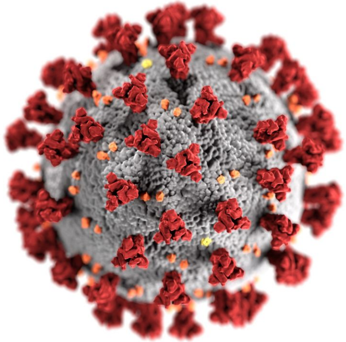
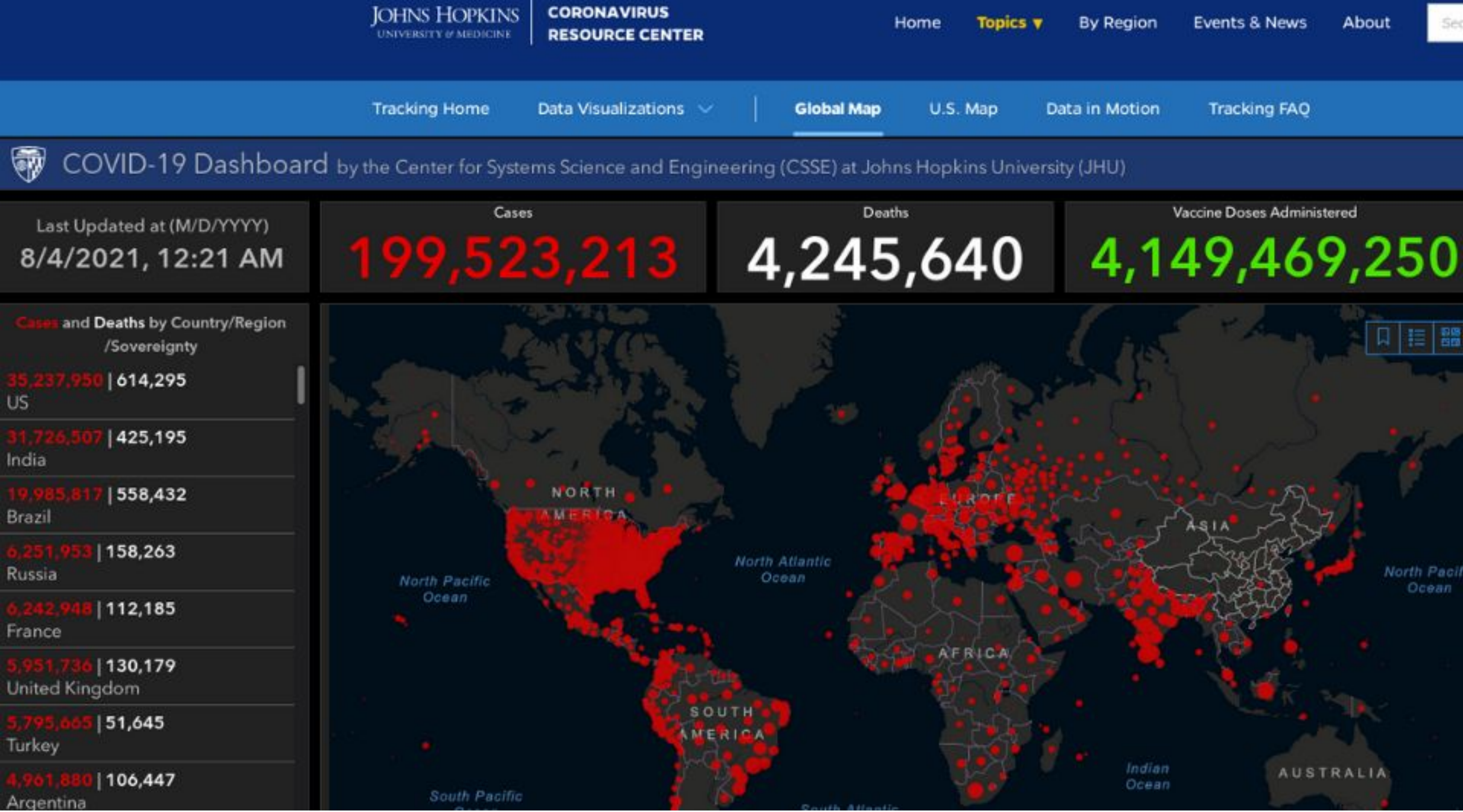


