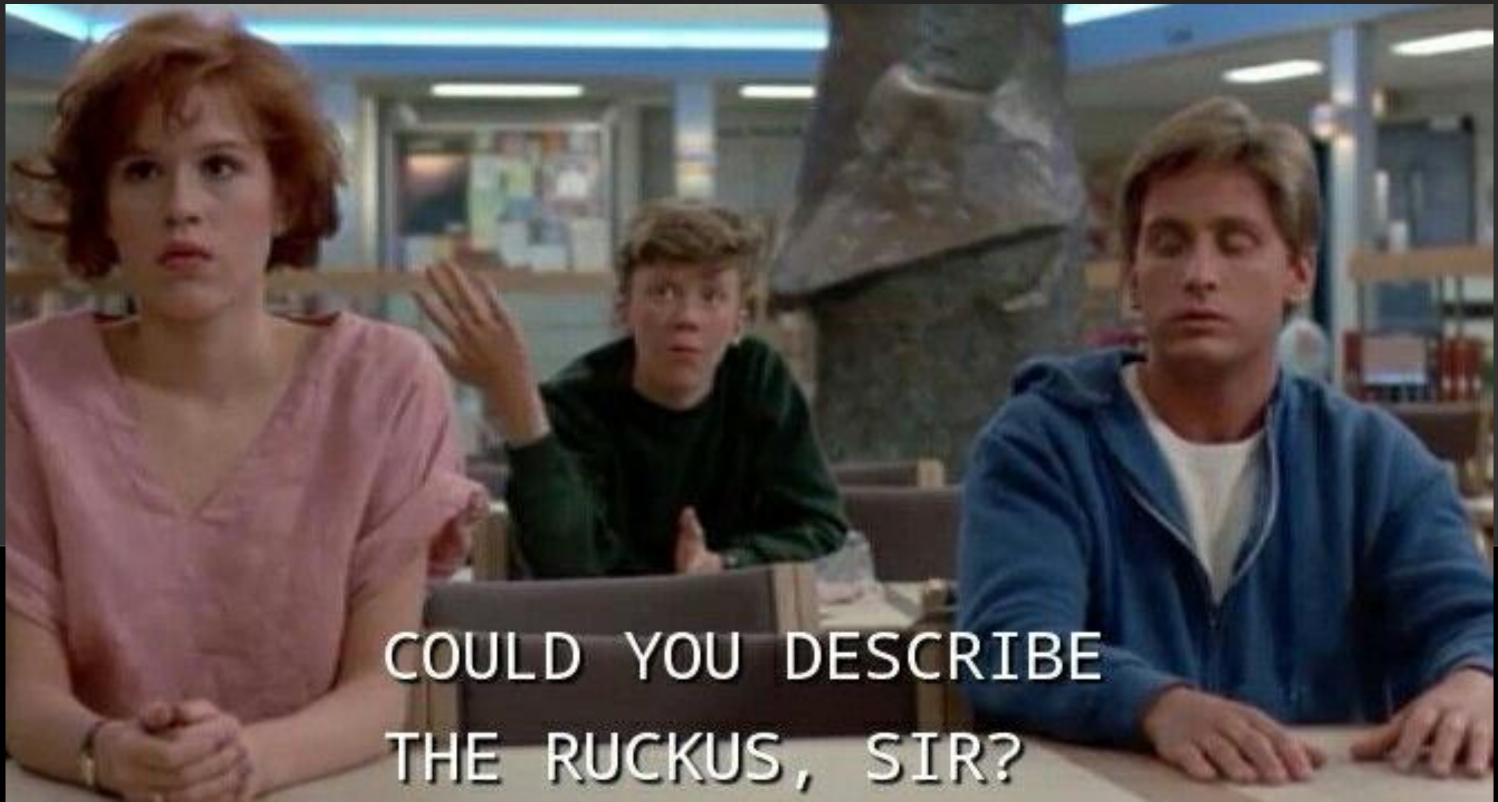
Reinvigorating Descriptive Epidemiology

Matthew Fox

mfox@bu.edu

@ProfMattFox mattpfox.bsky.social Free Associations, SERious Epi



https://academic.oup.com/aje/article/189/6/511/5811164

American Journal of Epidemiology
Vol. 189, No. 6
DOI: 10.1093/aje/kwaa030
Advance Access publication: March 24, 2020

