COVID-19 surveillance and metrics: Data for decisions

Michael A. Stoto, PhD
GLID 4/502
April 13, 2021

GEORGETOWN UNIVERSITY

Week 2 - COVID-19 surveillance and metrics: Data for Decisions

- ***Introduction to infectious disease surveillance***
- COVID-19 metrics
  - what do we need to know?
  - problems with current metrics
  - science-based alternatives – estimates vs. counts

Surveillance definition and uses

- Definition (CDC): Ongoing, systematic collection, analysis, and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control
- Uses
  - Identify individual ⇒ and local-level interventions to control epidemics
    - individual level: case finding, patient tracking and linking to care, partner notification, contact tracing
    - local level: identification and removal of contaminated food sources, environmental pollutants, COVID-19 superspreader sites
  - Identify issues ⇒ population-level interventions
    - epi research (routes of transmission), implementation of targeted programs, professional and public education
    - allocation of resources for programs, including formula funding

Case-based surveillance

- Focus on detecting individual cases of infectious disease and taking action
  - Monitoring, treatment, quarantine, contact tracing
- In US, largely a state responsibility
  - Federal reporting, interstate commerce
  - Implemented at the local level
- Basic mechanisms in place in 19th century
  - "Notification" required
  - Post card reporting
  - Weekly reports
  - Electronic reporting being developed
- *What about the HIPAA Privacy Rule?*

Broadening focus of surveillance

- Until 1950, “surveillance” restricted in public health practice to
  - Monitoring persons with serious communicable diseases and their contacts
  - Taking action with these individuals (or populations) to prevent further spread
- Since 1950s “surveillance” includes “statistical” aspects
  - Increasing interest in chronic diseases, etc.
  - Behavioral, environmental, other risk factors
  - *Healthy People 2000, 2010, 2020*
- Tensions between case and statistical surveillance approaches
  - HIV/AIDS, syndromic surveillance, COVID-19 metrics

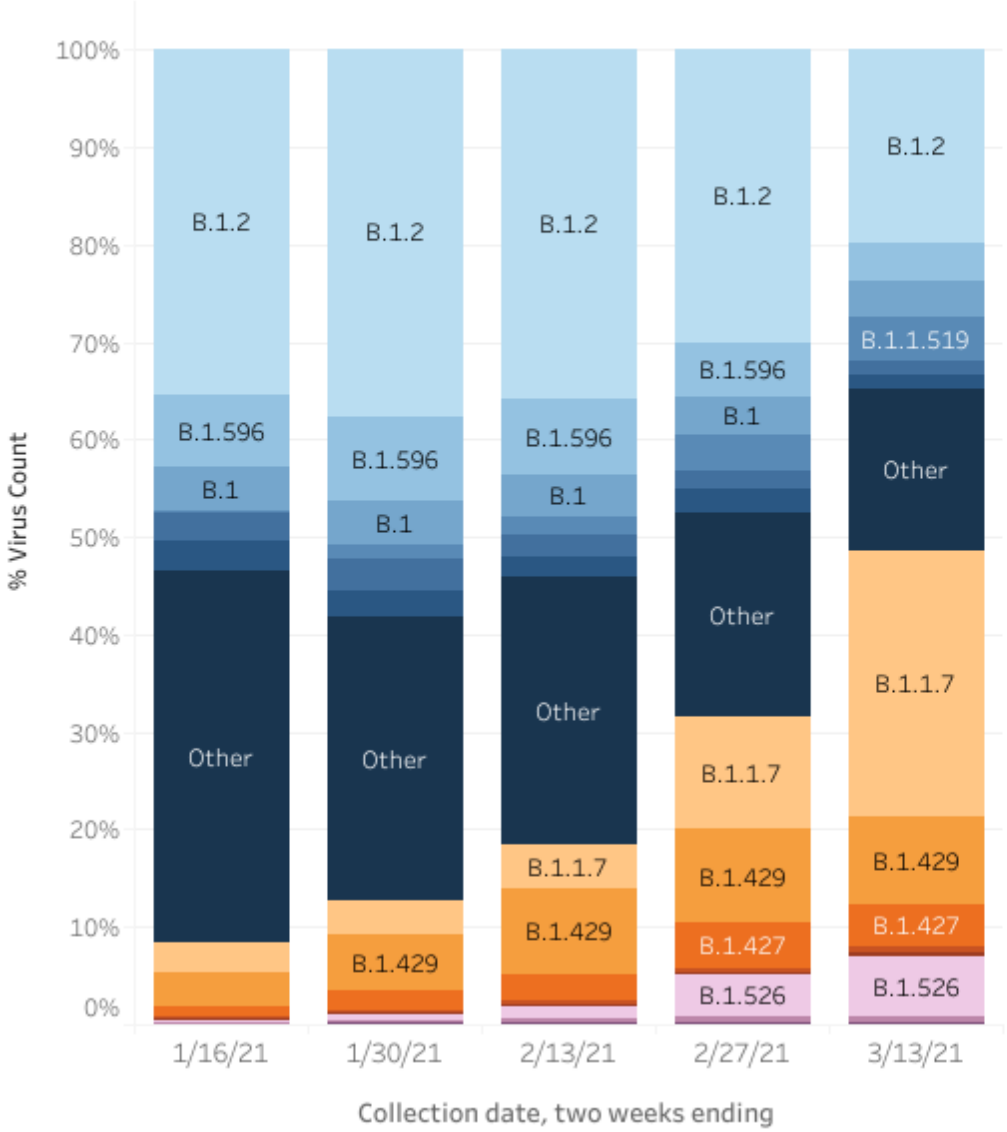
Indicator-based surveillance

- Traditional case-based surveillance
- Virological & genomic surveillance
- Syndromic surveillance, e.g. of hospital EDs
- Sentinel surveillance, e.g. statistical reports from sensitive sites

- Characterized by standardized, structured information, e.g. reports received on a regular basis and entered routinely into a disease-reporting database on the number of laboratory-confirmed cases of influenza identified at a hospital laboratory

Virologic/genomic surveillance

SARS-CoV-2 Variants Circulating in the United States, January 3 – March 13 2021



	Lineage	% Total	95%CI	Type
Most common lineages	B.1.1.7	27.2%	24.6-29.9%	VOC
	B.1.2	19.8%	17.3-22.6%	
	B.1.429	9.1%	6.8-12.1%	VOC
	B.1.526	6.2%	4.3-8.8%	VOI
	B.1.1.519	4.7%	3.8-5.8%	
	B.1.427	4.3%	3.3-5.6%	VOC
	B.1.596	3.9%	3.2-4.7%	
	B.1	3.7%	2.9-4.6%	
	B.1.234	1.4%	1.0-1.9%	
	B.1.243	1.4%	1.1-1.8%	
Additional VOI/VOC lineages	P.2	0.5%	0.3-0.7%	VOI
	P.1	0.5%	0.3-0.8%	VOC
	B.1.351	0.5%	0.3-0.7%	VOC
	B.1.525	0.3%	0.2-0.4%	VOI
Other*	Other	16.6%	14.4-19.2%	

Summary data that appear in the table include specimen collection dates from February 28 through March 13, 2021.
* Other represents >200 additional lineages, which are each circulating at <2% of viruses

Estimated weights based on # of RT-PCR tests performed & positive results stratified by state, specimen collection date, and by genomic surveillance data source.

Event-based surveillance

- Monitoring reports, stories, rumors, and other information about unusual health events that could signal outbreak or serious public health risk
  - unstructured information
  - non-standardized or subjective
- Information can come from
  - reports in the media
  - rumors on an internet blog
  - community, e.g.
    - reported by through a hotline
    - teacher notices an unusually high number of children absent from school with similar symptoms and reports it to a local health official

Published Date: 2019-12-30 23:59:00
Subject: PRO/AH/EDR> Undiagnosed pneumonia - China (HU): RF
Archive Number: 20191230.6864153

UNDIAGNOSED PNEUMONIA - CHINA (HUBEI): REQUEST FOR INFORMATION

****

Wuhan unexplained pneumonia has been isolated test results will be announced [as soon as available]

On the evening of [30 Dec 2019], an "urgent notice on the treatment of pneumonia of unknown cause" was issued, which was widely distributed on the Internet by the red-headed document of the Medical Administration and Medical Administration of Wuhan Municipal Health Committee.
On the morning of [31 Dec 2019], China Business News reporter called the official hotline of Wuhan Municipal Health and Health Committee 12320 and learned that the content of the document is true.
12320 hotline staff said that what type of pneumonia of unknown cause appeared in Wuhan this time remains to be determined.
According to the above documents, according to the urgent notice from the superior, some medical institutions in Wuhan have successively appeared patients with pneumonia of unknown cause. All medical institutions should strengthen the management of outpatient and emergency departments, strictly implement the first-in-patient responsibility system, and find that patients with unknown cause of pneumonia actively adjust the power to treat them on the spot, and

Surveillance data *not* a substitute for epidemiological studies

(Lipsitch *et al*., *NEJM*, February 2020)

Types of Evidence Needed for Controlling an Epidemic.

Evidence Needed	Study Type
No. of cases, including milder ones	Syndromic surveillance plus targeted viral testing
Risk factors and timing of transmission	Household studies
Severity and attack rate	Community studies
Severity "pyramid"	Integration of multiple sources and data types
Risk factors for infection and severe outcomes, including death	Case–control studies
Infectiousness timing and intensity	Viral shedding studies

Week 2 - COVID-19 surveillance and metrics: Data for Decisions

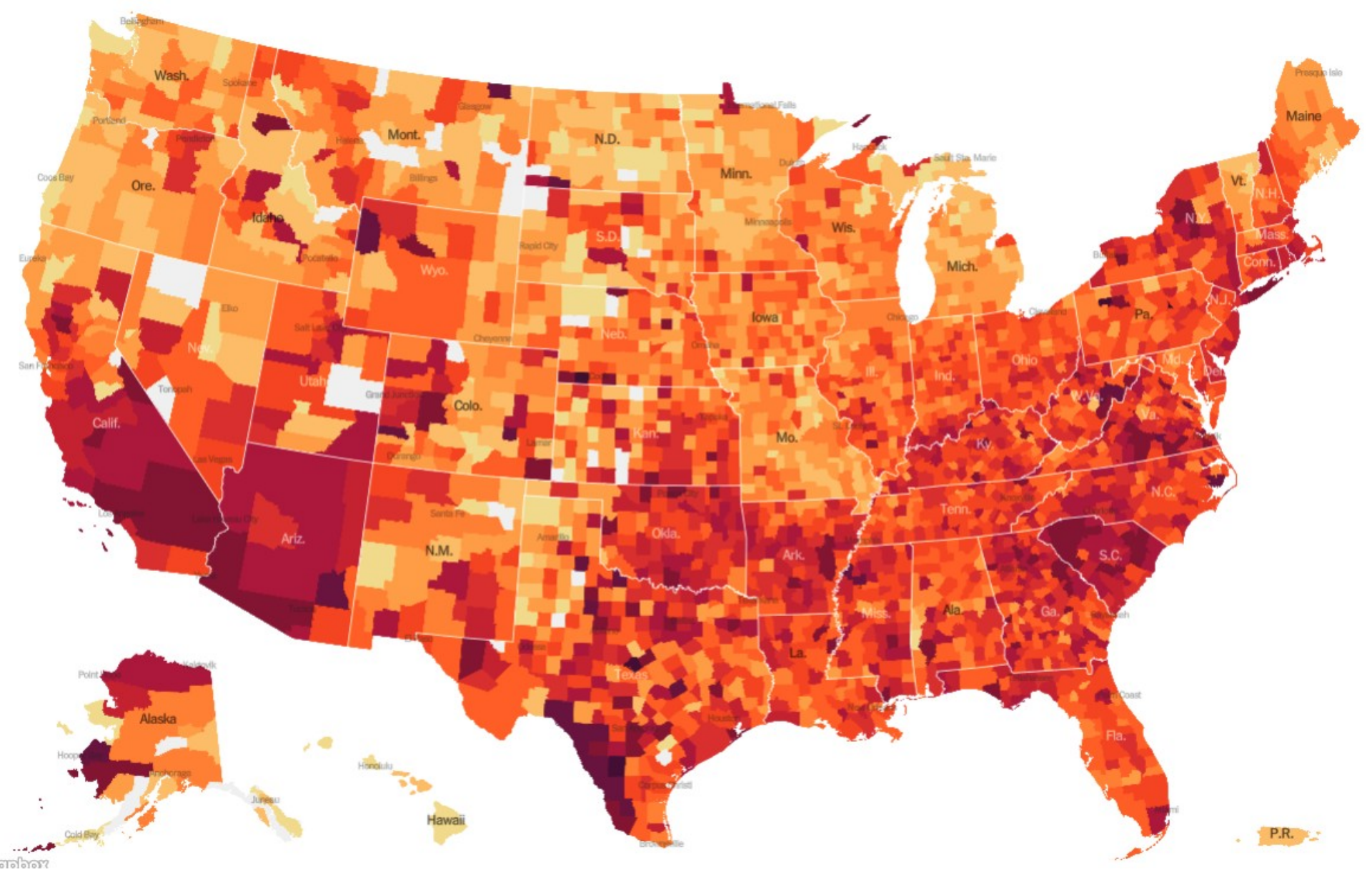
- Introduction to infectious disease surveillance
- ***COVID-19 metrics***
  - ***what do we need to know?***
  - problems with current metrics
  - science-based alternatives – estimates vs. counts

