldface. All comparisons are based on the original scale of hospitalizations released by HHS. Methods are sorted based on their average RMSE of one-week-ahead and two-week-ahead predictions. ARGO outperforms the best alternative method by approximately 8% in RMSE, 12% in MAE and 1% in Cor. Overall, ARGO is the best-performing prediction model compared with other listed models.

Table 2 summarizes the average error metrics of all methods' state-level predictions from January 4, 2021 to December 27, 2021. For the one-week-ahead predictions, ARGO is able to achieve uniformly best performance in all error metrics. Compared with the two benchmark models from COVID-19 forecast hub, ARGO yields roughly 35% error reduction in RMSE, around 30% error reduction in MAE and around 2% increase in Pearson correlation coefficient. For the two-week-ahead predictions, ARGO achieves around 12% error reduction in RMSE, approximate 8% error reduction in MAE and around 1% increase in Pearson correlation coefficient compared with best alternative models from COVID-19 forecast hub (table S5). Overall, ARGO gives the leading performance in state-level forecasts compared with the benchmark models, by efficiently utilizing relevant public search information and incorporating cross-state cross-region information as model features. The full comparison of state-level hospitalizations predictions is shown in SI Table S5, and the comparison tables and figures for each individual state are available in SI Table S6-S56 and Figure S4-S54. Figure 2 further demonstrates the accuracy and robustness of ARGO in

state-level hospitalizations predictions. The violin charts, which present the distributions of each model's predictions errors in all three error metrics, show that the 1 standard deviation ranges of ARGO are the smallest in RMSE and MAE, and are the second best in Cor.



**Figure 2.** The distributions of error metrics of one-week-ahead and two-week-ahead predictions from 2021-01-04 to 2021-12-27 over the 51 states of the 5 compared models. The embedded black dot indicates mean. The vertical line represents 1 standard deviation range which is trimmed to be inside normal range of RMSE, MAE and correlations coefficients.

## Discussion

The ARGO inspired model combines autoregressive COVID-19 related information and online search data. It is able to produce accurate, reliable real-time hospitalizations predictions at both national and state-level for 1-2 weeks ahead predictions. The state-level real-time hospitalizations predictions made by ARGO could help local public health officials make timely allocation decisions of healthcare resources, such as ventilator, ICU beds, personal protective equipment, personnel, etc, as well as account and promptly prepare for future surges of COVID-19 pandemics caused by new virus variants. Furthermore, our ARGO hospitalization prediction model is a straightforward adaptation from the original ARGO model for influenza<sup>9</sup>, which reduces the chance of overfitting and again demonstrates ARGO's robustness and flexibility.

Although ARGO shows strong performance in hospitalizations forecasts, its accuracy is controlled by the reliability of the inputs. Google search data can be noisy due to the instability of Google Trends' sampling approach and public fear. Especially for state-level Google search data, the lack of search intensity can make the search data unrepresentative of the real interest of the people. Luckily, the IQR filter<sup>19</sup> and moving average smoothing applied to Google search data are able to minimize the risk caused by such noisiness, and help ARGO produce robust output. To further account for the instability in the state-level Google search queries, the query terms are identified using the national level data where the search frequencies are more representative with lower noise and higher stability. ARGO selects the most representative search queries according to their Pearson correlation coefficients with hospital admission. In addition, the national level search frequency is directly used as input features for state-level predictions. An optimal delay between the search data and the hospital admissions is also identified for each query term. All together, ARGO is able to overcome the sparsity issues of Google search queries and produce robust

future estimations (Figure 1) while avoiding over-fitting.

Another challenge in using online search data to estimate hospitalizations is that the predictive information in Google search data die down as forecasts horizon expands (shown in Table 1 and 2). In our results of COVID-19 hospitalizations predictions, the optimal lags (delays) of some Google search terms are relative small shown in table 3 which indicates those Google search queries are more effective for short-term prediction of hospitalizations. Nevertheless, by leveraging COVID-19 related time series information, ARGO is able to adjust the focus between the time series and the Internet search information features when forecast horizon extends, thanks to the L1-norm penalty and the dynamic training that selects the most relevant Google search terms.