Commentary

The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job

Catherine R

There has
We believe t
that are not
effects. Desc
briefly articul
estimation pr
creates ambi
questions inc
causal effect
guide epidem
fallacies, mini
frameworks t
such researc

American Journal of Epidemiology
Vol. 191, No. 7
https://doi.org/10.1093/aje/kwac056
Advance Access publication: March 22, 2022

Commentary

On the Need to Revitalize Descriptive Epidemiology

Matthew

Nearly
componen
last time w
impression
Descriptive
We believe
and ensuri
subsequen

American Journal of Epidemiology
Vol. 191, No. 12
https://doi.org/10.1093/aje/kwac115
Advance Access publication: July 1, 2022

Practice of Epidemiology

A Framework for Descriptive Epidemiology

Catherine R. Lesko*, Matthew P. Fox, and Jessie K. Edwards





[redacted]

epidemiology

epidemiology

what happens when you don't get into med school the first time around.

What is epidemiology?

- Epidemiology is study of the **distribution** and **determinants** of disease states in human populations and the application of that knowledge to the control of disease

Distribution

Determinants

Catherine R. Lesko*, Alexander P. Keil, and Jessie K. Edwards

public health (12). Loosely speaking, these research goals fall along a spectrum with purely descriptive epidemiology at 1 end; hypothesis generation, prediction, and outbreak investigation somewhere in the middle; and causal effect estimation and program evaluation at the other end. Here, we https://journals.lww.com/epidem/abstract/2023/07000/complexity_in_epidemiology_and_public_health_.7.aspx

Complexity in Epidemiology and Public Health. Addressing Complex Health Problems Through a Mix of Epidemiologic Methods and Data

Naja Hulvej Rod,[a,b] *Alex Broadbent,*[c,d] *Morten Hulvej Rod,*[b,e,f] *Federica Russo,*[b,g,h] *Onyebuchi A. Arah*[i,j] *and Karien Stronks,*[b,k]