Better metrics for controlling COVID-19

- As communities adjust their policies to prevent the spread of COVID-19, the focus is on science-based and objective "metrics"
- To meet this demand, data are published by many different government health agencies, universities, and the media
  - These data sources vary in terms of which metrics they use and how they are defined
    - moreover, the definitions change over time
  - Many also have well-known biases such as "uncounted deaths"
  - As a result, despite all of these data, we seem to be "flying blind" in the fight against COVID-19

What do we need to know?

- When deciding on control strategies, metrics are needed for two distinct purposes
  - First, to gauge the *level of infections*
    - For example, we use the per capita rate of new infections to inform decisions about opening schools, etc., ("is it safe enough to …")
    - For example, these two maps show the number of new cases per 100,000 people in
      - the U.S. and California
      - at the end of January and mid-April, 2021



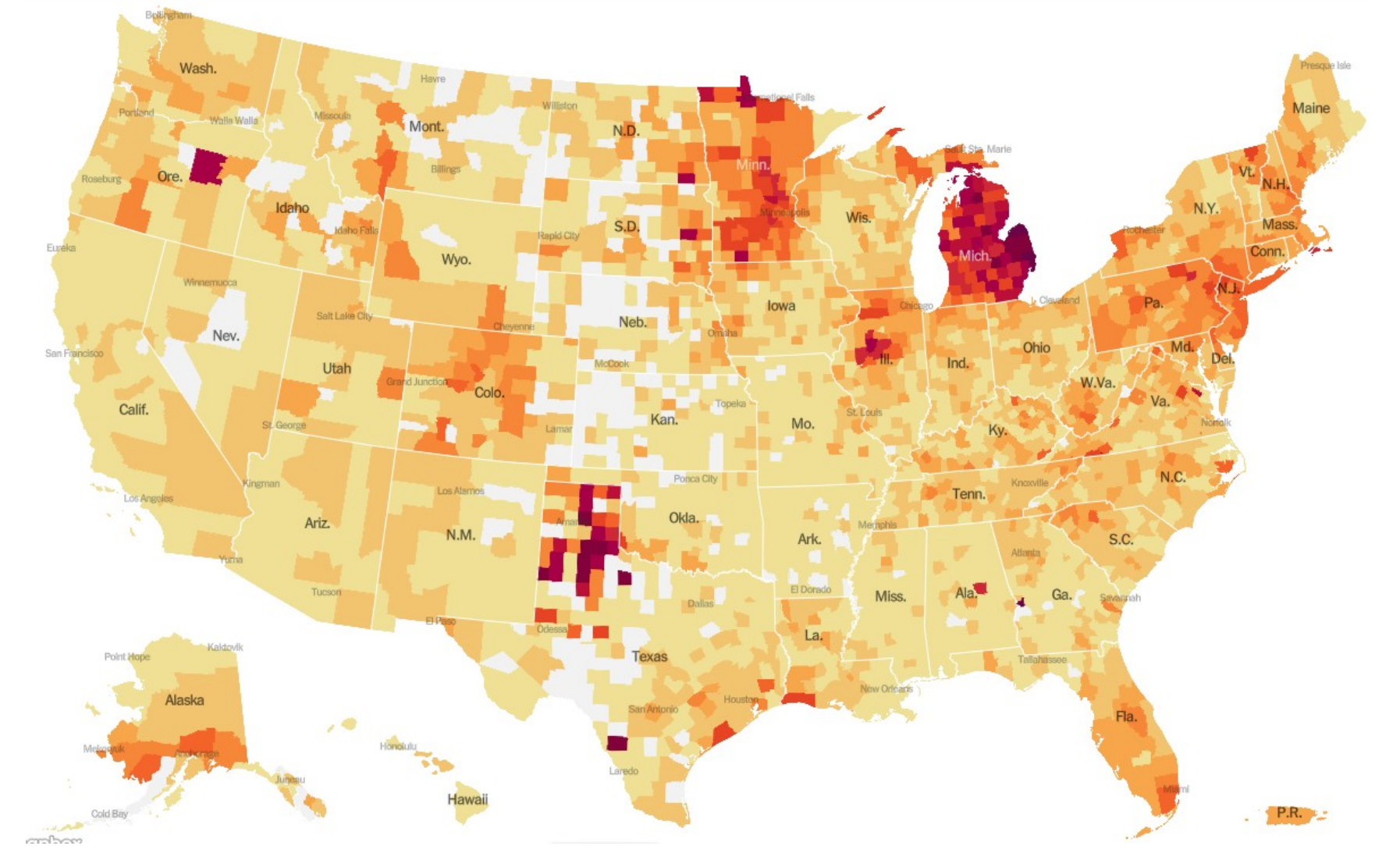
The New York Times

January 22, 2021

Average daily cases per 100,000 people in past week

10 30 50 70 100 250 Few or no cases

U.S. COVID-19 situation – April 13, 2021



Average daily cases per 100,000 people in past week

10 30 50 70 100 250 Few c

	TOTAL REPORTED	ON APRIL 12	14-DAY CHANGE
Cases	31.2 million+	72,286	+6% →
Deaths	562,007	476	–27% →
Hospitalized		43,735	+10% →

California Blueprint

Tier 1	Tier 2	Tier 3	Tier 4
Widespread Most nonessential indoor business operations are closed	**Substantial** Some nonessential indoor business operations are closed	**Moderate** Some business operations can open with modifications	**Minimal** Most business operations can open with modifications

January 30, 2021

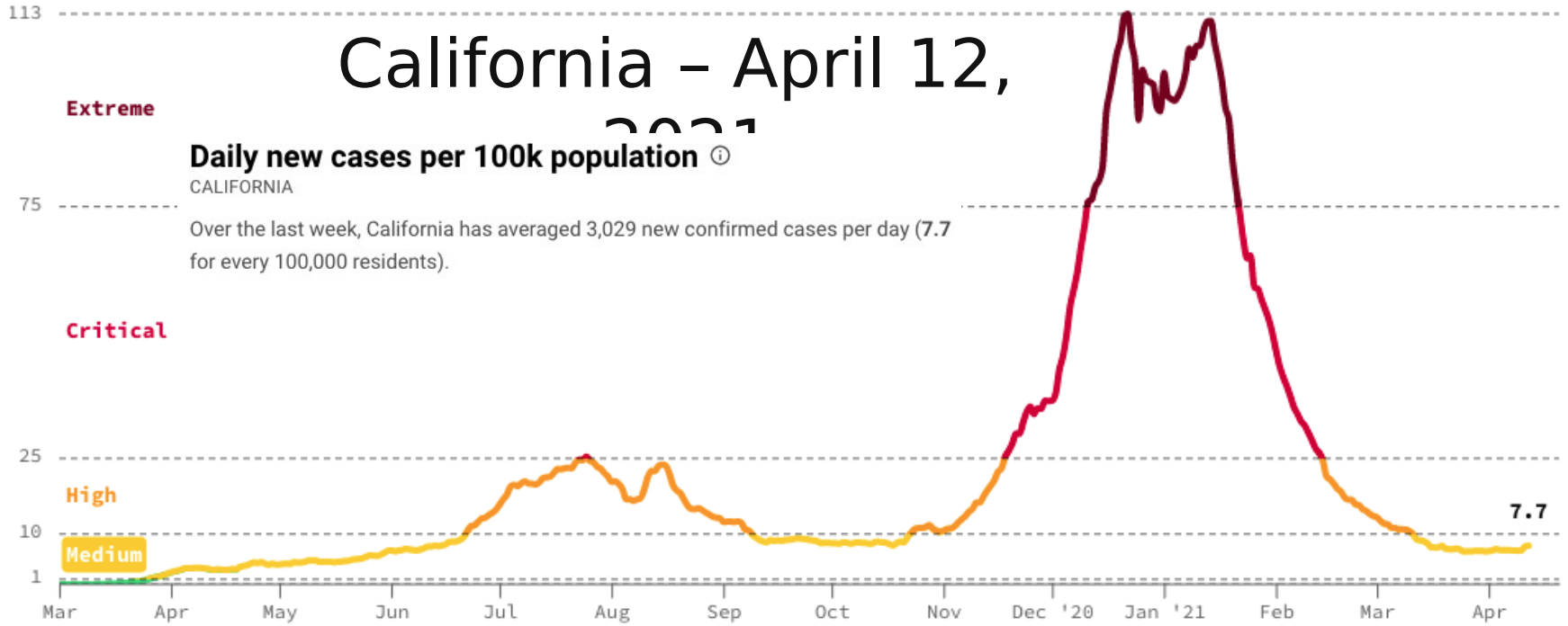
April 12, 2021

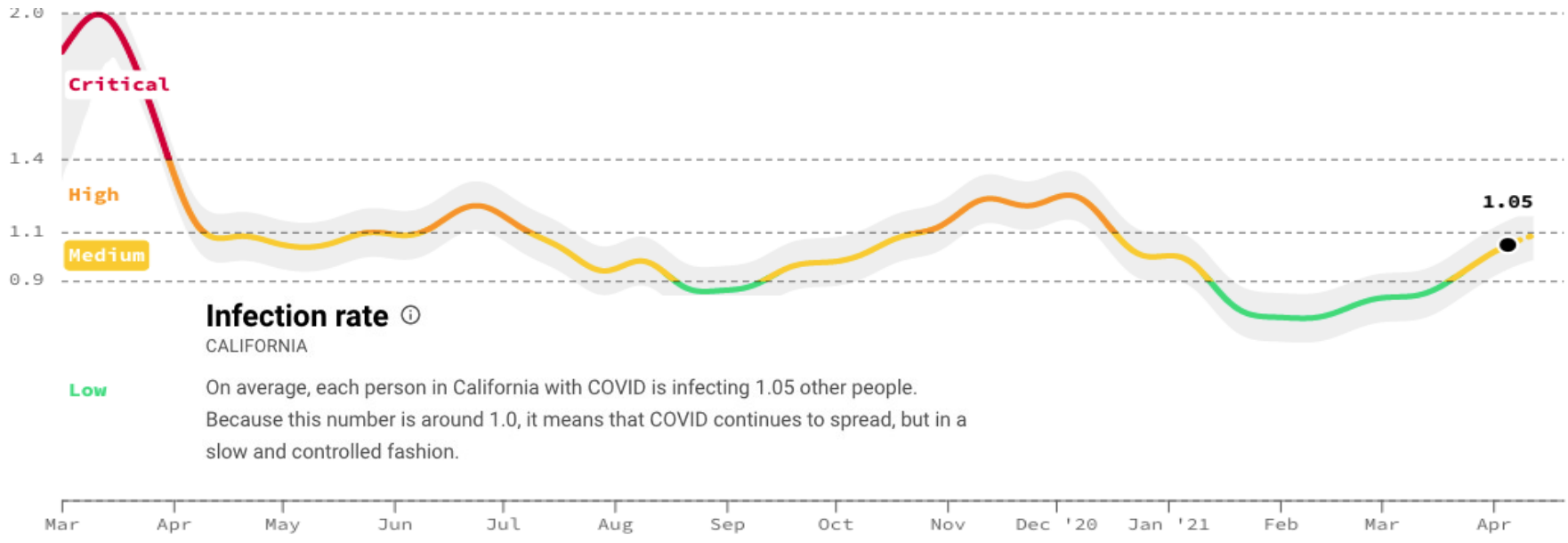
What do we need to know?

