INSTITUTE FOR DEFENSE ANALYSES

# DATAWorks 2021: Empirical Analysis of COVID-19 in the U.S.

Emily D. Heuring

April 2021

Approved for Public Release.
Distribution Unlimited.

IDA Document NS D-21549

Log: H 2021-000032

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task: BA-9-4869.0.0, "COVID-19 Empirical Tools and Analytical Support" for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of Rebecca Medlin, Matthew Avery, and Bram Lillard from the Operational Evaluation Division, and Kristina Guerrera and Zach Szlendak from the Cost Analysis and Research Division.

For more information:
Emily Heuring, Project Leader
eheuring@ida.org • 703-845-4374

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-21549

# DATAWorks 2021:
# Empirical Analysis of COVID-19 in the U.S.

Emily D. Heuring

# Executive Summary

The zoonotic emergence of the coronavirus SARS-CoV-2 at the beginning of 2020 and the subsequent global pandemic of COVID-19 has caused massive disruptions to economies and health care systems, particularly in the United States. This briefing to be presented at DATAWorks 2021 will describe IDA's empirical analysis of COVID-19 data within the U.S. general population.

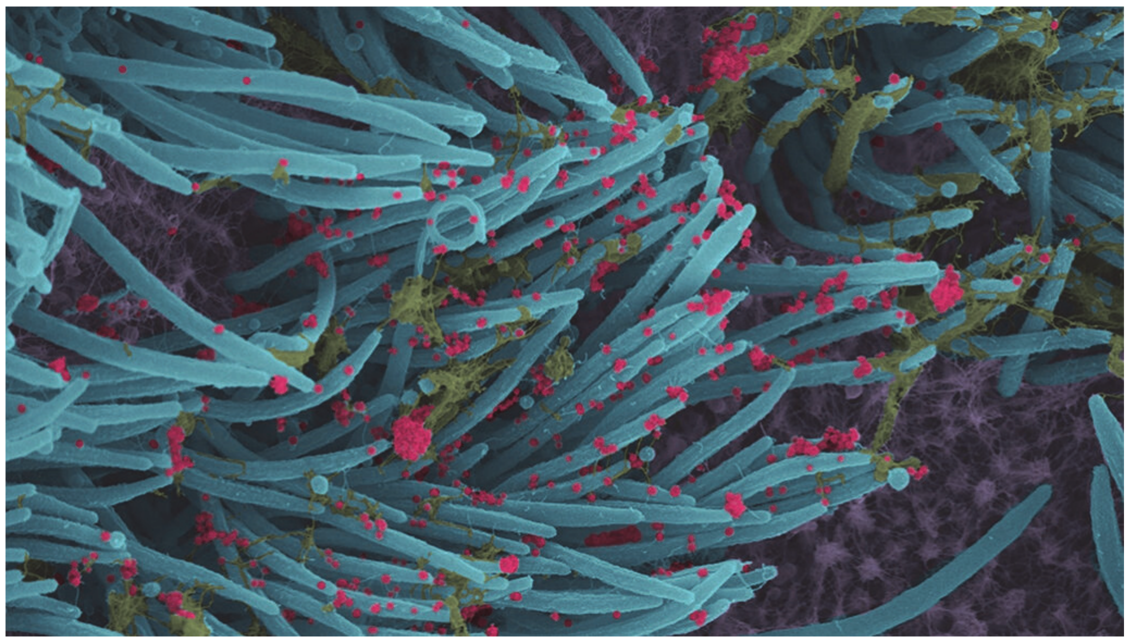
Using the results of serology testing, we have developed true prevalence estimates for COVID-19 case counts in the U.S. over time, which allows for more precise estimates of infection and case fatality rates throughout the course of the pandemic. To elucidate the behavioral, demographic, weather, and policy factors that contribute to or inhibit the spread of COVID-19, IDA compiled panel data sets of empirically derived, publicly available COVID-19 data. We then analyzed which factors were most highly correlated with both the increased and decreased spread within U.S. states and counties. Our analysis shows that mobility to retail and recreation locations (i.e., bars and restaurants) was highly correlated with increased COVID-19 case counts.

SARS-CoV-2 genomic analysis over the course of the pandemic demonstrates that different lineages have emerged and waned over the course of the pandemic.

Overall, the emergence of SARS-CoV-2 and COVID-19 has resulted in the public having unparalleled access to data for analyzing the pandemic. Such access has allowed for widespread real-time analysis to support the deployment of resources and implementation of policy. However, along with this unparalleled access to data come potential problems, as researchers may lack insight into data collection methods and experience with new types of data, which can lead to conflicting interpretations and conclusions about the pandemic.

# Empirical Analysis of COVID-19 in the U.S.

Dr. Emily Heuring



**Institute for Defense Analyses**
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

# Early in the pandemic, the sponsor identified a need for empirical data and analysis to support decision making

- Traditional epidemiology models (SEIR) show what can happen, not what will happen
  - Changes in behavior change the trajectory of the spread of the virus

- Sponsor identified a need to compile a large data set of empirical data from a variety of sources

- It is important to identify the factors most associated with both mitigation and spread of the virus