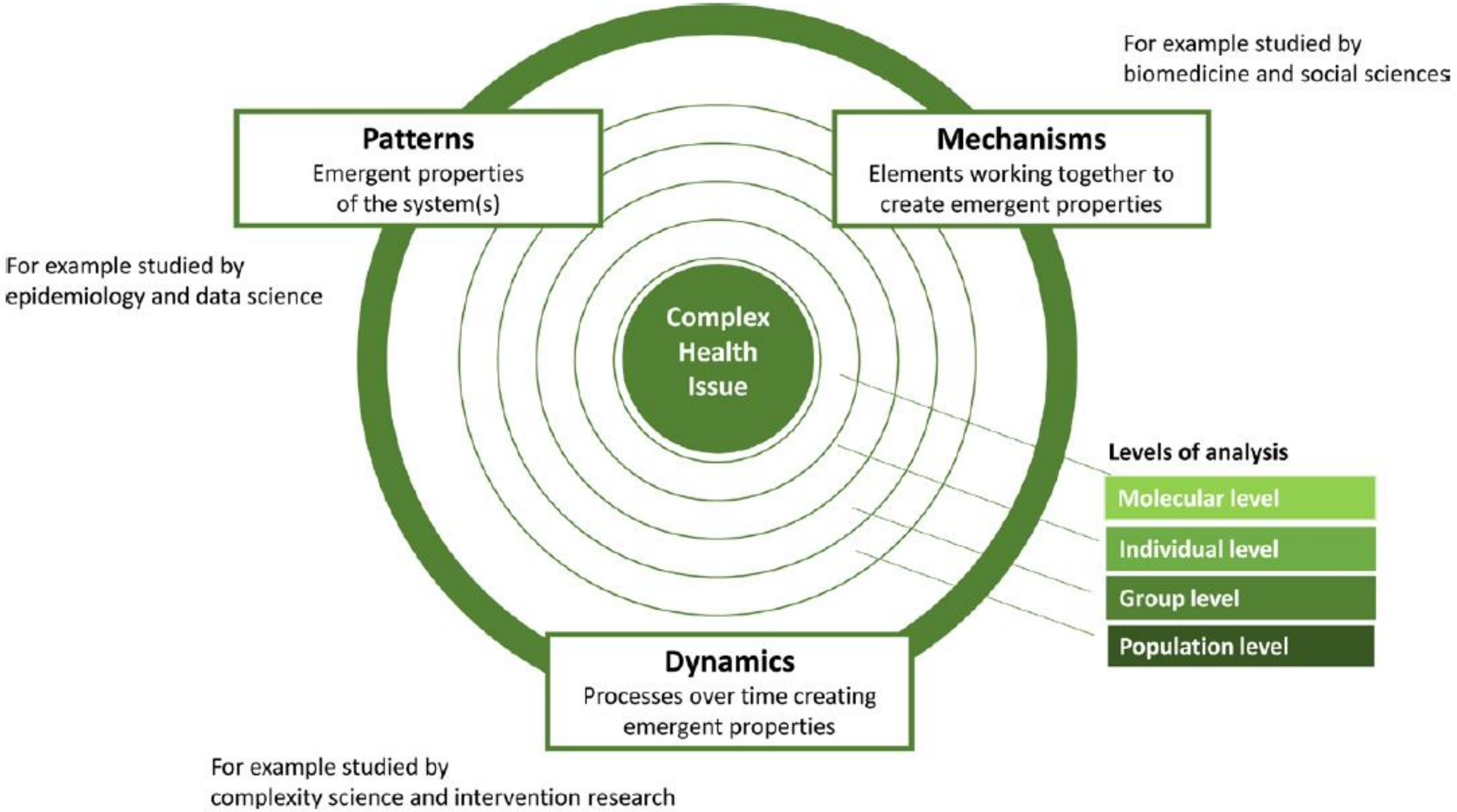


FIGURE 2. Overview of the interdisciplinary Health Complexity Framework for producing knowledge on complex health issues.

HEALTH COMPLEXITY FRAMEWORK

We suggest that a first, but important, step towards understanding and intervening in complex public health phenomena is to systematically generate and integrate the knowledge of the system(s) that give rise to these phenomena. To operationalize this for public health, we propose an interdisciplinary framework that organizes this knowledge production according to three core dimensions, capturing seven critical features of complex systems (Figure 2). The framework builds upon the idea of methodologic pluralism,[18,27,28] and is intended as an overarching framework for interdisciplinary and collaborative research.

The three dimensions involve:

1. patterns: describing the health patterns that emerge from complex systems;
2. mechanisms: understanding the mechanisms that produce these emergent patterns; and
3. dynamics: exploring the dynamics that make mechanisms and patterns change over time.

Each of these dimensions is directly related to key features of complex systems, which are often highlighted in relation to public health.[21,22] These features include emergence, interactions, nonlinearity, interference, feedback loops, adaptation, and evolution.