- Metrics are needed for two distinct purposes
  - First, to gauge the *level of infections*
  - Second, to assess the *rate of increase in new infections*
    - This allows us to monitor the effectiveness of masking, social distancing and other population-level control measures ("are they working …")
    - so policy makers know when to dial them up or down
    - For example, the next two figures show trends in California and Michigan in terms of the change over 14 days and the estimated reproduction rate R_t

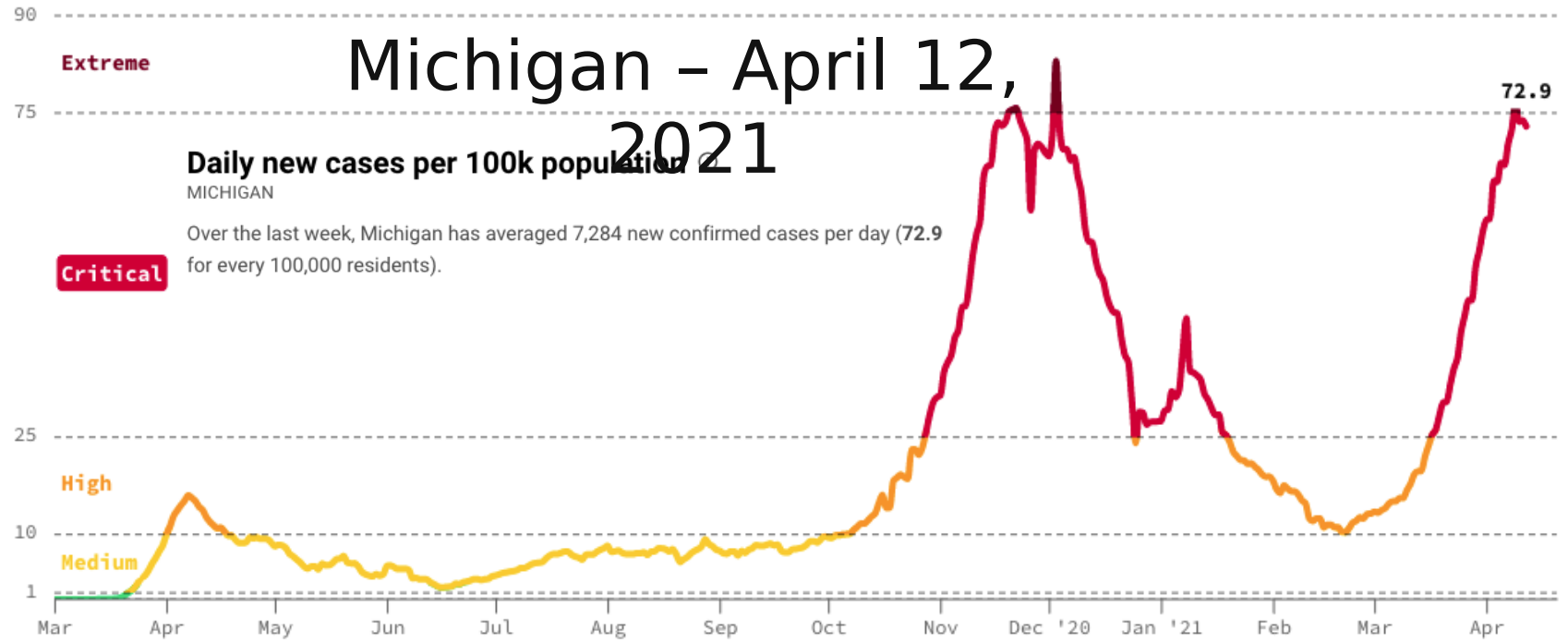
      > 1 indicates that the epidemic is growing

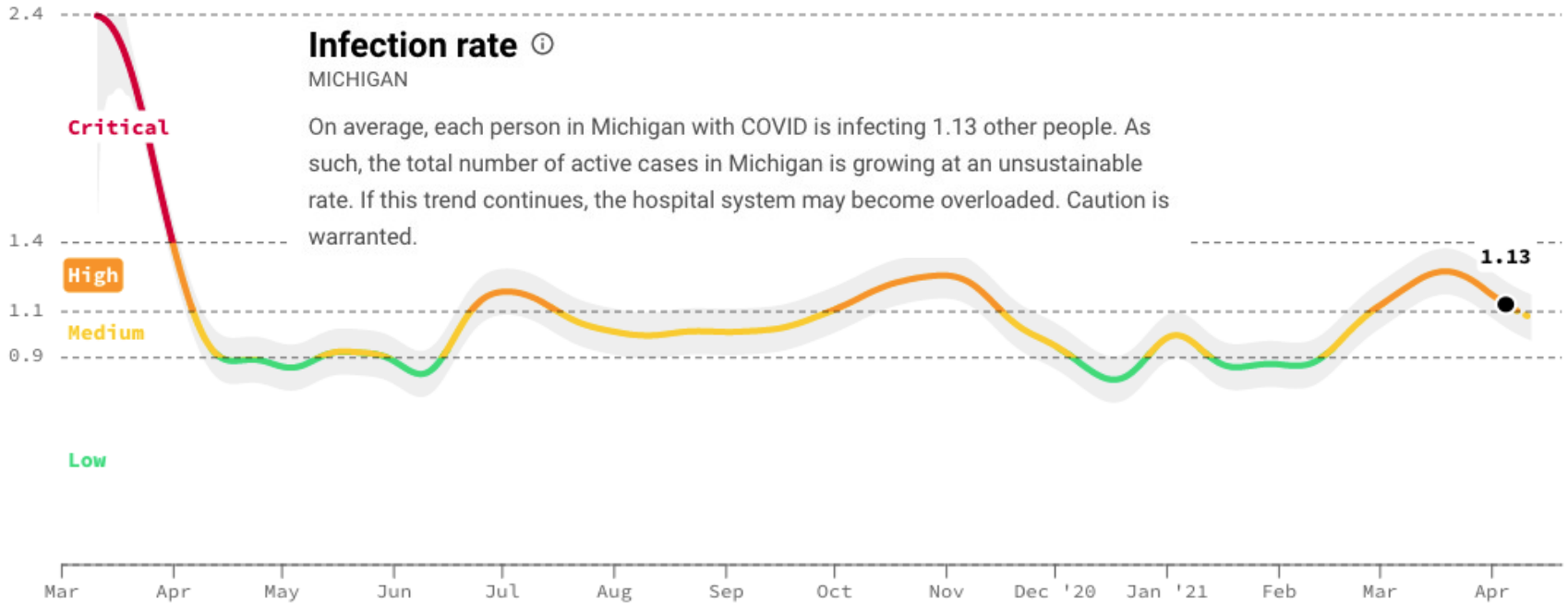
California – April 12, 2021





Michigan – April 12, 2021





What do we need to know?

- Metrics are needed for two distinct purposes
  - First, to gauge the *level of infections*
  - Second, to assess *rate of increase in infections*
  - These are distinct but obviously related over time
- Need to know these both
  - for the entire population of a community
  - for specific subgroups so policy makers know which populations to focus on
    - nursing homes, colleges and universities, and so on
    - by age, because of impact on schools
    - by groups defined by race, ethnicity, socioeconomic status, occupation, etc.
- Other metrics assess capacity for testing and contact tracing, hospital surge capacity, vaccine deployment, …
  - but are beyond scope of this presentation

Week 2 - COVID-19 surveillance and metrics: Data for Decisions

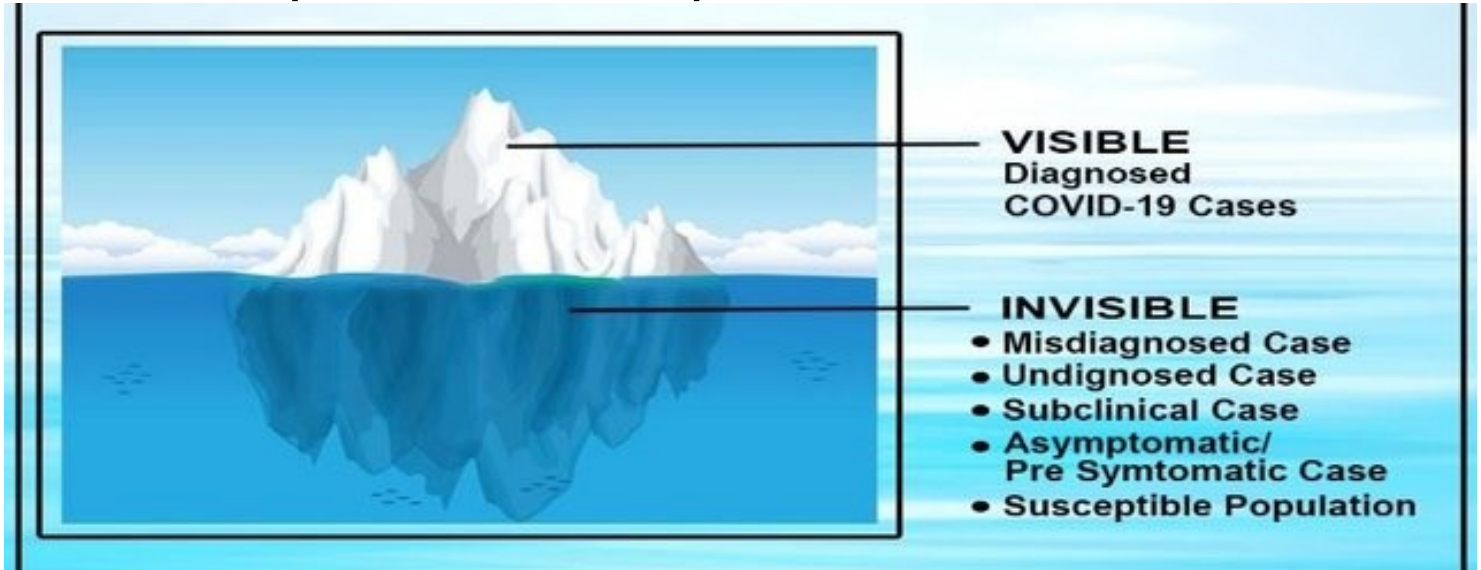
- Introduction to infectious disease surveillance
- COVID-19 metrics
  - what do we need to know?
  - ***problems with current metrics***
  - science-based alternatives – estimates vs. counts

Where do the data come from?