# COVID-19 is the first pandemic in history in which vast amounts of data are openly accessible to the public

# True Prevalence of COVID-19 vs Reported Case Counts

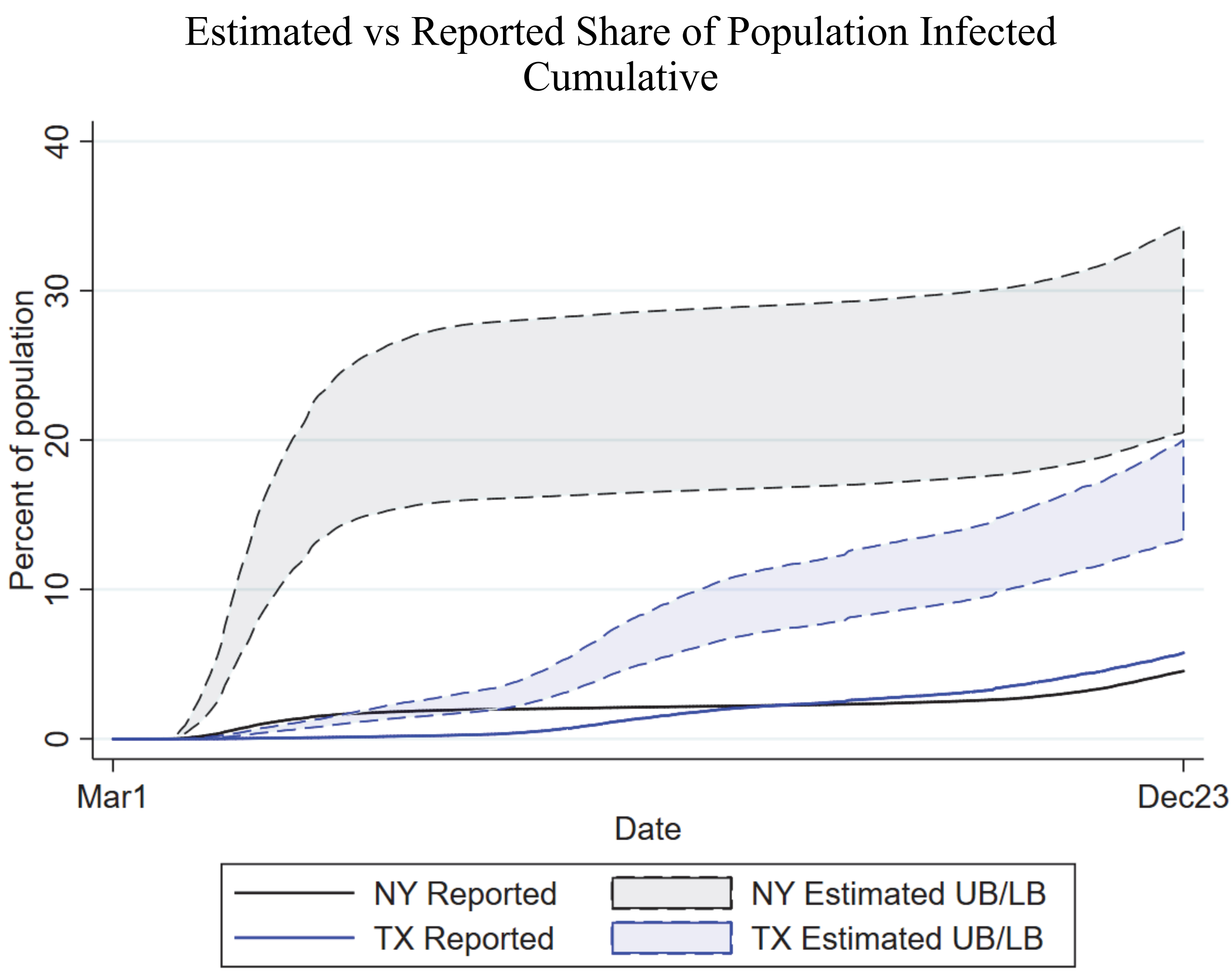
# COVID-19 case counts have been chronically underreported from the beginning of the pandemic

- Causes of underreporting:
  - Inadequate testing supplies and access during initial months
  - Those who had mild/asymptomatic cases weren't aware that they had COVID-19 and did not get tested

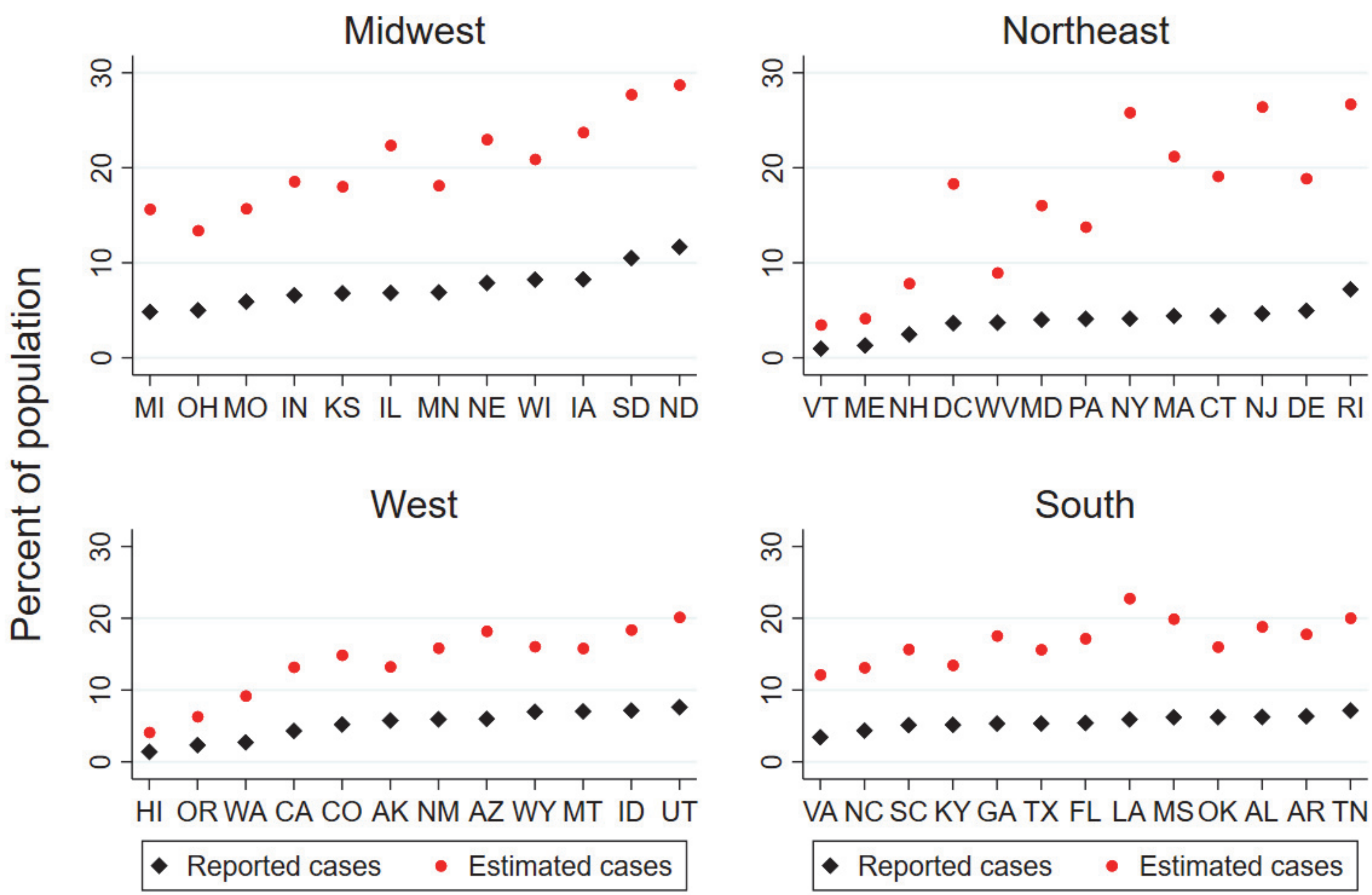
- We use serology data to estimate true COVID-19 prevalence in the U.S.
  - Serology testing detects antibodies to SARS-CoV-2 from previous infection
    - Although many city and state-level seroprevalence estimates were published, many had biased sampling techniques
    - We prefer CDC's seroprevalence methods: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/geographic-seroprevalence-surveys.html