By effectively combining the Google search data and COVID-19 related time series information, ARGO has a stable model structure that is able to make accurate and robust national and state level 1-2 weeks ahead hospitalizations predictions. The simple structure of ARGO makes it universal adaptable to other COVID-19 related forecasts. With its simplicity and strengthened accuracy over other benchmark methods, ARGO could help public health decision making for the local monitoring and control of COVID-19 to better prepare for future surges of hospitalizations, patients in ICU, deaths and cases caused by new COVID-19 variants such as alpha, delta, omicron, etc.

## Data and Methods

We focused on national hospital admission predictions and state-level predictions of 51 states in the United States, including Washington D.C.. The inputs consist of confirmed incremental cases, percentage of vaccinated population, confirmed new hospital admissions, and Google search query frequencies. Both state-level data and national data were directly obtained from respective data sources outlined in this section. Our prediction method is inspired by ARGO<sup>9</sup>, with details presented in this section as well.

### Data Availability

All data used in this study are publicly available from respective data sources outlined here. For completeness, the datasets analysed during the current study are also deposited in the Harvard dataverse repository, DOI: 10.7910/DVN/XXXXXXXX.

### COVID-19 Related Data

We used reported COVID-19 confirmed incremental cases from JHU CSSE data<sup>20</sup>, percentage of vaccinated population from Centers for Disease Control and Prevention (CDC)<sup>21</sup> and COVID-19 confirmed new hospital admissions from HHS<sup>17</sup>. The data sets were collected from July 15, 2020 to January 15, 2021.

### Google Search Data

Google Trends provides estimated Google search frequency for the specified query term<sup>22</sup>. We obtained online search data from Google Trends<sup>22</sup> for the period from July 15, 2020 to January 15, 2021. To retrieve the time series search frequencies of a desired query, one needs to specify the query's geographical information and time frame on Google Trends. The returned frequency from Google Trends is obtained by sampling all raw Google search frequencies that contain this query<sup>22</sup>. We collected daily Google Search frequencies of 256 top searched COVID-19-related terms based on the previous work of COVID-19 death forecasts<sup>19</sup>.

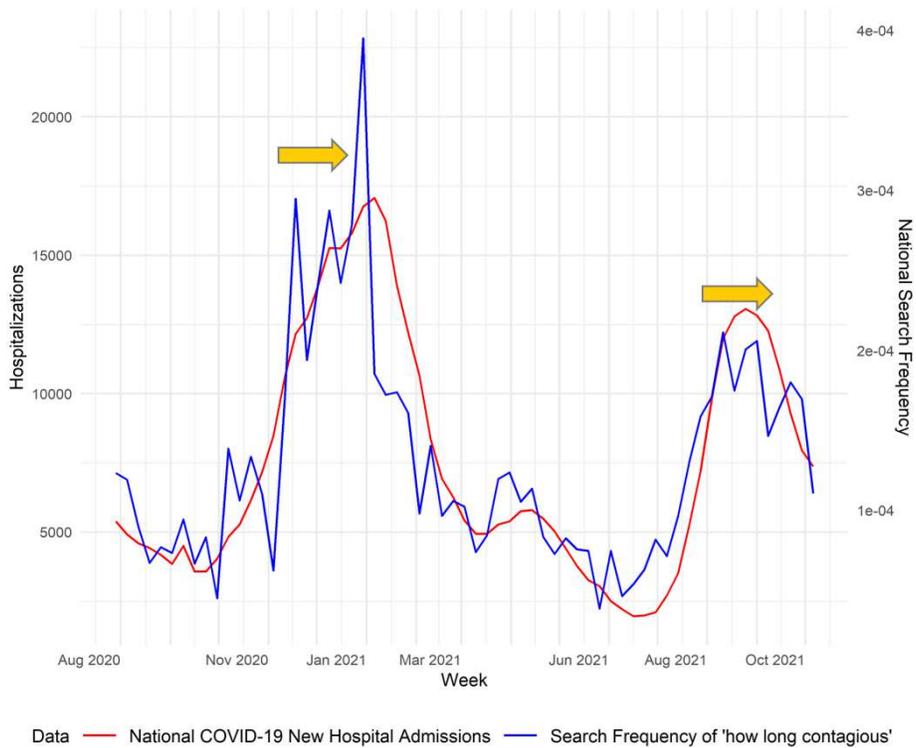
### *Inter-Quantile Range (IQR) Filter and Optimal Lags for Google Search Data*

The raw Google search frequencies obtained from Google Trends<sup>22</sup> are observed to be unstable and sparse<sup>19</sup>. Such instability and sparsity can negatively affect prediction performance of linear regression models which are sensitive to outliers. To deal with such outliers in Google search data, we used an IQR filter<sup>19</sup> to remove and correct outliers on a rolling window basis. The search data that is beyond 3 standard deviation from the past 7-day average are examined and removed<sup>19</sup>.