- Mostly from “case surveillance”
  - doctors who diagnose a “case” notify health department
  - which then takes steps to control it
- Contact tracing requires identifying specific individuals who have the disease
  - cases who are symptomatic and/or test positive
  - their contacts while they were infectious
- These data facilitate epidemiologic investigations
  - characterize clinical disease course and factors influencing risk of transmission, including socio-demographic factors
  - identify local transmission risks (e.g. specific locations such as bars where super-spreader events may occur)
  - health officials take action, either focused on the specific location, or changing policy, e.g. delay reopening bars
- But case surveillance data have problems

Problems with reliance on reported cases

- Iceberg effect: # reported cases < # infected
  - individuals with mild or no symptoms, depending on
    - whether they seek care & referred to testing
    - test availability
  - proportion of cases reported varies ("iceberg bobs")
    - over time (changing test availability, etc.)
    - between states based on differences in definitions, policies, systems for reporting, etc.
  - similar problems for reported deaths

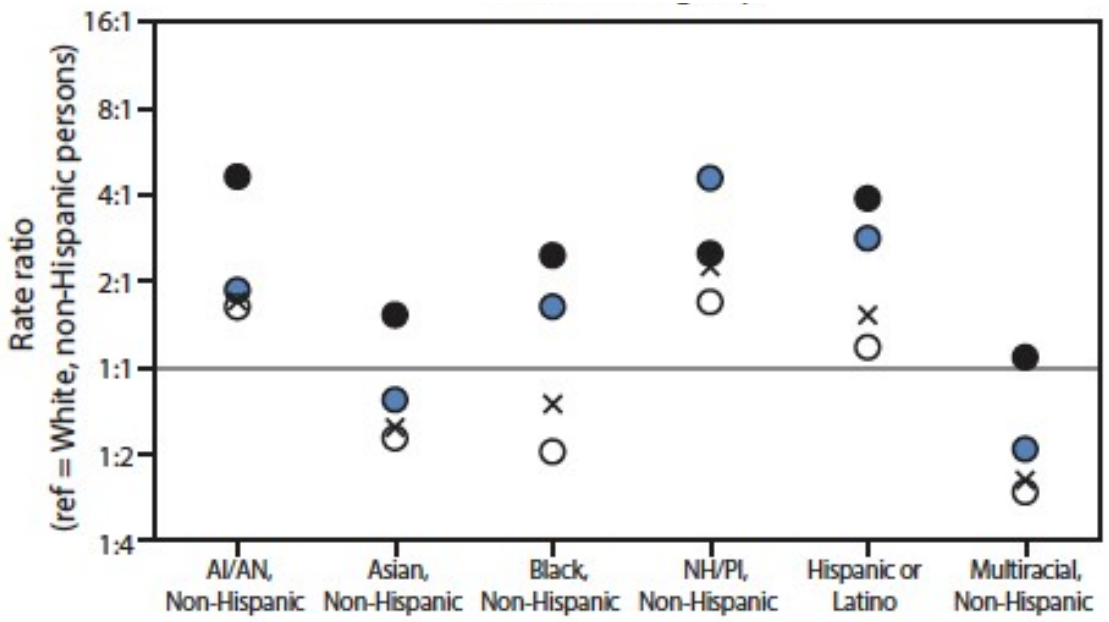


Problems with reliance on reported cases

- Socio-demographic data not needed for operational purposes (e.g. contact tracing)
  - *so often not available for statistical purposes*

Centers for Disease Control and Prevention
MMWR
Morbidity and Mortality Weekly Report
Early Release / Vol. 70
March 10, 2021

Racial and Ethnic Disparities in COVID-19 Incidence by Age, Sex, and Period Among Persons Aged <25 Years — 16 U.S. Jurisdictions, January 1–December 31, 2020



- Based on 16 jurisdictions with >70% completeness of race and ethnicity information
- only 30% of U.S. population

We need more testing!

Yes, but the right kind of testing, not just “more”

- Diagnosis and treatment of individual cases
  - public health case reports and contact tracing
  - requires sensitive and specific rtPCR viral test
    - with rapid turnaround
- Surveillance of infection in the population
  - viral, antigen, or antibody tests
  - representative samples of whole population
    - or subsets: people giving blood, nursing home residents
- Screening individuals in high-risk situations
  - appropriate when physical distancing not possible
    - limit asymptomatic transmission by removing infected
  - requires frequent (2x week), inexpensive, antigen testing
- Alternative testing sites

We need more testing!

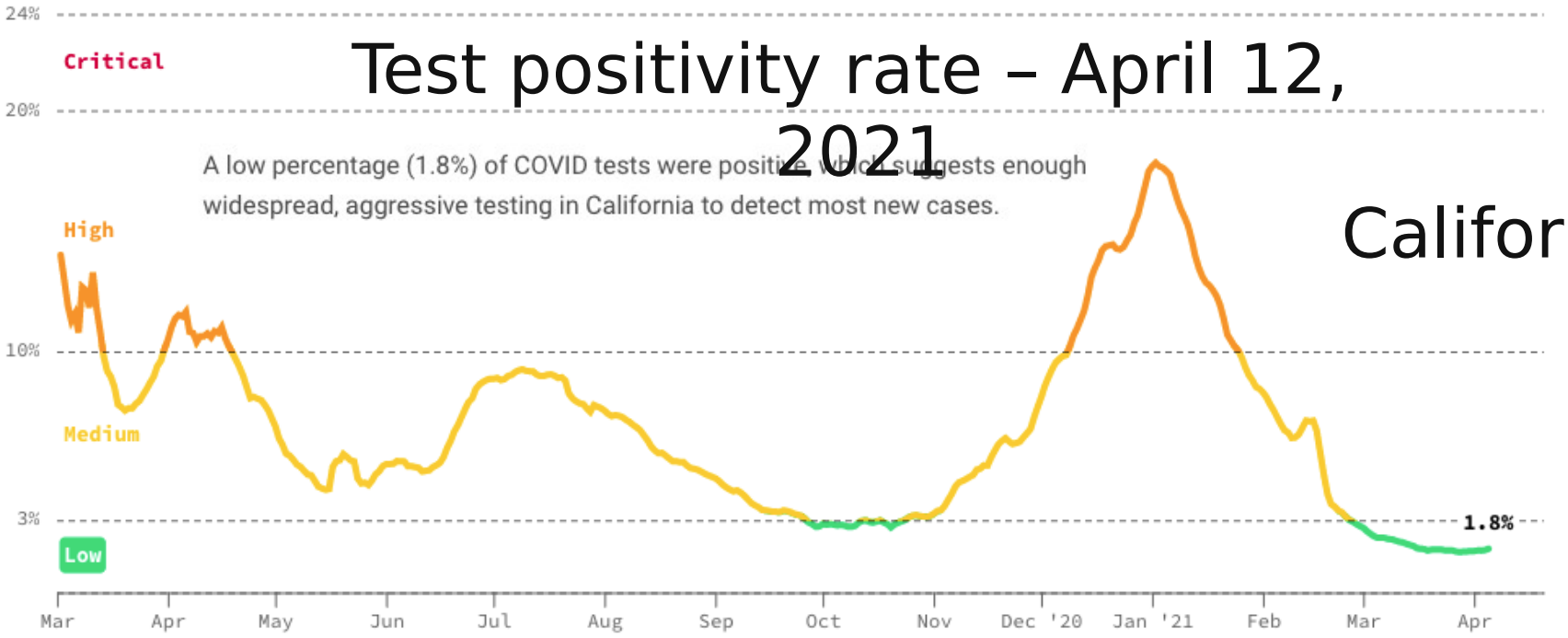
Yes, but the right kind of testing, not just "more"

- Diagnosis and treatment of individual cases
- Surveillance of infection in the population
- Screening individuals in high-risk situations
- Alternative testing sites
  - Front-line workers concerned about possible exposures
  - Test required for work
  - Individuals visiting vulnerable relatives
  - Release from travel-related quarantine requirements
  - Immunity passports
  - Peace of mind
- Each situation has a different pre-test probability
  - which effects interpretation of positive & negative results
- Different purposes require different kinds of tests

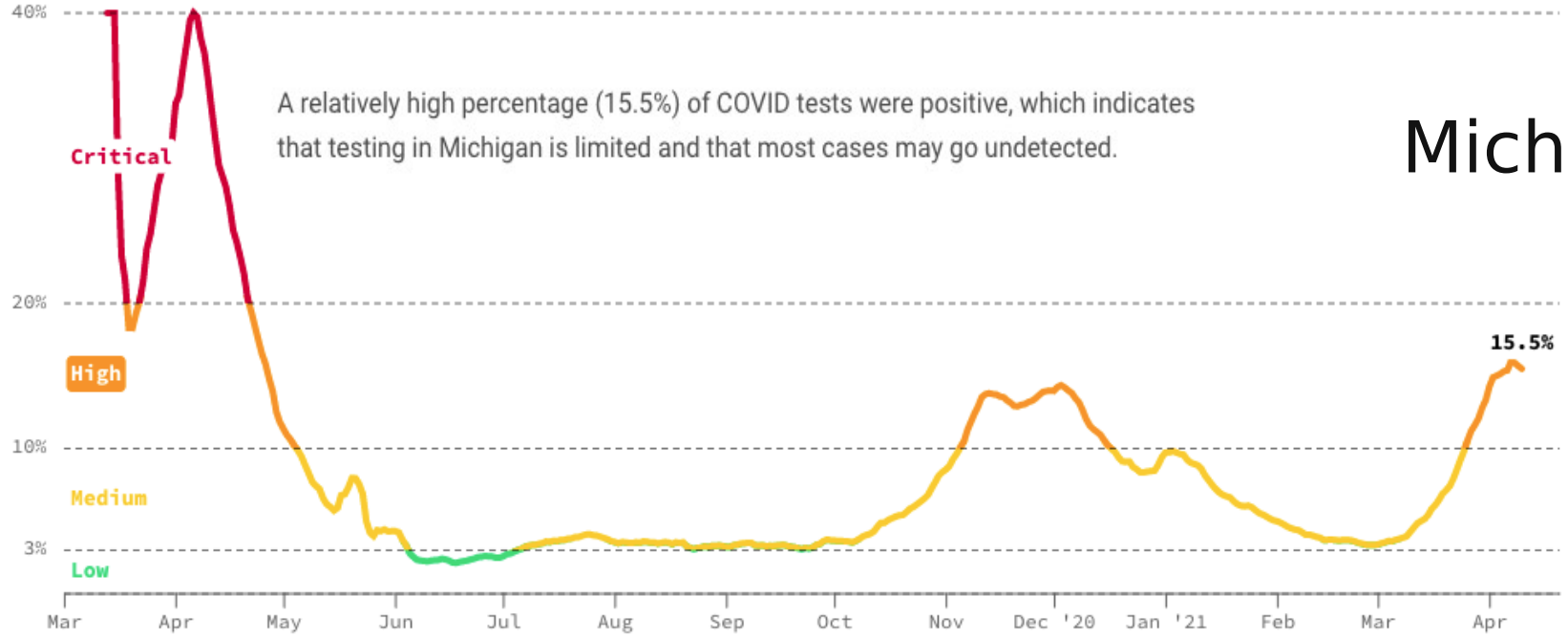
Test positivity rate

- Introduced as an *ad hoc* solution to the problem that cases were being missed
  - testing capacity was limited, so likely cases were prioritized for testing
- < 5% target adopted from another use, i.e. whether a wide enough net was being thrown in contact tracing
- Were some cases intentionally not tested?