Predicting Covid-19 Deaths

By Miriam Sosa

Wed, August 4

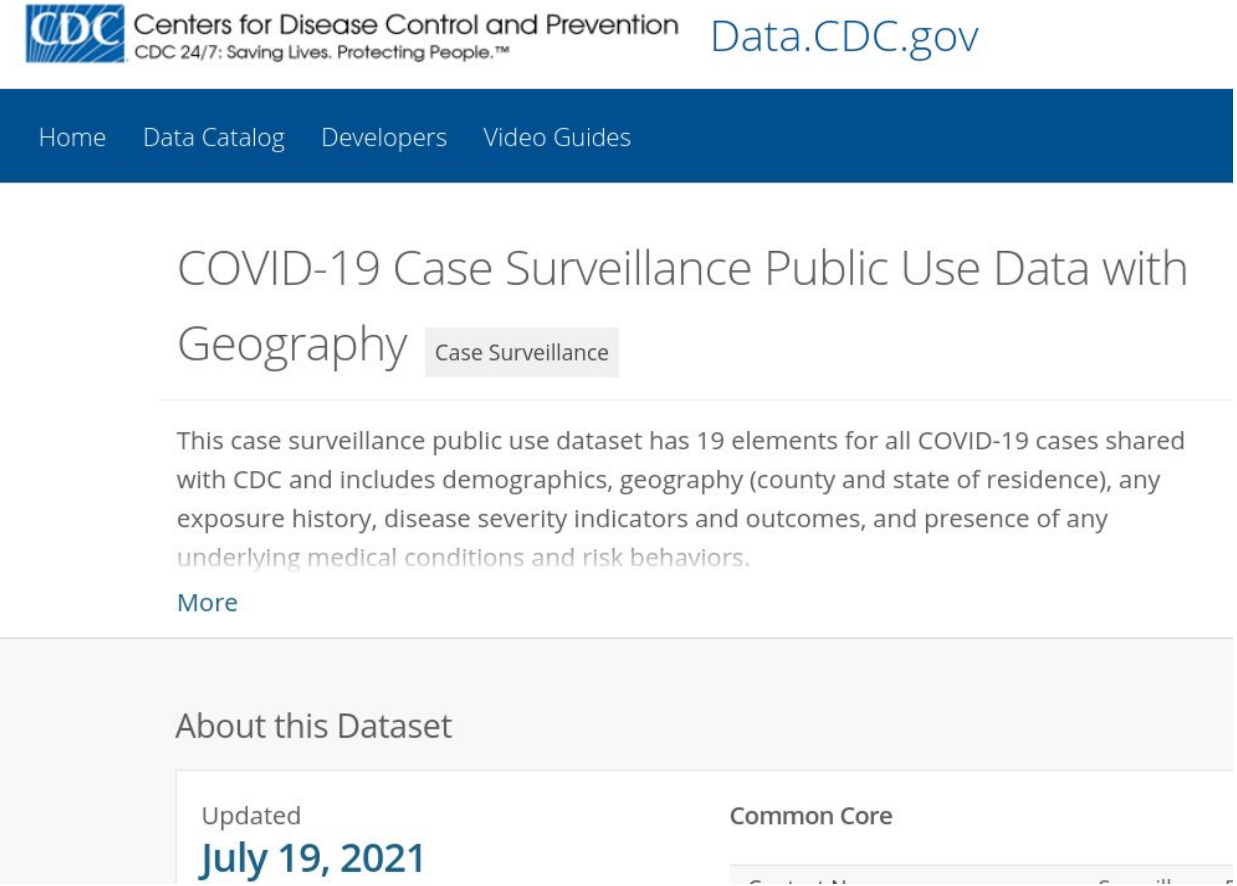


COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Last Updated at (M/D/YYYY)	Cases	Deaths	Vaccine Doses Administered
8/4/2021, 12:21 AM	199,523,213	4,245,640	4,149,469,250

Cases and Deaths by Country/Region/Sovereignty

Cases	Deaths	Country/Region
35,237,950	614,295	US
31,726,507	425,195	India
19,985,817	558,432	Brazil
6,251,953	158,263	Russia
6,242,948	112,185	France
5,951,736	130,179	United Kingdom
5,795,665	51,645	Turkey
4,961,880	106,447	Argentina



Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™

Data.CDC.gov

COVID-19 Case Surveillance Public Use Data with Geography

Case Surveillance

This case surveillance public use dataset has 19 elements for all COVID-19 cases shared with CDC and includes demographics, geography (county and state of residence), any exposure history, disease severity indicators and outcomes, and presence of any underlying medical conditions and risk behaviors.

About this Dataset

Updated
July 19, 2021

Common Core

Data Source

N = 26,887,803 cases

279,097 deaths

Case fatality **~1%**

Subset

n = 100,000 cases randomly sampled

966 deaths

Problem Statement:

To predict mortality outcomes of COVID cases in the US, dated Jan 2020-June 2021

Classification Approach

twitter@Chelsea Parlett-Pelleriti

Random Forest

Undersampling majority cases (non-death)

		Predicted	
		No Death Recorded	Death
Actual	No Death Recorded	**TN** 22,189	**FP** 2,570
	Death	**FN** 16	**TP** 225

Predictions: HIGH RECALL, low precision

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{225 + 22189}{25000} = \mathbf{90\%}$$