The trends of Google search frequencies are often a few days ahead of hospitalizations, indicating that the search data might contain predictive information about hospitalizations. Figure 3 demonstrates the delay behavior between Google search query frequencies and national hospitalizations. To fully utilize the predictive information in national Google search terms, we found and applied optimal lags<sup>19</sup> to filtered Google search frequencies to match the trends of national hospitalizations. For each query, a linear regression of COVID-19 new hospitalizations is fitted against lagged Google search frequency, considering a range of lags. The lag results in lowest mean square error is selected as the optimal lag for that query. The data used to find optimal lags are from August 1, 2020 to December 31, 2020.

### *Selection of Google Search Terms*

Among the 256 COVID-19 related terms, we further selected the queries that have correlation coefficients with national COVID-19 hospitalizations larger than 0.6 for the period from August 1, 2020 to December 31, 2020. We applied 7-day



**Figure 3. Google search query “how long contagious” and COVID-19 weekly new hospitalizations** Illustration of delay in peak between Google search query search frequencies (how long contagious in blue) and COVID-19 national level weekly new hospitalizations (red). Y-axis are adjusted accordingly.

moving average to further smooth out weekly fluctuations in the selected Google search queries. Table 3 shows the 11 selected important terms as well as their optimal lags. Table 3 supports our hypothesis that when people get infected, they would probably search for general query like "symptoms of the covid-19" first as this query has relative large optimal lag. After the symptoms develop, people might begin to search for specific symptoms such as "loss of smell" which has relatively smaller optimal lag.

**Table 3.** Optimal Lags of Selected Important Terms

Selected Google Search Term	Optimal Lag (in days)
how long contagious	4
loss of smell	4
loss of taste	9
covid-19 vaccine.	7
Cough	7
pneumonia	7
how long covid-19	4
sinus	21
symptoms of the covid-19	7
contagious coronavirus	4
coronavirus vaccine	4

### ARGO Inspired Prediction

Let  $\hat{y}_{t,r}$  be the daily hospital admissions of region  $r$  on day  $t$ ;  $X_{k,t}$  be the Google search data of term  $k$  on day  $t$ ;  $c_{t,r}$  be the JHU COVID-19 incremental confirmed cases on day  $t$  of region  $r$ ;  $v_{t,r}$  be the cumulative percent of people who get vaccinated by day  $t$  of region  $r$ ;  $\mathbb{I}_{\{t,d\}}$  be the weekday indicator for day  $t$  (i.e.  $\mathbb{I}_{\{t,1\}}$  indicates day  $t$  being Monday). Standing on day  $T$ , to predict  $l$ -day-ahead hospital admission of state  $r$ ,  $\hat{y}_{T+l,r}$ , we used penalized linear estimator as following:

$$\hat{y}_{T+l,r} = \hat{\mu}_{y,r} + \sum_{i=0}^I \hat{\alpha}_{i,r} y_{T-i,r} + \sum_{j \in J} \hat{\beta}_{j,r} c_{T+l-j,r} + \sum_{m \in M_r} \hat{\gamma}_{m,r} y_{T,m} + \sum_{q \in Q} \hat{\phi}_{q,r} v_{T+l-q,r} + \sum_{k=1}^K \hat{\delta}_{k,r} X_{k,T+l-\hat{O}_k} + \sum_{d=1}^6 \hat{\tau}_{d,r} \mathbb{I}_{\{T+l,d\}} \quad (1)$$