https://www.cdc.gov/eis/field-epi-manual/chapters/Describing-Epi-Data.html

Describing Epidemiologic Data

Print

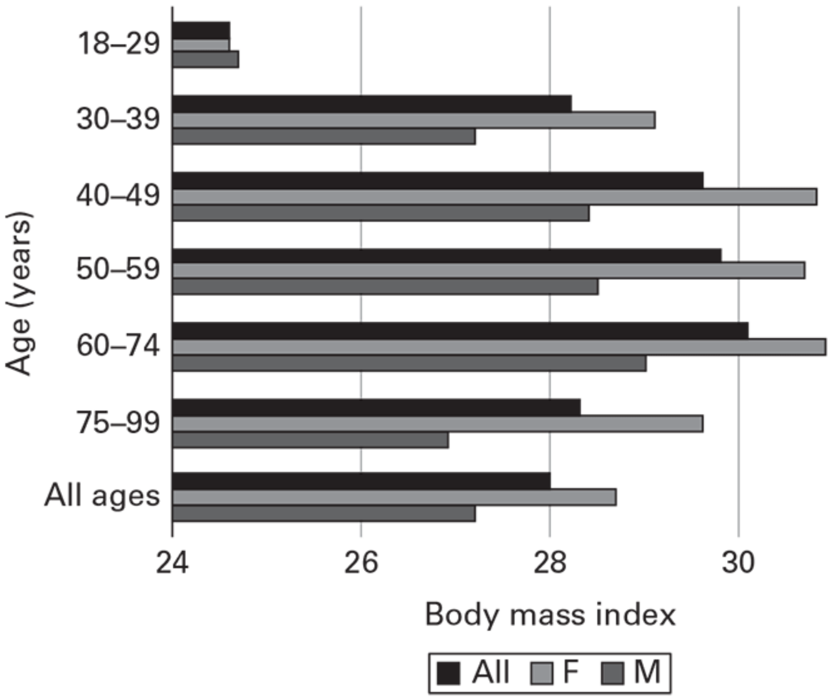
Robert E. Fontaine

"this task, called descriptive epidemiology, answers the following questions about disease, injury, or environmental hazard occurrence: What? How much? When? Where? Among whom?" (4, p. 106).

BU

Person

Mean BMI among adults by Age and Sex in Jordan 2012



Place

Time

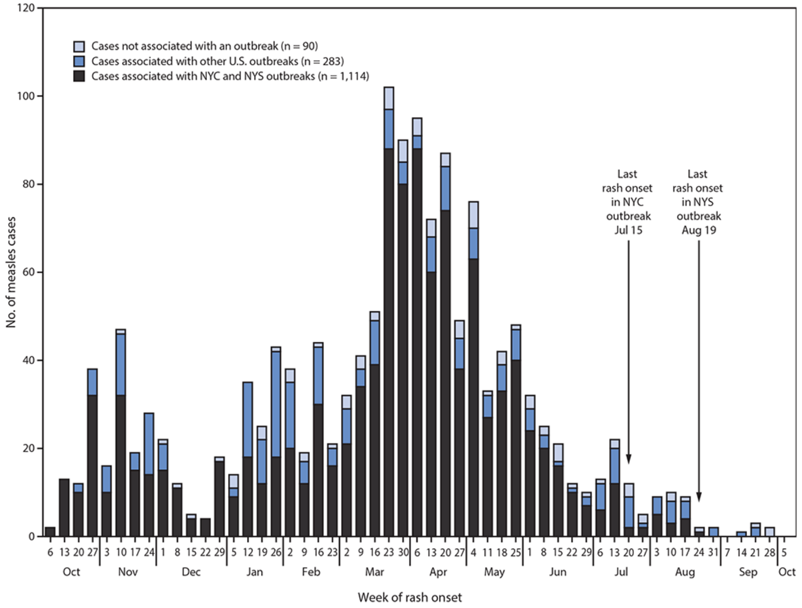
Morbidity and Mortality Weekly Report (*MMWR*)

National Update on Measles Cases and Outbreaks — United States, January 1–October 1, 2019