- Compare the antibody prevalence to the number of reported cases at multiple points in time
  - Ten sample locations initially (has since been expanded)
  - Assume that for a given day, all regions in the U.S. had the same underreporting ratio
  - Assume no antibody decay (conflicting data in the literature)

# True prevalence estimates create a clearer picture of the pandemic

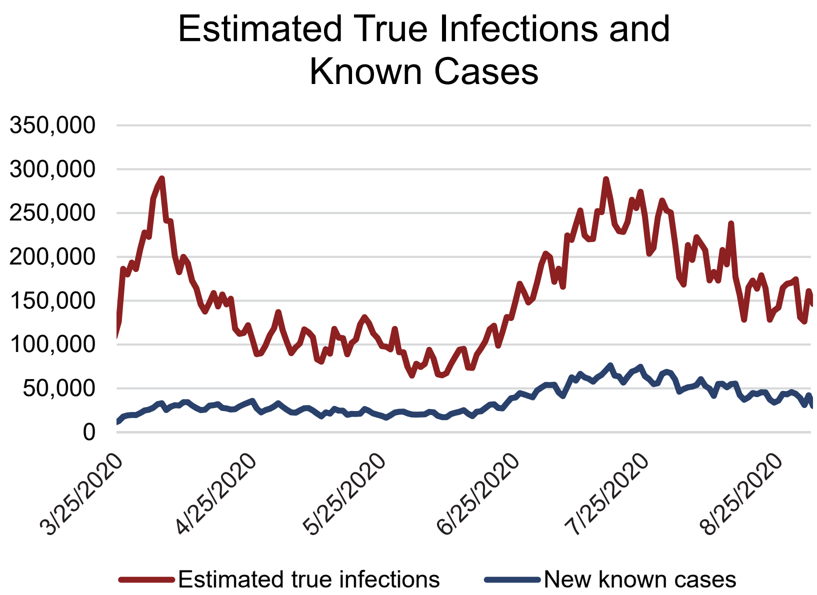


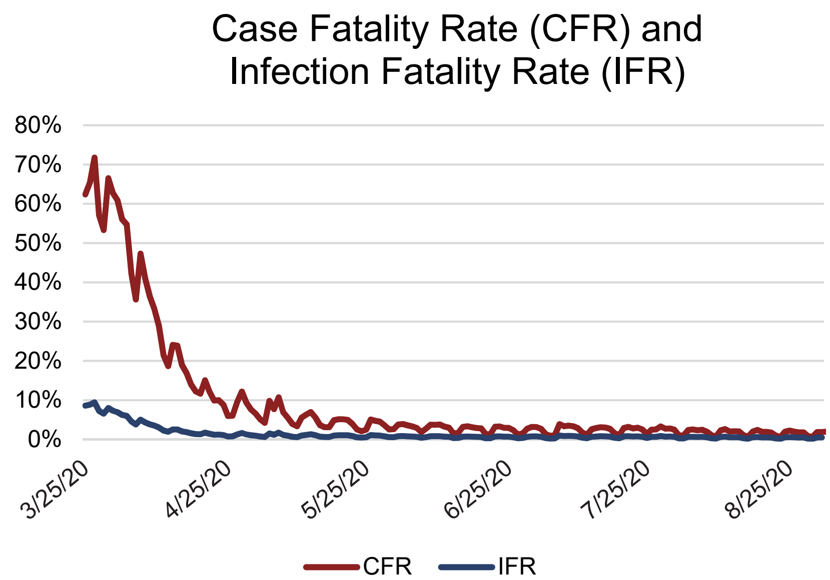
# Using seroprevalence data, we can estimate the percentage of the population in each state who have been infected with COVID-19




Data as of December 16

# With true prevalence estimates, we can calculate infection fatality rates for the U.S.




| **Period-average fatality rate** | **CFR** | **IFR** |
|---|---|---|
| Mar 26-Apr 30 | 31.56% | 3.69% |
| May 1-31 | 5.07% | 0.92% |
| Jun 1-30 | 2.79% | 0.62% |
| Jul 1-31 | 2.53% | 0.65% |
| Aug 1-Sep 8 | 1.66% | 0.46% |
| | | |
| ***Total deaths/total cases or infections during Mar 6-Sep 2*** | ***2.89%*** | ***0.69%*** |