Test positivity rate – April 12, 2021



California



Michigan

Test positivity rate

- Through the summer and fall, testing options expanded, increasing the number of people tested (denominator)
  - population tested changed
    - from mostly people with symptoms or close contact
    - to include back-to-work testing, travelers seeking to avoid quarantine, people living with at-risk relatives & worried well
  - universities, schools, workplaces began frequent screening
  - test types (serum/antibodies, rapid antigen) expanded
- $\frac{+}{+}$ positivity rate is *not* a desirable metric
  - both numerator and denominator change in ways that don't reflect transmission of COVID-19 in the population

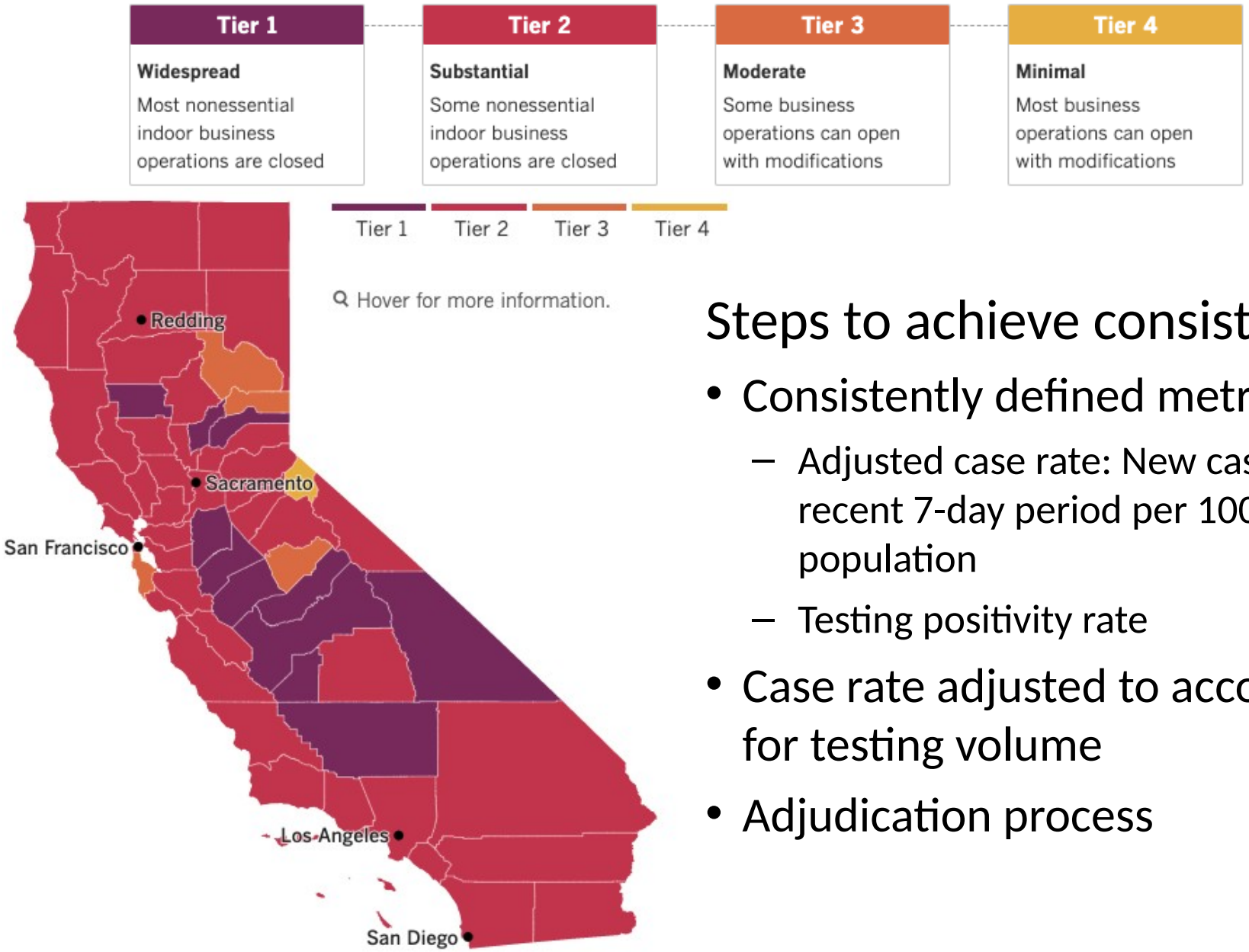
Principles for using metrics

- Metrics neither right or wrong, but imperfect "indicators" of the epidemiologic situation
  - often point to the need for more in-depth analysis
- Avoid
  - focus on daily changes, e.g. "largest daily number of deaths since May"
  - hard cut-offs, e.g. "schools can't reopen until positivity rate is less than 3%"
- A balanced portfolio of metrics can help diagnose the epidemiologic situation, but is not a substitute for detailed epi studies

Principles for using metrics

- Consistency is more important than "accuracy" (metrics only indictors) comparing
  - states & counties: LEVEL OF INFECTIONS
  - over time: RATE OF INCREASE
- Lack of consistent definitions, processes, etc. provides an opportunity to pick and choose among the options based on political aims
- Seek a constant reference population
  - should not reflect changes in test availability, public perceptions, etc.
- Standardize definitions and processes

California Blueprint



Steps to achieve consistency

- Consistently defined metrics
  - Adjusted case rate: New cases in recent 7-day period per 100k population
  - Testing positivity rate
- Case rate adjusted to account for testing volume
- Adjudication process

Recommendations

- Starting with current reported cases and deaths, CDC should standardize
  - case definitions (as already done)
  - measurement processes, e.g.
    - how to handle cases tested in one jurisdiction who live in another
    - electronic death registration systems, dropdown menus, etc.
  - metric definitions
    - including which tests to count (PCR, antibody, screening programs; pooled tests; first test only?; …)
  - time periods for averaging
  - and so on
- We're fortunate that Hopkins and other universities, NY Times and other publications, and private groups are publishing COVID-19 data and metrics
  - but this should really be CDC's responsibility

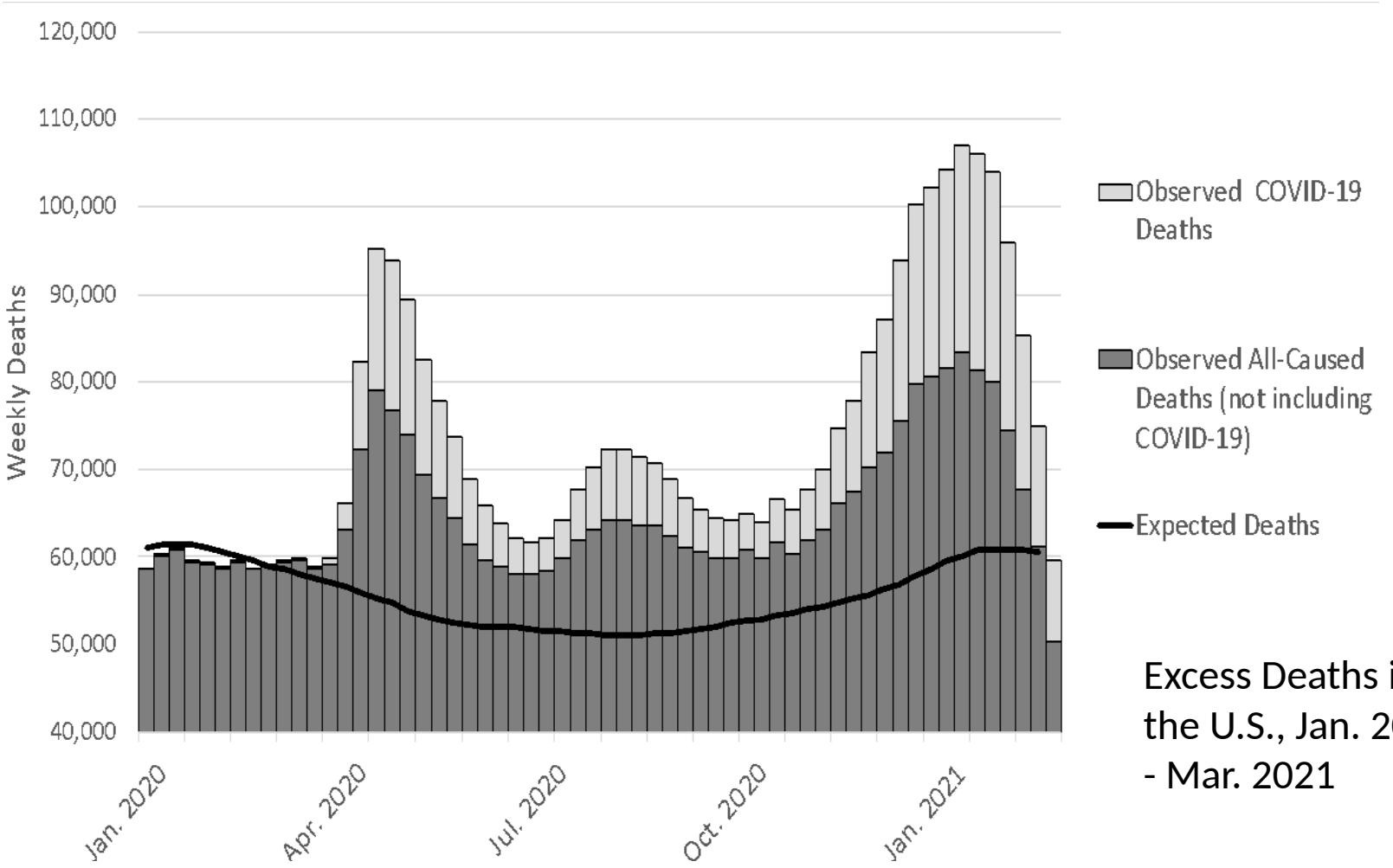
Week 2 - COVID-19 surveillance and metrics: Data for Decisions

- Introduction to infectious disease surveillance
- COVID-19 metrics
  - what do we need to know?
  - problems with current metrics
  - ***science-based alternatives – estimates vs. counts***

Beyond the current metrics

- Early in outbreak we go with the best we have
  - as the pandemic presses on, we must do better
  - trying to *count* all cases is not necessarily best
- Going forward, a new NASEM report suggests three statistical *estimation* methods to complement counts
  - excess mortality methods
  - syndromic surveillance
  - surveys based on representative samples
- Shift focus
  - from tracking day-to-day changes
  - to long-term trends and patterns & better understanding

- Excess Mortality = actual deaths - predicted deaths
  - includes deaths
    - directly caused by COVID-19 infection (whether attributed or not)
    - indirectly caused by COVID-19