Weekly / October 11, 2019 / 68(40);893–896

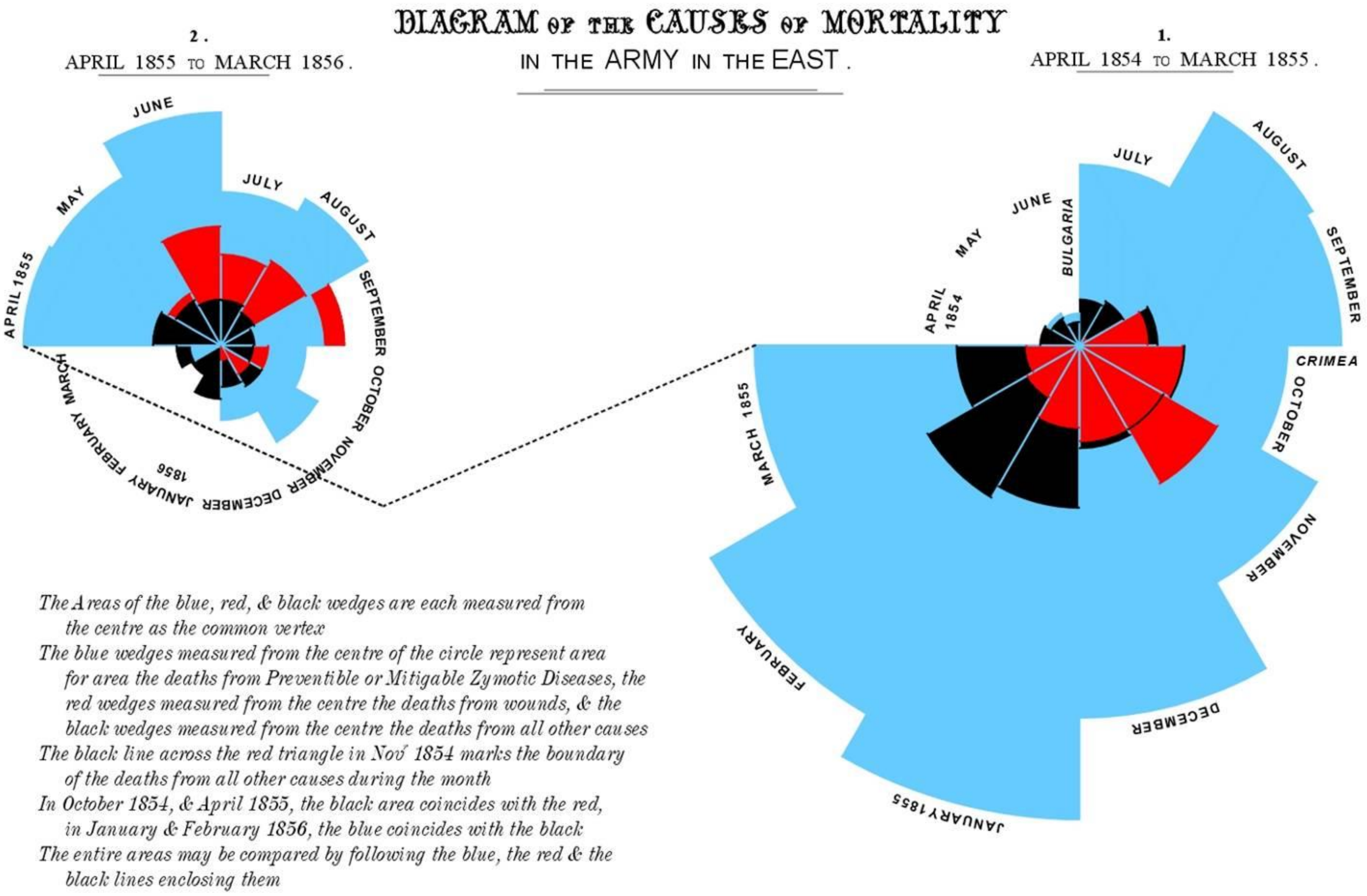
On October 4, 2019, this report was posted online as an MMWR *Early Release.*

Manisha Patel, MD[1]; Adria D. Lee, MSPH[1]; Nakia S. Clemmons, MPH[1]; Susan B. Redd[1]; Sarah Poser[1]; Debra Blog, MD[2]; Jane R. Zucker, MD[3,4]; Jessica Leung, MPH[1]; Ruth Link-Gelles, PhD[1]; Huong Pham, MPH[1]; Robert J. Arciuolo, MPH[3]; Elizabeth Rausch-Phung, MD[2]; Bettina Bankamp, PhD[1]; Paul A. Rota, PhD[1]; Cindy M. Weinbaum, MD[4]; Paul A. Gastañaduy, MD[1] (VIEW AUTHOR AFFILIATIONS)