Where  $I = 6$  considering consecutive one week lagged daily hospital admissions;  $J = \max(\{7, 28\}, l)$ , considering lagged confirmed cases;  $M_r$  is the set of geographical neighboring states of state  $r$ ;  $Q = \max(7, l)$ , considering vaccination data lagged by one week;  $\hat{O}_k = \max(O_k, l)$  is the adjusted optimal lag for term  $k$ ;  $K = 11$ , considering 11 selected Google search terms. The coefficients for  $l$ -day-ahead predictions of region  $r$ ,  $\{\mu_{y,r}, \alpha = (\alpha_{1,r}, \dots, \alpha_{6,r}), \beta = (\beta_{1,r}, \dots, \beta_{|J|,r}), \gamma = (\gamma_{1,r}, \dots, \gamma_{|M_r|,r}), \phi = \phi_{\max(7,l),r}, \delta = (\delta_{1,r}, \dots, \delta_{11,r}), \tau = (\tau_{1,r}, \dots, \tau_{6,r})\}$ , were computed by

$$\begin{aligned} \operatorname{argmin}_{\mu_{y,r}, \alpha, \beta, \gamma, \phi, \delta, \tau, \lambda} \sum_{t=T-M-l+1}^{T-l} \omega^{T-l-t+1} & \left( y_{t+l,r} - \mu_{y,r} - \sum_{i=0}^6 \alpha_{i,r} y_{t-i,r} - \sum_{j \in J} \hat{\beta}_{j,r} c_{t+l-j,r} - \sum_{m \in M_r} \hat{\gamma}_{m,r} y_{t,m} \right. \\ & \left. - \sum_{q \in Q} \hat{\phi}_{q,r} v_{t+l-q,r} - \sum_{k=1}^5 \hat{\delta}_{k,r} X_{k,t+l-\hat{O}_k} - \sum_{d=1}^6 \hat{\tau}_{d,r} \mathbb{I}_{\{t+l,d\}} \right)^2 \\ & + \lambda_\alpha \|\alpha\|_1 + \lambda_\beta \|\beta\|_1 + \lambda_\gamma \|\gamma\|_1 + \lambda_\phi \|\phi\|_1 + \lambda_\delta \|\delta\|_1 + \lambda_\tau \|\tau\|_1 \end{aligned} \quad (2)$$

$M = 56$  which is the length of our training period;  $\omega = 0.8$  is the exponentially time-decaying weight which assigns higher weight on more recent observation. Region  $r$  consists of U.S. and its 51 states, including Washington D.C.. For U.S. national level training, the hospitalizations of neighboring states,  $y_{t,m}$ , and their coefficients,  $\gamma$ , are excluded. To address the sparsity of Google search data, we used penalty of L1-norm. For simplicity, the hyperparameters  $\lambda = (\lambda_\alpha, \lambda_\beta, \lambda_\gamma, \lambda_\phi, \lambda_\delta, \lambda_\tau)$  for L1-norm penalty were set to be equal and obtained via 10-folds cross-validation.

With the formulation above, on each Monday from January 4, 2021 to December 27, 2021, we iteratively trained our model and made national and state-level retrospective out-of-sample hospitalizations predictions up to 14 days into future. We then aggregated daily predictions into one-week-ahead and two-week-ahead predictions. For example,  $\hat{y}_{T+1:T+7,r} = \sum_{i=1}^7 \hat{y}_{T+i,r}$  and  $\hat{y}_{T+8:T+14,r} = \sum_{i=8}^{14} \hat{y}_{T+i,r}$  are the one-week-ahead prediction and two-week-ahead prediction on day  $T$  of region  $r$ , respectively.

### Evaluation Metrics

Root Mean Squared Error (RMSE) between a hospitalization estimate  $\hat{y}_t$  and the true value  $y_t$  over period  $t = 1, \dots, T$  is  $\sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}$ . Mean Absolute Error (MAE) between an estimate  $\hat{y}_t$  and the true value  $y_t$  over period  $t = 1, \dots, T$  is

$\frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t|$ . Correlation is the Pearson correlation coefficient between  $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_T)$  and  $\mathbf{y} = (y_1, \dots, y_T)$ . All estimates  $\hat{y}_t$  and the true value  $y_t$  were weekly aggregated before calculating RMSE, MAE and Cor.