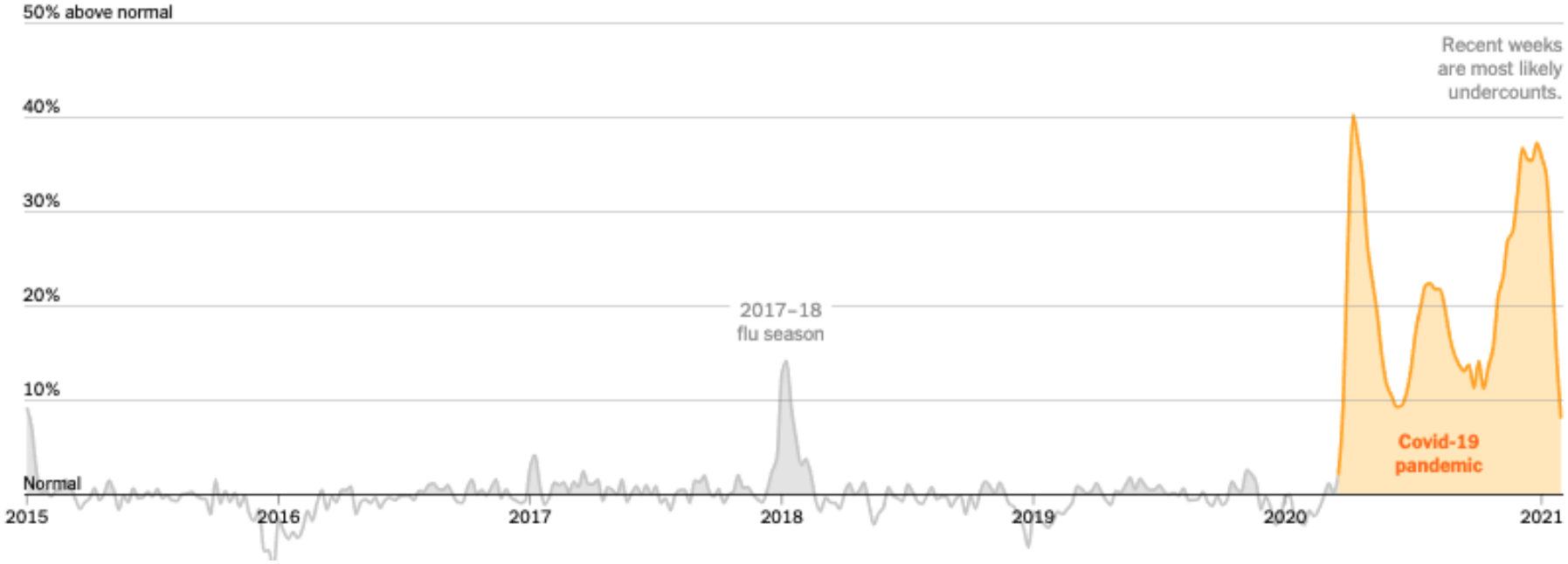


Excess Deaths in the U.S., Jan. 2020 - Mar. 2021

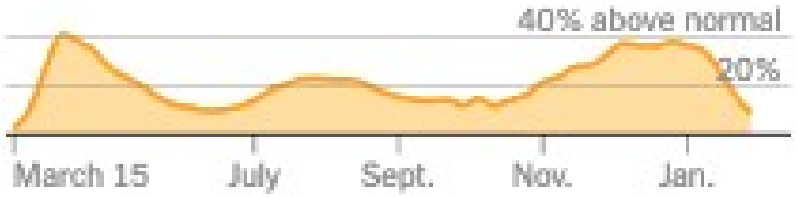
544,000 More U.S. Deaths Than Normal Since Covid-19 Struck

By Josh Katz, Denise Lu and Margot Sanger-Katz Updated March 3, 2021

Weekly deaths above and below normal in the U.S. since 2015



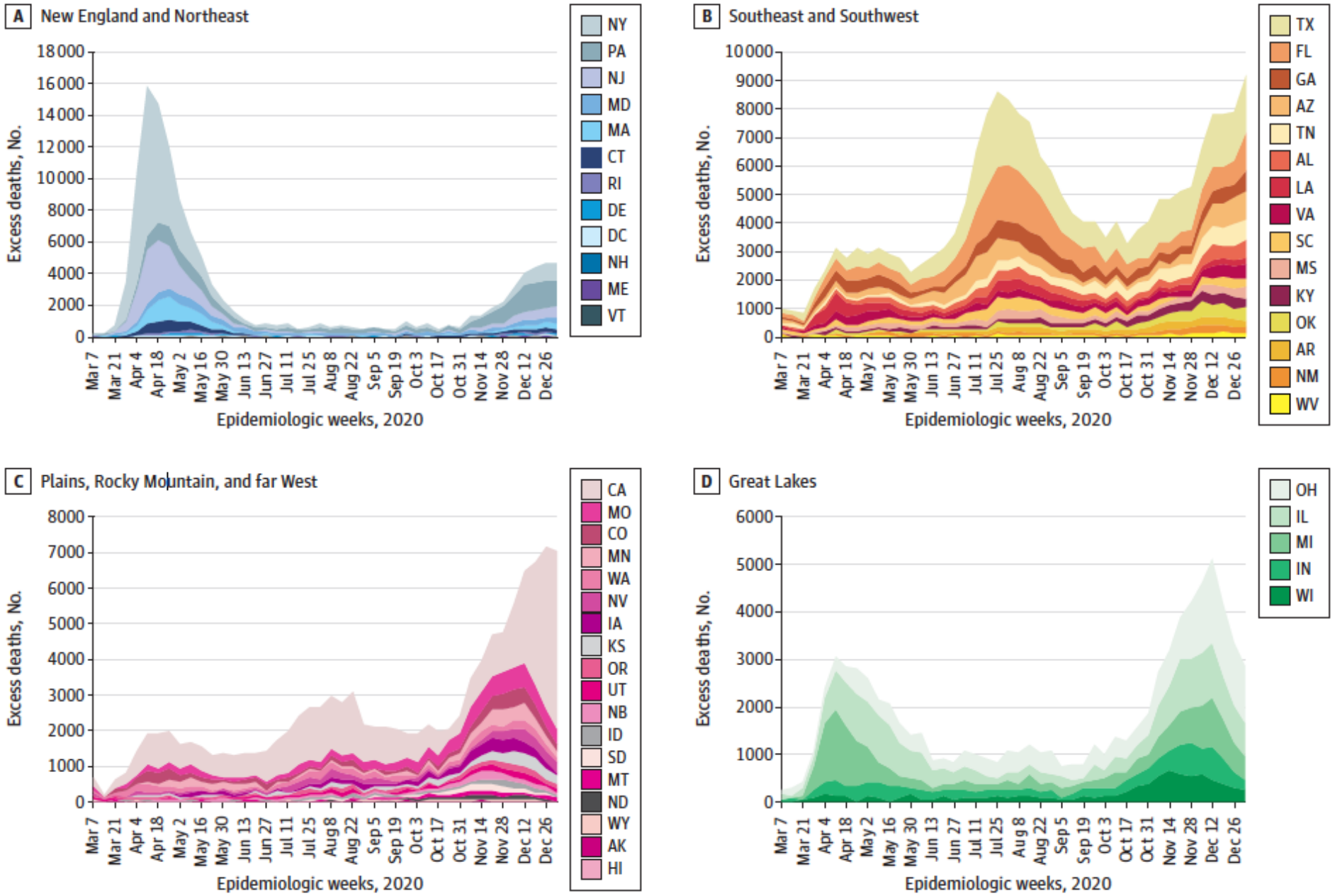
	REPORTED COVID-19 DEATHS	TOTAL EXCESS DEATHS	TOTAL ABOVE NORMAL
United States March 15 – Feb. 6	461,992	544,700	21%



The New York Times

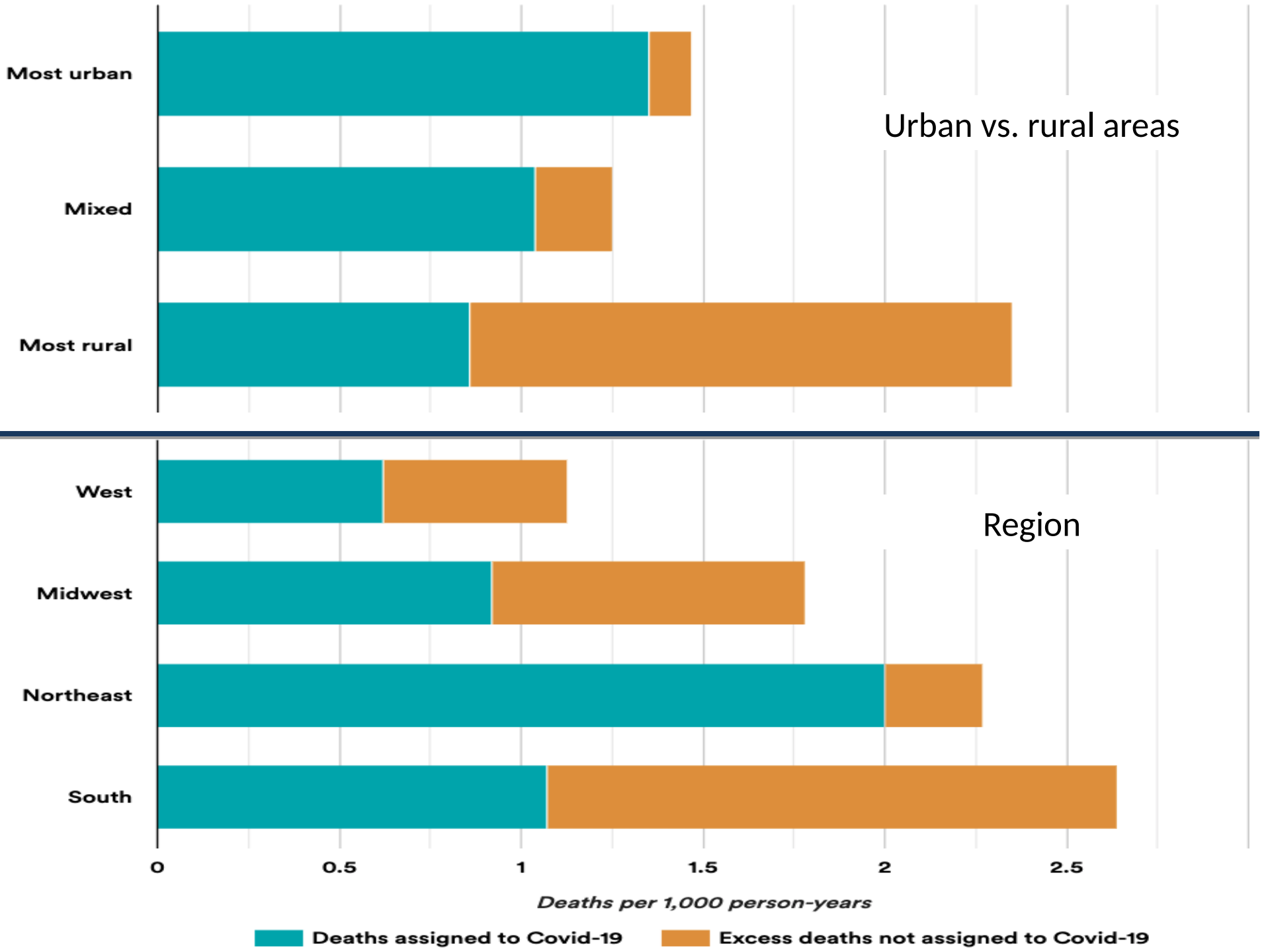
Woolf et al., JAMA, Apr. 2, 2021

Figure. Excess Deaths by Regions, March 1, 2020, to January 2, 2021



Excess mortality differentials

- Age
  - U.S. March-July, 2020 - 25-44 age group
  - 12,000 excess deaths (small fraction of 225,000 in all ages)
  - only 38% attributed to COVID-19
  - relative increase of 26.5% greater than any other age
- Race and Ethnicity
  - relative excess mortality (11.9% for Whites)
    - vs. Latinxs 53.6%, Blacks 32.9%, Asians 36.6%
  - If these groups died at the same rate as Asians or Whites
    - 19,500 Black, 8,400 Latinx & 600 Indigenous people would still be alive
    - attributable to (consistent with existing inequities)
      - more likely to have "essential" jobs
      - more comorbidities


Most urban
Mixed
Most rural
Urban vs. rural areas
West
Midwest
Northeast
South
Region
0
0.5
1
1.5
2
2.5
Deaths per 1,000 person-years
Deaths assigned to Covid-19
Excess deaths not assigned to Covid-19