# Effect of mitigation policies on the spread of COVID-19

# We examined how mitigation policies over the course of the pandemic affected COVID-19 case growth or decline

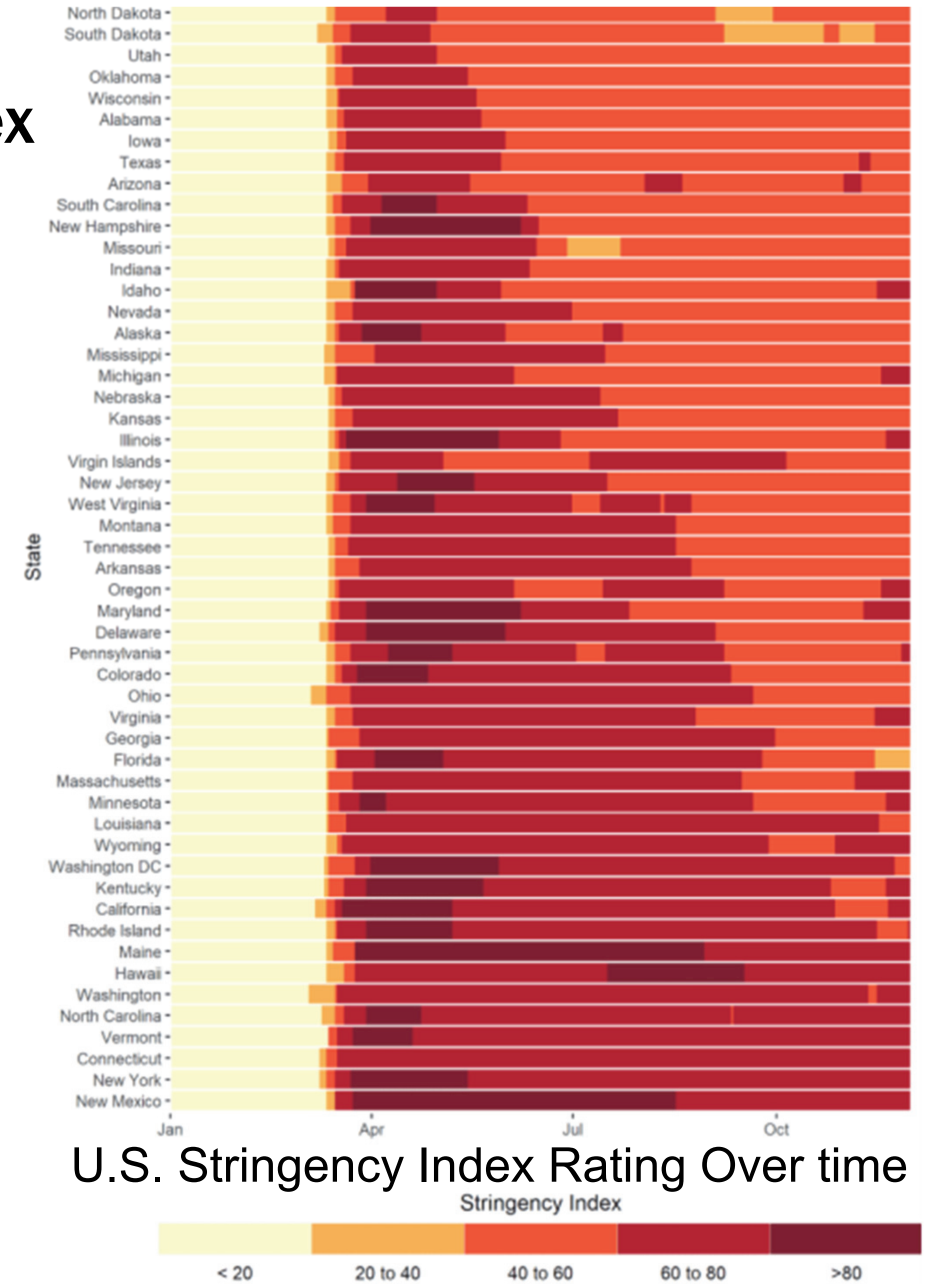
- To reduce bias from many mitigation policies being implemented simultaneously, we use a two-stage approach:
  - How policies affect mobility (first stage)
  - How changes in mobility from policies affect new cases (second stage)

- We find evidence that policies aimed at retail and recreation traffic <u>decrease</u> new cases

# Example policy data – Oxford's Policy Stringency Index

## Oxford US State Policy Stringency Index

| Measure | Name | Type | Targeted/general? |
|---|---|---|---|
| Containment and closure | | | |
| C1 | School closing | Ordinal | Geographic |
| C2 | Workplace closing | Ordinal | Geographic |
| C3 | Cancel public events | Ordinal | Geographic |
| C4 | Restrictions on gathering size | Ordinal | Geographic |
| C5 | Close public transport | Ordinal | Geographic |
| C6 | Stay-at-home requirements | Ordinal | Geographic |
| C7 | Restrictions on internal movement | Ordinal | Geographic |
| C8 | Restrictions on international travel | Ordinal | No |
| Health systems | | | |
| H1 | Public information campaign | Ordinal | Geographic |



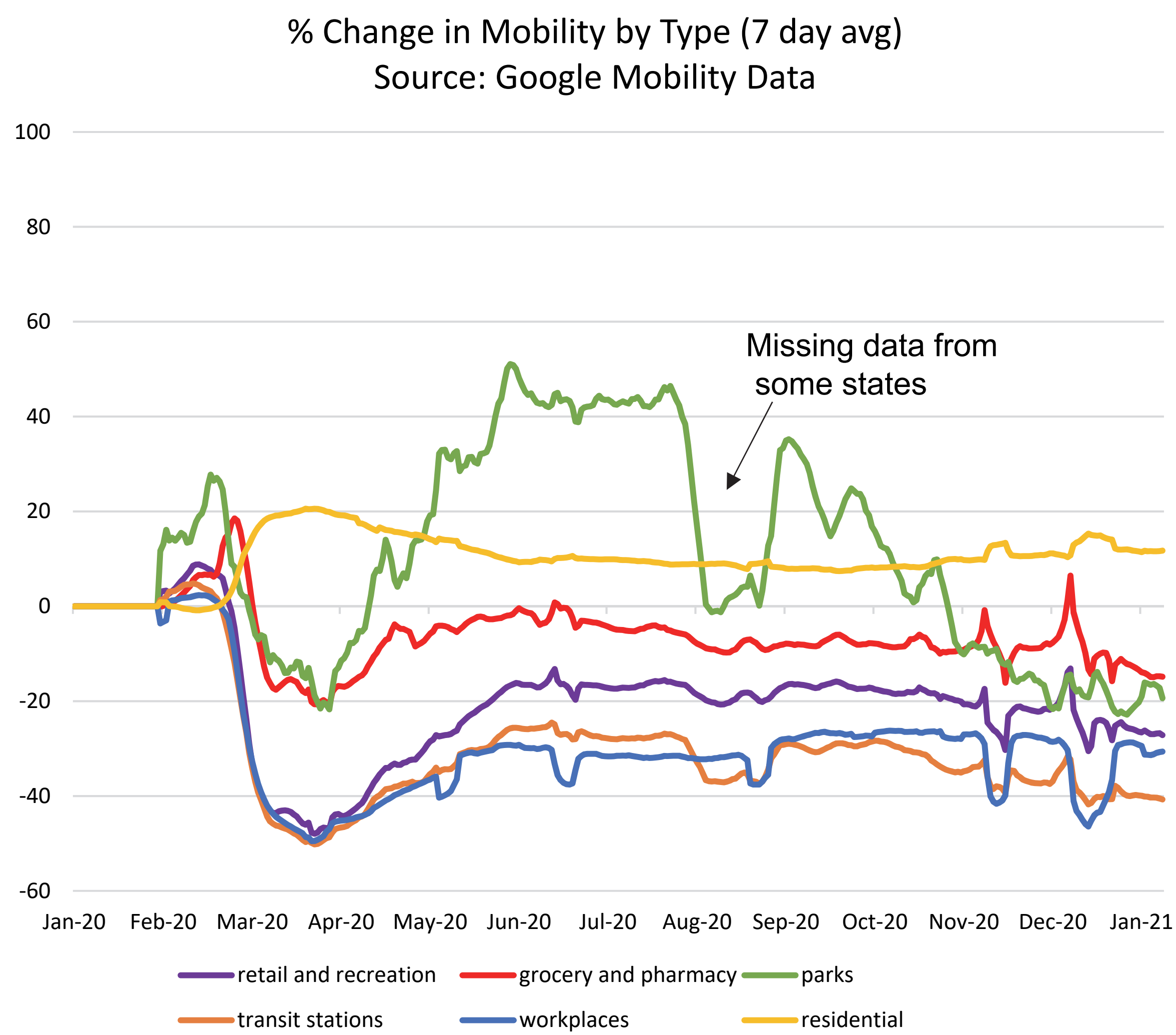

U.S. Stringency Index Rating Over time

Laura Hallas, Ariq Hatibie, Saptarshi Majumdar, Monika Pyarali, Thomas Hale. "Variation in US states' responses to COVID-19 2.0." Blavatnik School of Government Working Paper. Available: www.bsg.ox.ac.uk/covidtracker