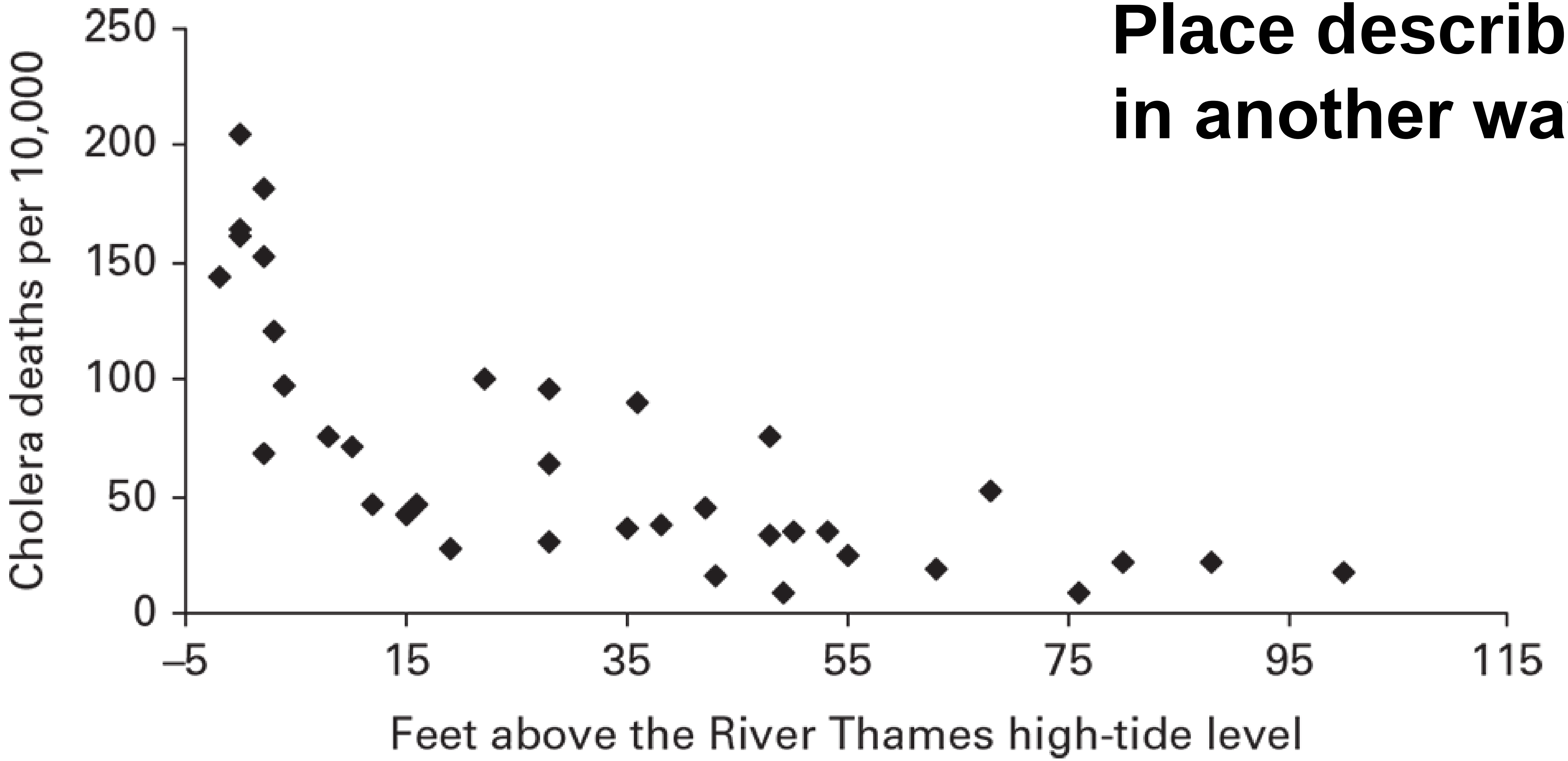


http://www.florence-nightingale-avenging-angel.co.uk/blog/wp-content/uploads/2012/01/Rose.jpg