Recall

TP / (TP + FN) = Sensitivity

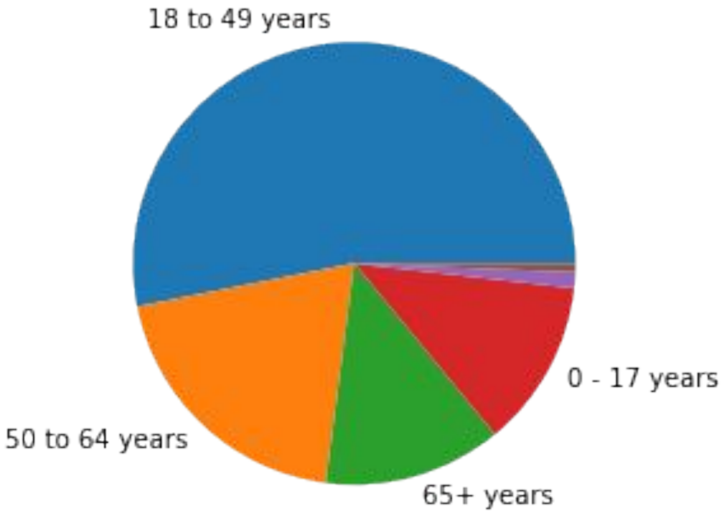
225 / (225+16) = **93% of deaths predicted**

Precision

TP / (TP + FP) = Specificity

225 / (225+2,570) = **8%**

Predictors - Feature Importance



- Age Groups (65+, 18-49, 50-64)
- Hospitalization Yes/No
- Intensive care unit (ICU)
- Case Month (no specific pattern)

Model Limitations

Risks:

- Low precision &

  Many false positives

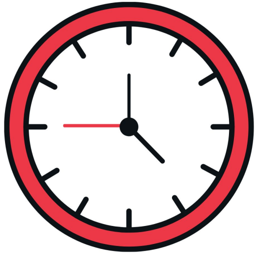
- What additional variables or screening might help overcome this without compromising our high recall?

Challenges

Many factors that contribute to complex health outcomes

Rarely measured or poorly understood:

- Variants
- Viral load
- Duration or intensity of exposure
- Severity of other diseases/complications



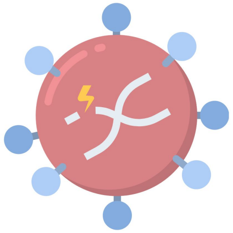
Not reported or not included in dataset

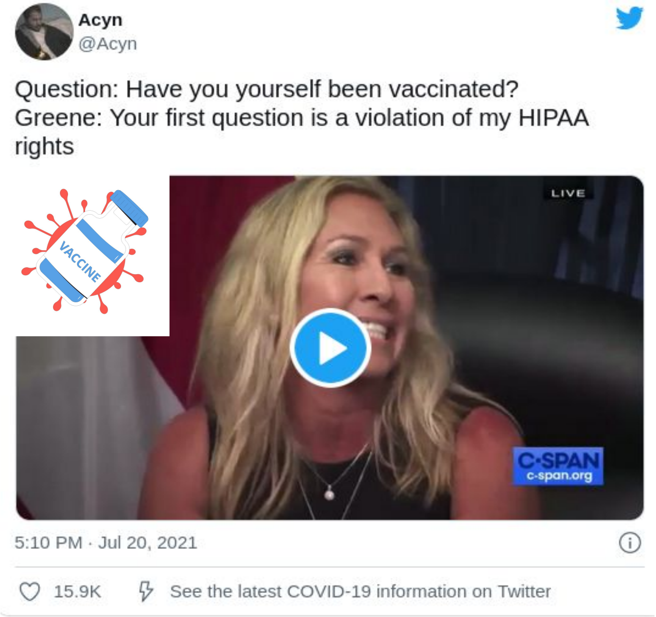
- Time from symptoms to hospitalization
- Duration of hospitalization
- Vaccination status

Another major element

- **Data missing or suppressed due to privacy rules**
- HIPAA! HIPAA! HIPAA!

- **This HURT the Model**





Acyn
@Acyn

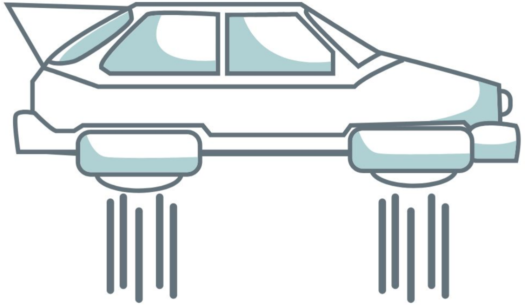
Question: Have you yourself been vaccinated?
Greene: Your first question is a violation of my HIPAA rights

5:10 PM · Jul 20, 2021

♡ 15.9K See the latest COVID-19 information on Twitter

Model applications

In the Future . . .



Potential screening tool for risk of death due to COVID-19

Concerns:

- High number of false positives
- Missing data and variables

Actions:

- Optimize to further improve specificity (false positives)
- Vaccination status - model would ideally include this. Vaccination substantially decreases risk of death

Thank You