IDA

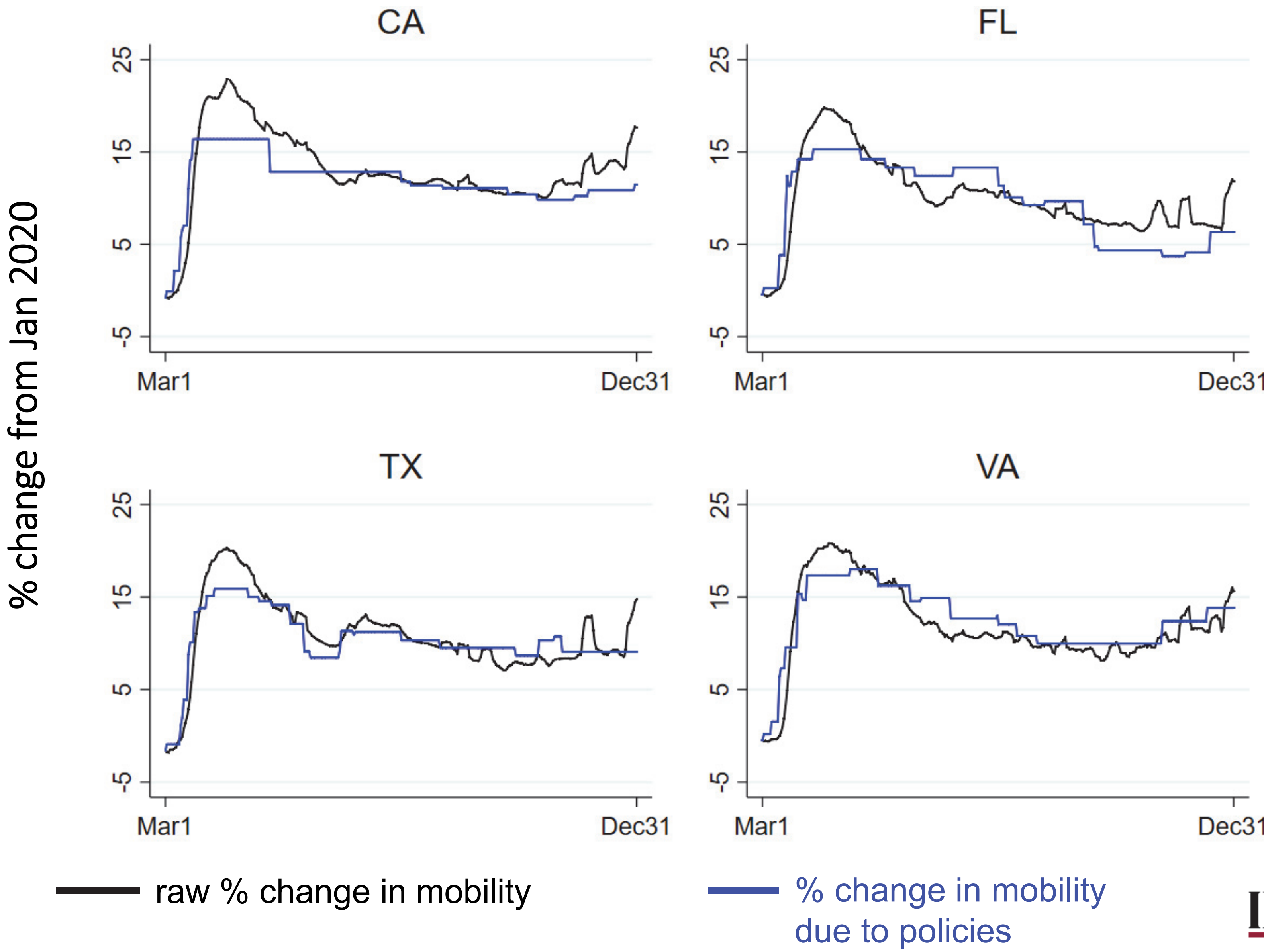
# In the U.S. in 2020, mobility to most places fell below baseline except in parks and residential locations

- Mobility to most places was below the baseline for most of 2020
- Parks and residential mobility was above baseline as people stayed home or went outside
- Gradual decline at end of CY20/early CY21 as cases increased