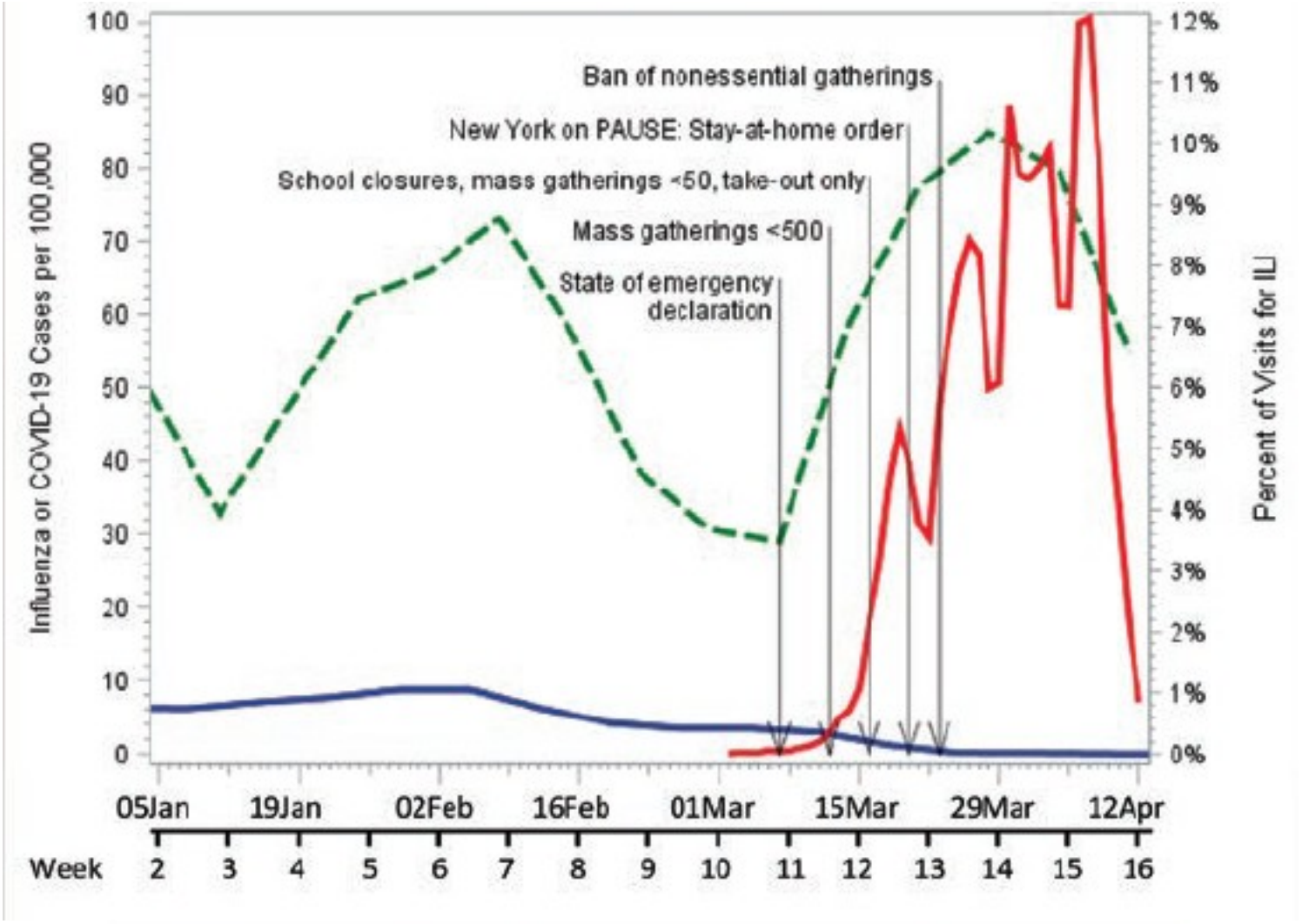
Excess mortality methods

- Compare deaths to similar period in the past
  - can look at cause of death, demographic and socioeconomic characteristics, etc.
  - research still needed, e.g. on
    - how to estimate expected deaths
      - e.g. the proper base period with which to compare
    - how to deal with people who moved because of the pandemic

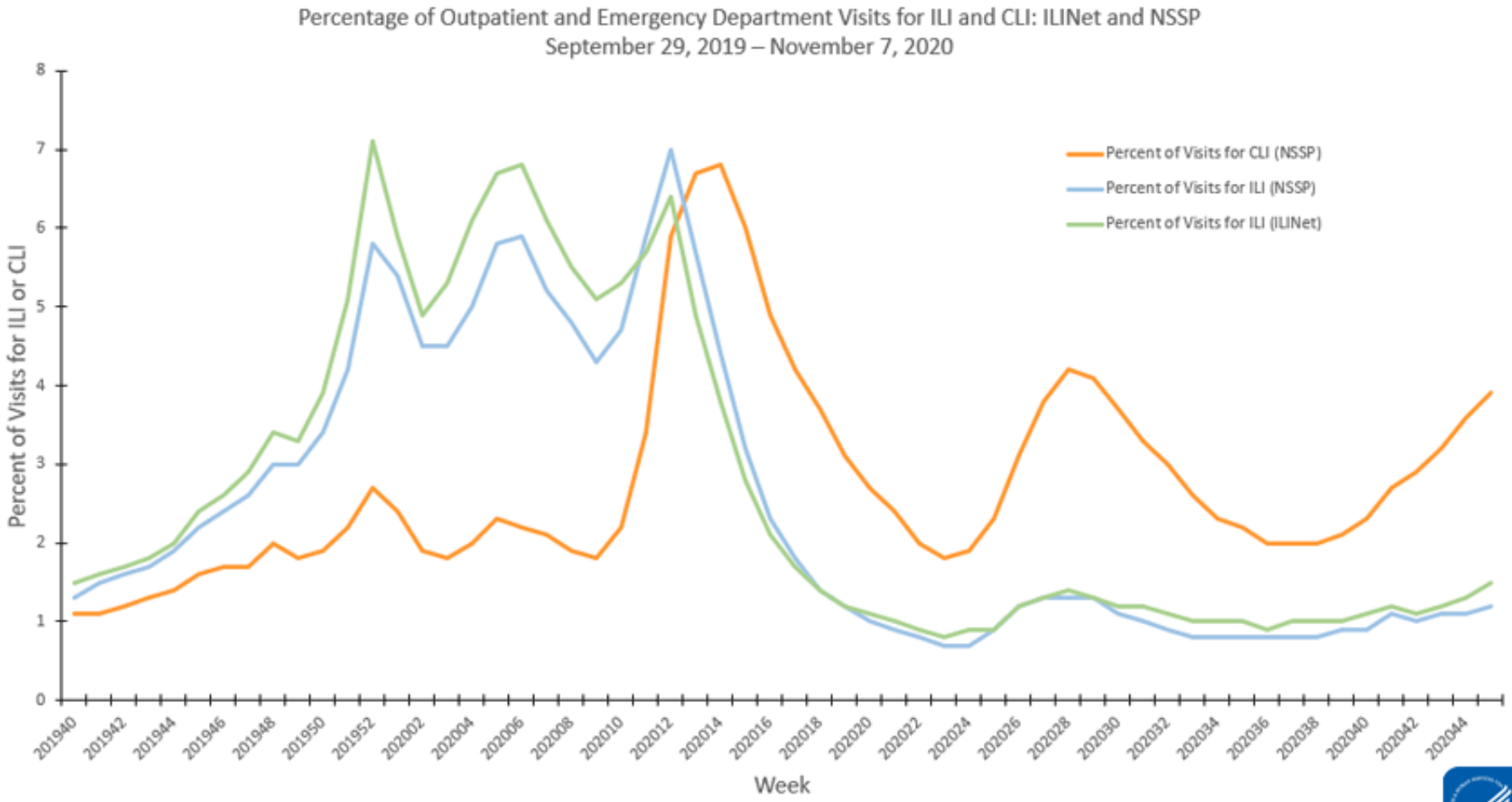
Syndromic surveillance

- Don’t wait for a formal diagnosis and case reporting processes, but rather track existing data that might indicate when people are having symptoms consistent with COVID-19 (“COVID-19-like illness”)
  - builds on an approach health officials have been using for years for influenza-like-illness (ILI)
  - based on
    - hospital ED visits (NSSP)
    - outpatient visits (ILINet)

New York City Metro Region, Jan. 1–April 12, 2020, Rosenberg *et al.*, 2020



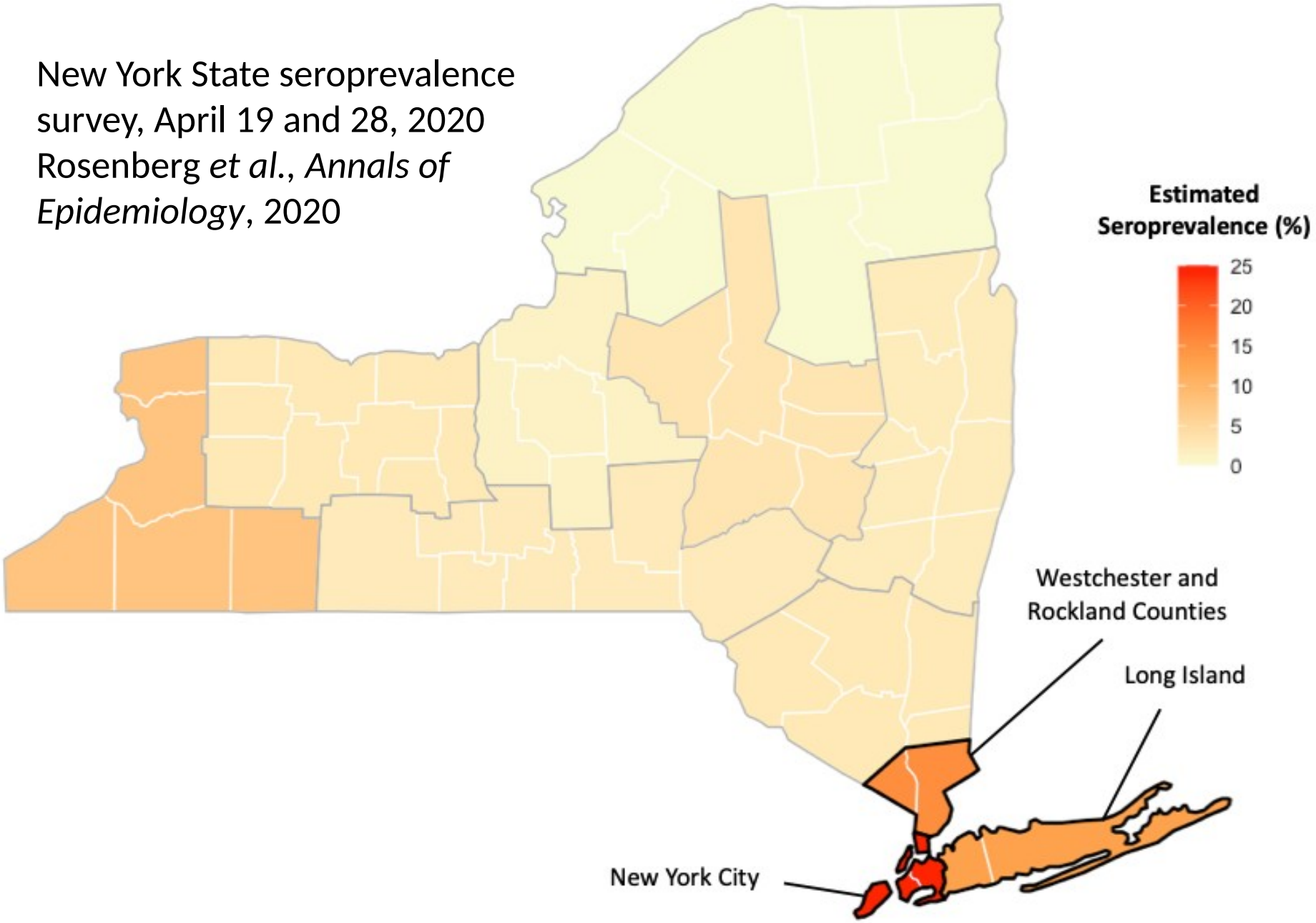
Percentage of outpatient and Emergency Department Visits for influenza-like illness (ILI) and COVID-19-like illness (CLI), U.S. September, 29, 2019 – November 7, 2020



Surveys based on representative samples

- Don’t need to count every case, or be sure that every case is “valid”
  - but do want a consistent reference population (denominator)
  - can sometimes adjust to be more representative
- Seroprevalence surveys
  - population-based
  - blood donations
  - clinic-based (dialysis, OB-GYN)