## References

1. Ritchie, H. *et al.* Coronavirus pandemic (covid-19). *Our World Data* (2020). <https://ourworldindata.org/coronavirus>.
2. Lurye, S. States with the biggest hospital staffing shortages. *US NEWS* (2022).
3. Cramer, E. Y. *et al.* The united states covid-19 forecast hub dataset. *medRxiv* DOI: [10.1101/2021.11.04.21265886](https://doi.org/10.1101/2021.11.04.21265886) (2021).
4. Rodríguez, A. *et al.* Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting, DOI: [10.1101/2020.09.28.20203109](https://doi.org/10.1101/2020.09.28.20203109) (2020).
5. Jin, X., Wang, Y.-X. & Yan, X. Inter-series attention model for covid-19 forecasting. *arXiv preprint arXiv:2010.13006* (2020). <https://arxiv.org/abs/2010.13006>.
6. Chinazzi, M. *et al.* Global epidemic and mobility model (2020). Accessed = 2021-11-03.
7. Kinsey, M. *et al.* Jhuapl-bucky (2021). Accessed = 2021-11-03.
8. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–4, DOI: [10.1038/nature07634](https://doi.org/10.1038/nature07634) (2008).
9. Yang, S., Santillana, M. & Kou, S. C. Accurate estimation of influenza epidemics using google search data via argo. *Proc. Natl. Acad. Sci.* **112**, 14473–14478, DOI: [10.1073/pnas.1515373112](https://doi.org/10.1073/pnas.1515373112) (2015). <https://www.pnas.org/content/112/47/14473.full.pdf>.
10. Santillana, M., Nguyen, A., Dredze, M., Paul, M. & Brownstein, J. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology* **11**, DOI: [10.1371/journal.pcbi.1004513](https://doi.org/10.1371/journal.pcbi.1004513) (2015).
11. Lu, F., Hattab, M., Clemente, C., Biggerstaff, M. & Santillana, M. Improved state-level influenza nowcasting in the united states leveraging internet-based data and network approaches. *Nat. Commun.* **10**, DOI: [10.1038/s41467-018-08082-0](https://doi.org/10.1038/s41467-018-08082-0) (2019).
12. Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D. & Weinstein, R. A. Using internet searches for influenza surveillance. *Clin. infectious diseases* **47**, 1443–1448 (2008).
13. Yuan, Q. *et al.* Monitoring influenza epidemics in china with search query from baidu. *PLoS one* **8**, e64323 (2013).
14. Young, S. D. & Zhang, Q. Using search engine big data for predicting new hiv diagnoses. *PLoS one* **13**, e0199527 (2018).
15. Yang, S. *et al.* Advances in using internet searches to track dengue. *PLoS computational biology* **13**, e1005607 (2017).
16. Turk, P. J., Tran, T. P., Rose, G. A. & McWilliams, A. A predictive internet-based model for covid-19 hospitalization census. *Sci. Reports* **11**, DOI: [10.1038/s41598-021-84091-2](https://doi.org/10.1038/s41598-021-84091-2) (2021).
17. U.S. Department of Health and Human Services. Healthdata.gov covid-19 reported patient impact and hospital capacity by state timeseries. <https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh> (2021). Last accessed on 2021-11-03.
18. Hyndman, R. *et al.* *forecast: Forecasting functions for time series and linear models* (2021). R package version 8.15.
19. Ma, S. & Yang, S. Covid-19 forecasts using internet search information in the united states. *arXiv preprint arXiv:2106.12160* (2021). <https://arxiv.org/abs/2106.12160>.
20. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect Dis* **20** (2020).
21. Centers for Disease Control and Prevention. Covid-19 vaccinations in the united states, county. <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh> (2021). Last accessed on 2021-11-03.
22. Faq about google trends data. [https://support.google.com/trends/answer/4365533?hl=en&ref\\_topic=6248052](https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052). Accessed: 2021-11-03.

## Author contributions statement

T.W., S.M., S.B. and S.Y. designed the research; T.W., S.M., S.B. and S.Y. performed the research; T.W. and S.B. analyzed data and conducted the experiment(s); T.W. and S.M. analysed the results. T.W., S.M. and S.Y. wrote the paper. All authors reviewed the manuscript.

## **Competing interests**

The authors declare no competing interests.

## **Additional information**

Correspondence and requests for materials should be addressed to S.Y.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ARGOCOVID19HospitalizationForecasts0221SI.pdf](#)