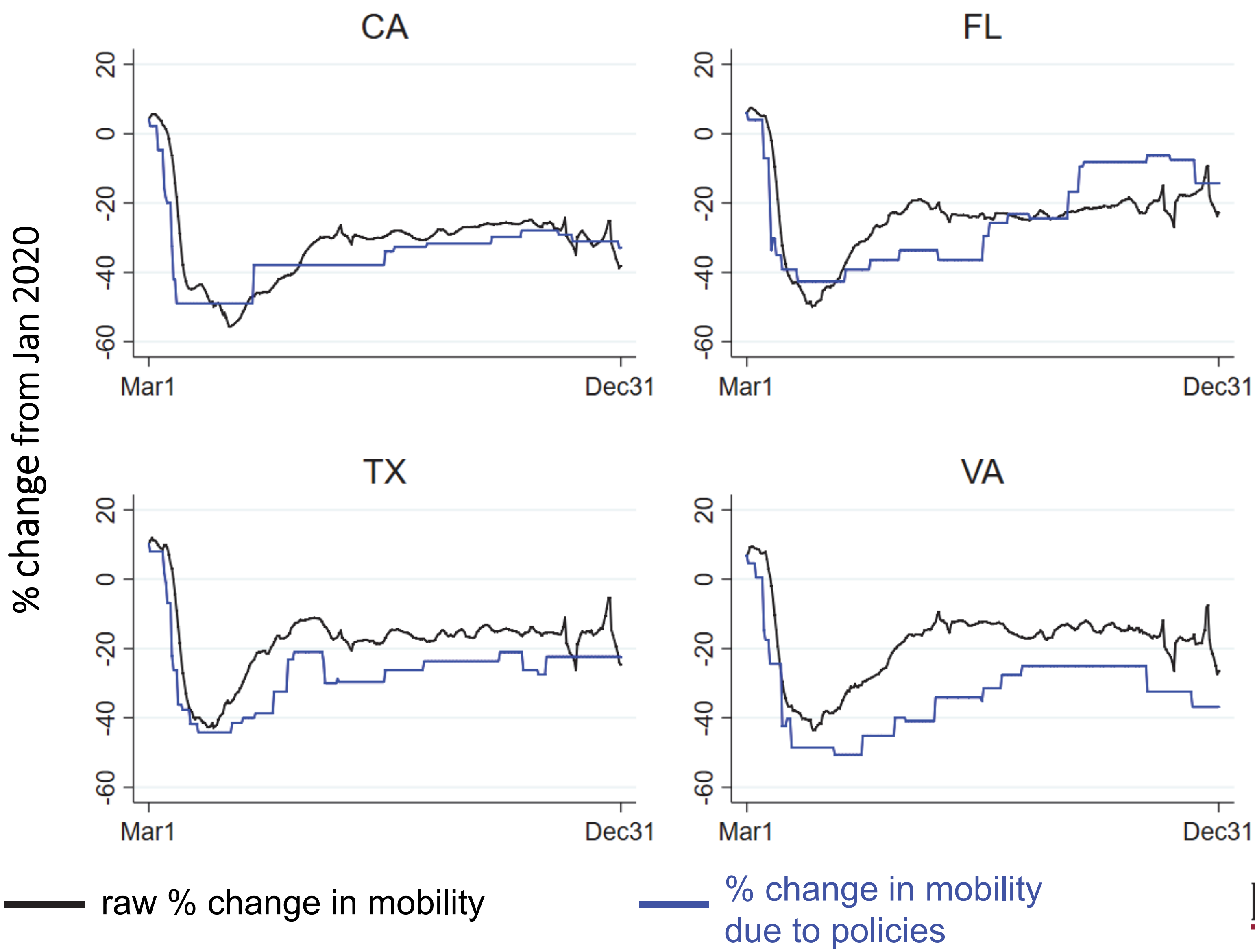

Use cell-phone mobility data as a proxy for adherence to mitigation policies

# Mitigation policy stringency is strongly correlated with increased residential mobility



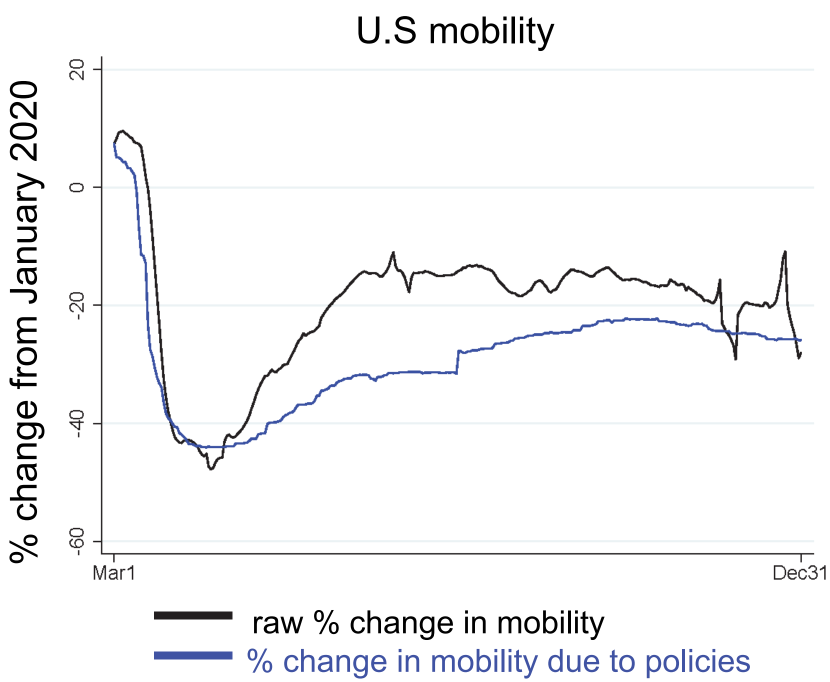

IDA

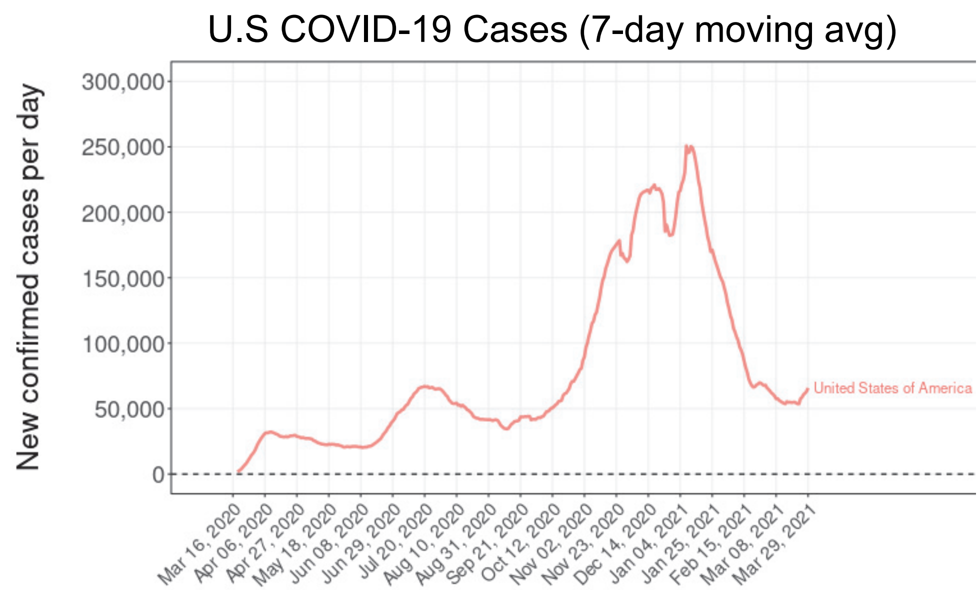
# Mitigation policy stringency is also strongly correlated with reduced mobility to retail and recreation locations




IDA

# How do reductions in mobility to retail and recreation locations affect COVID-19 case counts?




https://test-science.shinyapps.io/covid-19-tracker/

# Mitigation policies that decreased traffic to retail and recreation locations decreased COVID-19 case counts in the U.S.