New York State seroprevalence survey, April 19 and 28, 2020
Rosenberg *et al.*, *Annals of Epidemiology*, 2020



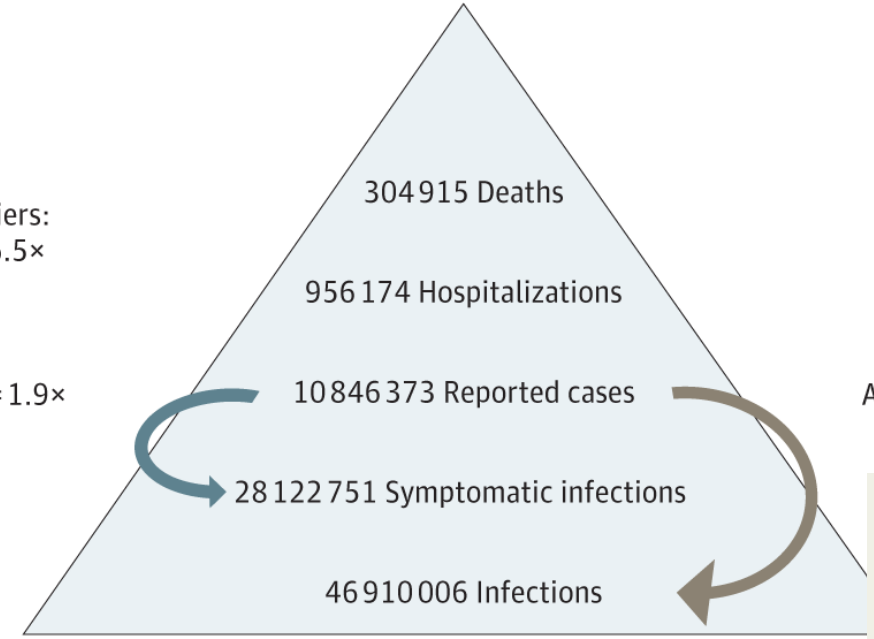
JAMA Network | Open

Frederick J. Angulo, Lyn Finelli, & David Swerdlow, January 5, 2021

Original Investigation | Infectious Diseases

Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys

Symptomatic underreporting multipliers:
January 21-April 30 = 6.5×
May 1-31 = 2.7×
June 1-30 = 3.2×
July 1-31 = 2.4×
August 1-November 15 = 1.9×



Infection underreporting multipliers:
January 21-April 30 = 10.8×
May 1-31 = 4.5×
June 1-30 = 5.4×
July 1-31 = 3.9×
August 1-November 15 = 3.2×

Proportion of population infected = 14.3%
Proportion of population with symptomatic infection = 8.6%
Infection hospitalization rate = 2.0%
Symptomatic fatality ratio = 1.1%

Meaning Findings of this study suggest that although more than 14% of the US population was infected with SARS-CoV-2 by mid-November, a substantial gap remains before herd immunity can be reached.

Mar. 5, 2021 estimate: 37%

Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study Anand *et al.*, *Lancet*, 2020

Seroprevalence substantially higher in Zip codes with

- Black & Hispanic populations
- high levels of poverty
- high population density

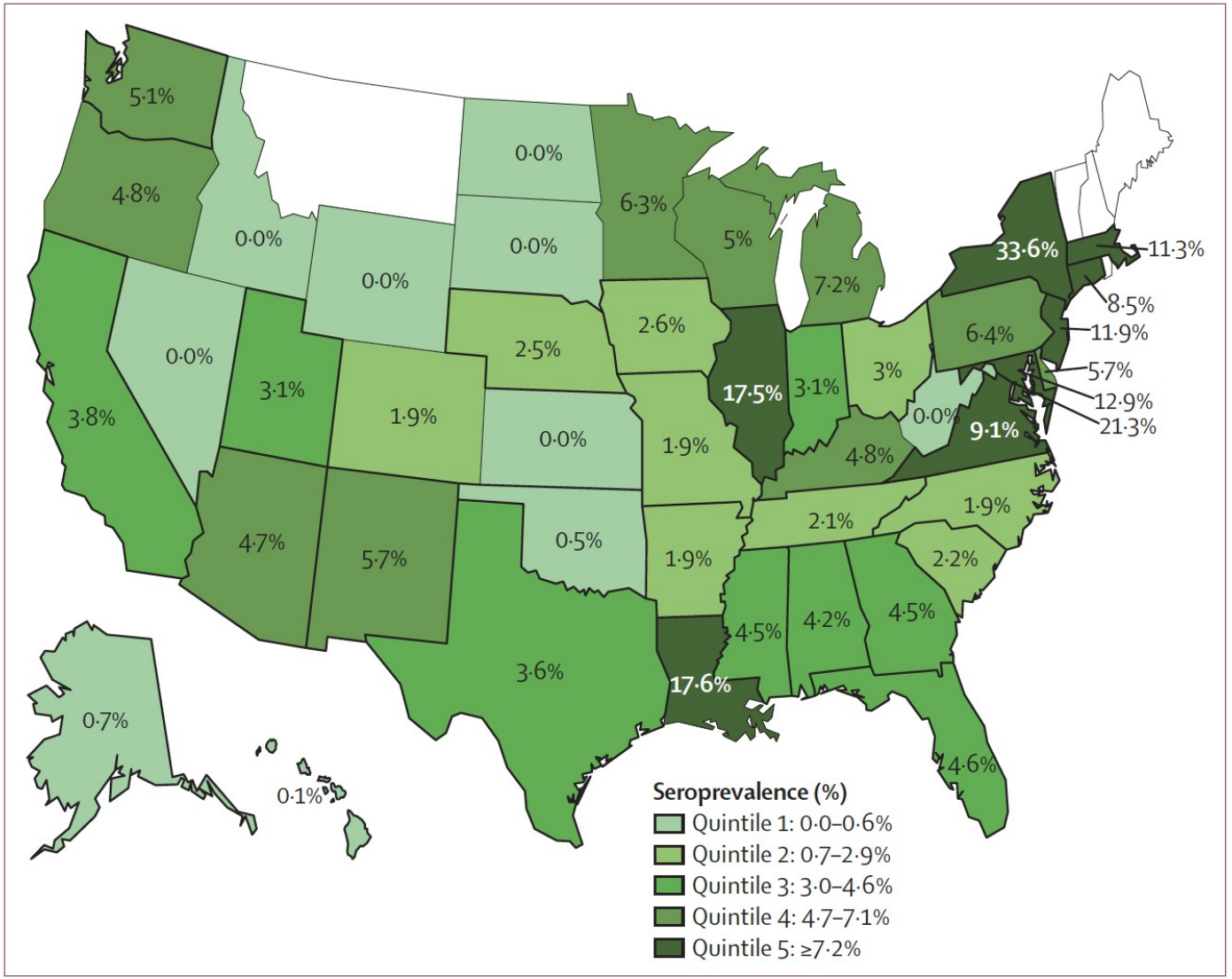


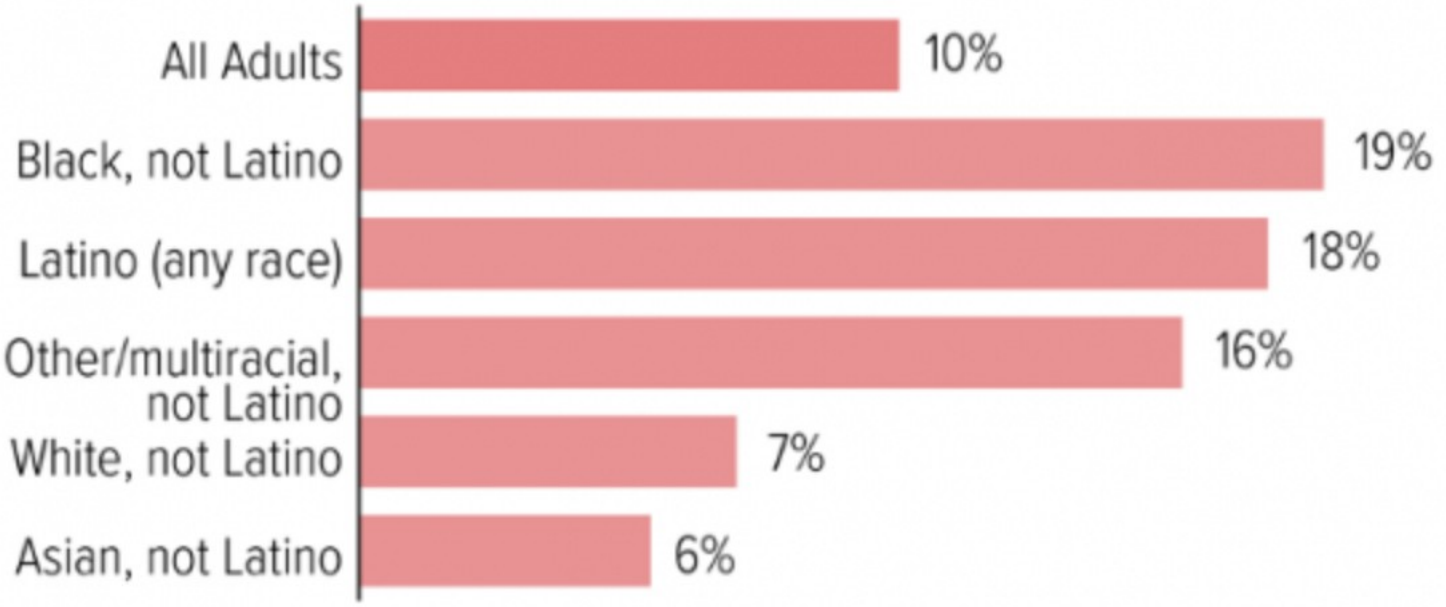
Figure 2: **Prevalence of SARS-CoV-2 antibodies in sampled population, by state**
Bolded borders represent states with more than 100 patients in the sample. The median number of patients sampled by state was 176 (IQR 83–536). States in white were not sampled. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.

Surveys based on representative samples

- Surveys can also be use to estimate mental health and social consequences of COVID-19
- Example: Census Bureau Pulse survey
  - designed to deploy quickly and efficiently, collecting data to measure household experiences during the coronavirus pandemic
  - Food insufficiency (share of households that sometimes or often did not have enough to eat in the last 7 days) concentrated in
    - the South and Southwest
    - Blacks and Latinos

Black and Latino Households Likelier to Experience Food Insufficiency During Pandemic

Share of adults saying that their household sometimes or often did not have enough to eat in the last 7 days, as of September 2-14, 2020



Note: Other/Multiracial not Latino = people identifying as American Indian, Alaska Native, Native Hawaiian or Pacific Islander, or more than one race. Percentages are based on reporting distributions and do not include the populations that did not respond to the question.

Source: CBPP analysis of Census Bureau Household Pulse Survey

Surveillance data *not* a substitute for epidemiological studies

(Lipsitch *et al*., *NEJM*, February 2020)

Types of Evidence Needed for Controlling an Epidemic.

Evidence Needed	Study Type
No. of cases, including milder ones	Syndromic surveillance plus targeted viral testing
Risk factors and timing of transmission	Household studies
Severity and attack rate	Community studies
Severity "pyramid"	Integration of multiple sources and data types
Risk factors for infection and severe outcomes, including death	Case–control studies
Infectiousness timing and intensity	Viral shedding studies

Conclusions

- Managing the COVID-19 pandemic
  - requires detailed, objective data on the level and rate of increase in new infections
- Better information starts with standardizing current case definitions, measurement processes, and metric definitions (i.e. good research methods)
- Research-based *estimation* methods can supplement and complement *counts* of cases and deaths
  - excess mortality
  - syndromic surveillance
  - surveys based on representative samples
- Still largely experimental, so research needed into best methods, standardization, etc.