First stage results

| U.S. | Avg Oxford Stringency Index Rating | Avg. percent change in retail & recreation mobility from policies | Avg Oxford Stringency Index Rating | Avg. percent change in retail & recreation mobility from policies |
|---|---|---|---|---|
| | March 2020 – July 2020 | | Aug 2020- Dec 2020 | |
| | 68.8 | -25.2% | 58.3 | -15.4% |

- Policies reduced traffic to retail and recreation locations by ~15-25%

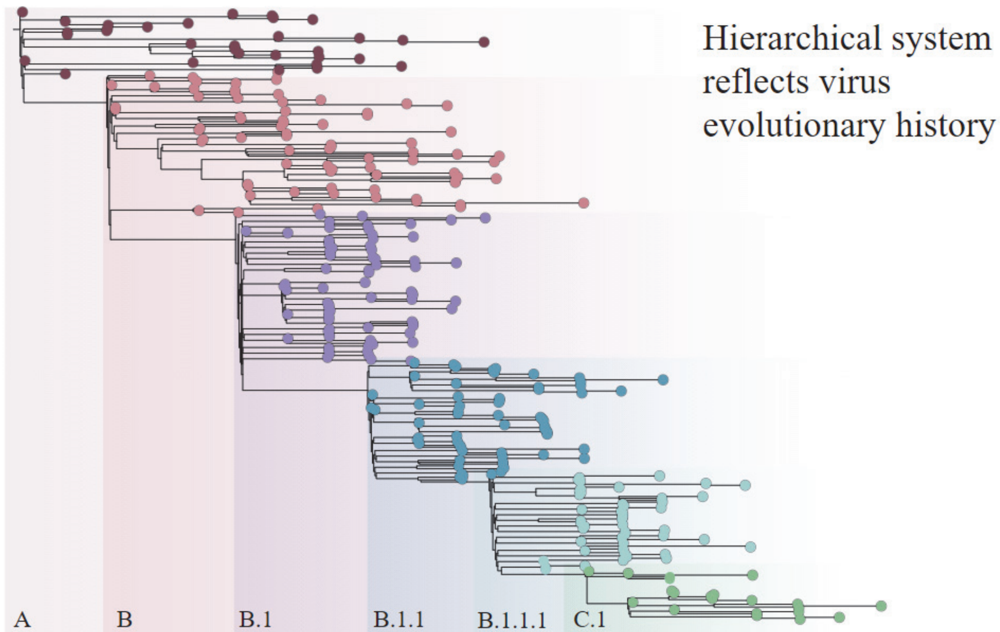
Second stage results

| U.S. | Avg. percent change in retail & recreation mobility from policies | Avg change in cases per 100k from policies | Avg. percent change in retail & recreation mobility from policies | Avg change in cases per 100k from policies |
|---|---|---|---|---|
| | March 2020 – July 2020 | | Aug 2020- Dec 2020 | |
| | 25.2% | -17.8 | 15.4% | -10.9 |

- The reduction in traffic to retail and recreation locations reduced COVID-19 case counts by ~10-17 cases per 100,000 per day

# Genomic sequencing surveillance

# Rapid, high-throughput genomic sequencing has resulted in unprecedented insight into how SARS-CoV-2 is evolving




*Note that C.1 is equivalent to B.1.1.1.1, with C being an alias for B.1.1.1 to shorten the name so it does not become infinitely long while still preserving the link to the parent lineages

**PANGO (Rambaut et al. 2020 *Nat Microbiol*) is the algorithm behind the new lineages making the news.**

**Lineages reflect:**

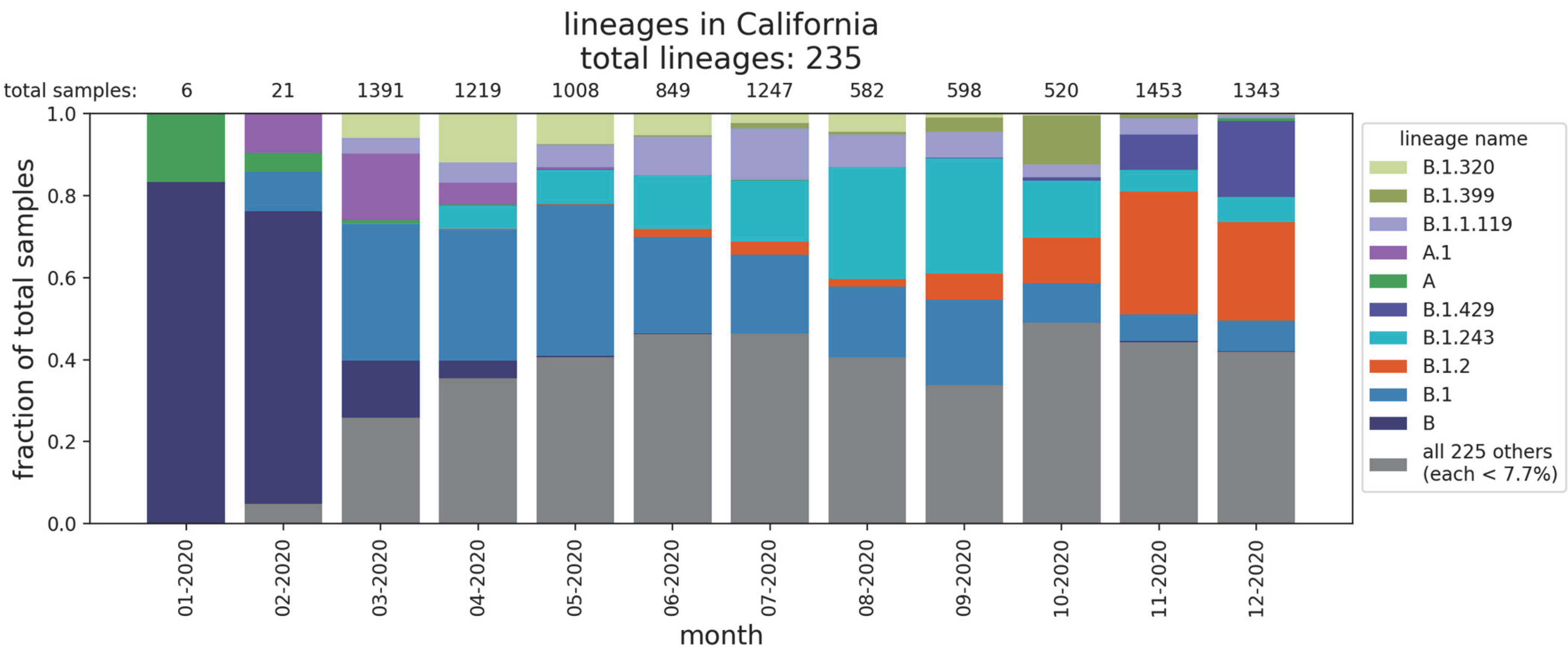
1) A virus's ancestry
2) Geographic location
3) A new mutation (in that lineage)

**Lineages DO NOT reflect:**

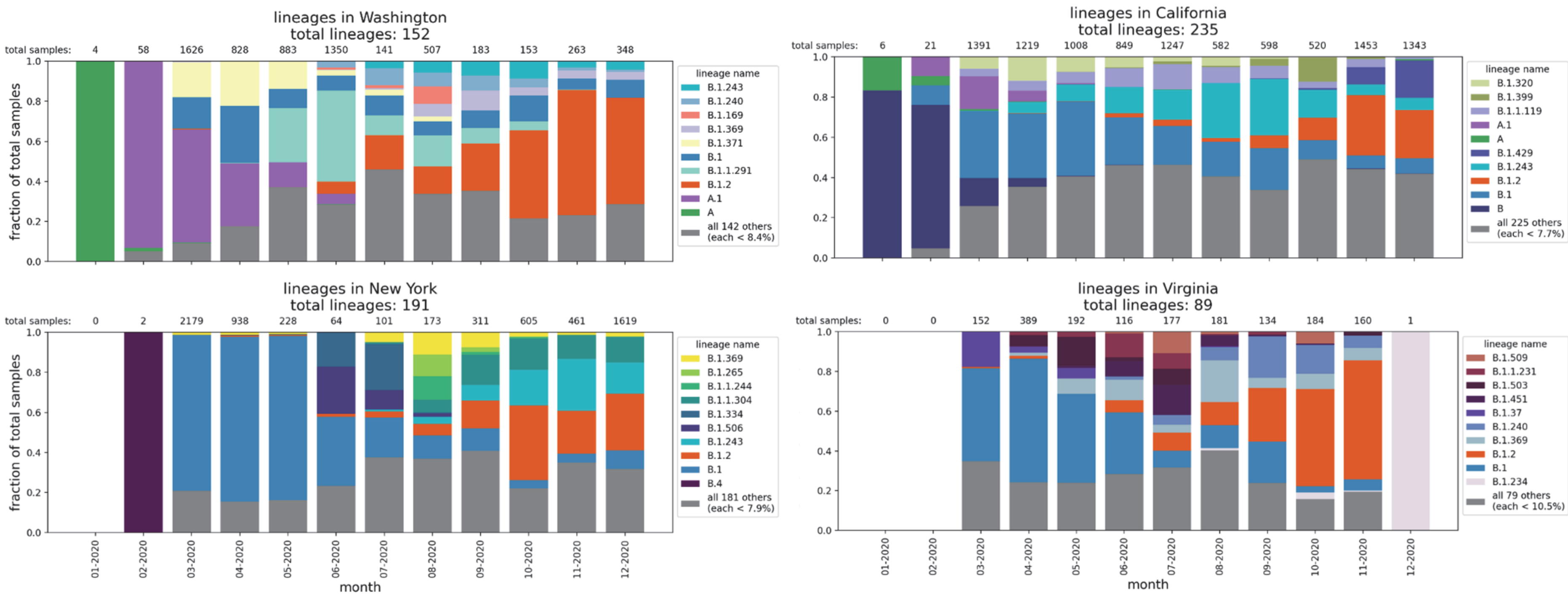
1) Whether a mutation is meaningful
2) Whether a mutation is unique (the same mutation can occur in different lineages)

IDA

# Using genomic sequence data, we can track lineages over time in the U.S.



# SARS-CoV-2 lineages have ebbed and flowed over the course of the pandemic in the U.S.




- It is easy to determine what a virus's genetic sequence is. It is more difficult to tell what the sequence means. ***Lineage names reflect sequence, not meaning***

- Things that may cause a lineage to be come predominant:
  - Founder effect – the reduction in genetic variation when a small subset of a large population is used to establish a new population
  - One or repeated superspreader events with that lineage
  - Small-number statistics/undersampling/biased sampling
  - A mutation with a biological mechanism that facilitates higher infectivity/spread

IDA

# Thoughts on a pandemic in the time of "Big Data"

- More data is good!
  - Allows for real-time analysis to support deployment of resources and policy implementation

- Where more data or new data might complicate things:
  - Lack of insight into data collection issues, important caveats, issues that might affect analysis and interpretation
    - Example: Early in the pandemic, some states were reporting both PCR and serology test results into a rolled-up number, providing a false picture of real-time spread of the virus

  - Lack of experience with certain data types
    - Example: Genomic sequencing generating concern about emerging variants

# Institute for Defense Analyses COVID-19 Team

Mr. Nathaniel Cleaves

Mr. Daniel Chaiken

Dr. Katherine Fisher

Dr. Kristen Guerrera

Mr. Ryan Murphy

Mr. Chris Oswald

Dr. Bryan Roberts

Dr. Zach Szlendak

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** *(DD-MM-YYYY)*
4-2021

**2. REPORT TYPE**
Final

**3. DATES COVERED** *(From - To)*

**4. TITLE AND SUBTITLE**
DATAWorks 2021: Empirical Analysis of COVID-19 in the U.S.

**5a. CONTRACT NUMBER**
HQ0034-19-D-0001

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Emily D. Heuring (OED)

**5d. PROJECT NUMBER**
BA-9-4869

**5e. TASK NUMBER**
4869

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

**8. PERFORMING ORGANIZATION REPORT NUMBER**
D-21549 NS

H 2021-000032

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
OSD/CAPE
1800 Defense Pentagon
Washington, DC 20301-1800

**10. SPONSOR/MONITOR'S ACRONYM(S)**
OSD-CAPE

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for Public Release. Distribution Unlimited.

**13. SUPPLEMENTARY NOTES**
Project Leader: Heuring, Emily D.

**14. ABSTRACT**
The zoonotic emergence of the coronavirus SARS-CoV-2 at the beginning of 2020 and the subsequent global pandemic of COVID-19 has caused massive disruptions to economies and health care systems, particularly in the United States. Using the results of serology testing, we have developed true prevalence estimates for COVID-19 case counts in the U.S. over time, which allows for more precise estimates of infection and case fatality rates throughout the course of the pandemic. To elucidate the behavioral, demographic, weather, and policy factors that contribute to or inhibit the spread of COVID-19, IDA compiled panel data sets of empirically derived, publicly available COVID-19 data and analyzed which factors were most highly correlated with increased and decreased spread within U.S. states and counties. These analyses can lead to some recommendations concerning future pandemic response preparedness.

**15. SUBJECT TERMS**
COVID-19, pandemic, empirical analysis, genomics

| 16. SECURITY CLASSIFICATION OF: a. REPORT | b. ABSTRACT | c. THIS PAGE | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)* |
|---|---|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited | 30 | Emily D. Heuring / (703) 845-4374 |

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39.18