Florence Nightingale’s Crimean War Rose Diagram



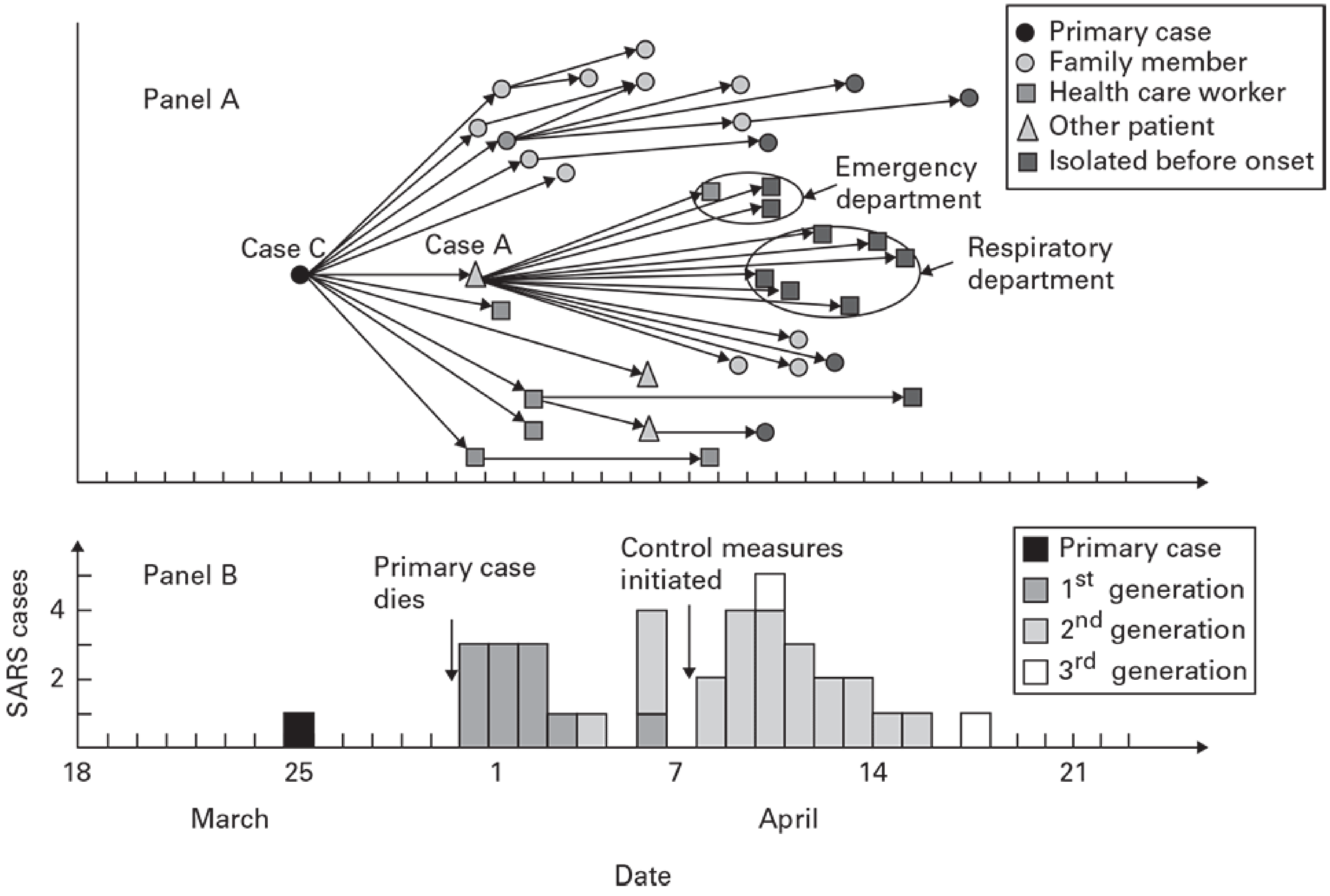
Place described in another way…



Cholera deaths per 10,000 and altitude above high tide level by district, London, 2849

Time

Contact between severe acute respiratory syndrome (SARS) cases among a group of relatives and health care workers: Beijing, China, 2003.



https://www.pnas.org/doi/pdf/10.1073/pnas.1510502113

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]

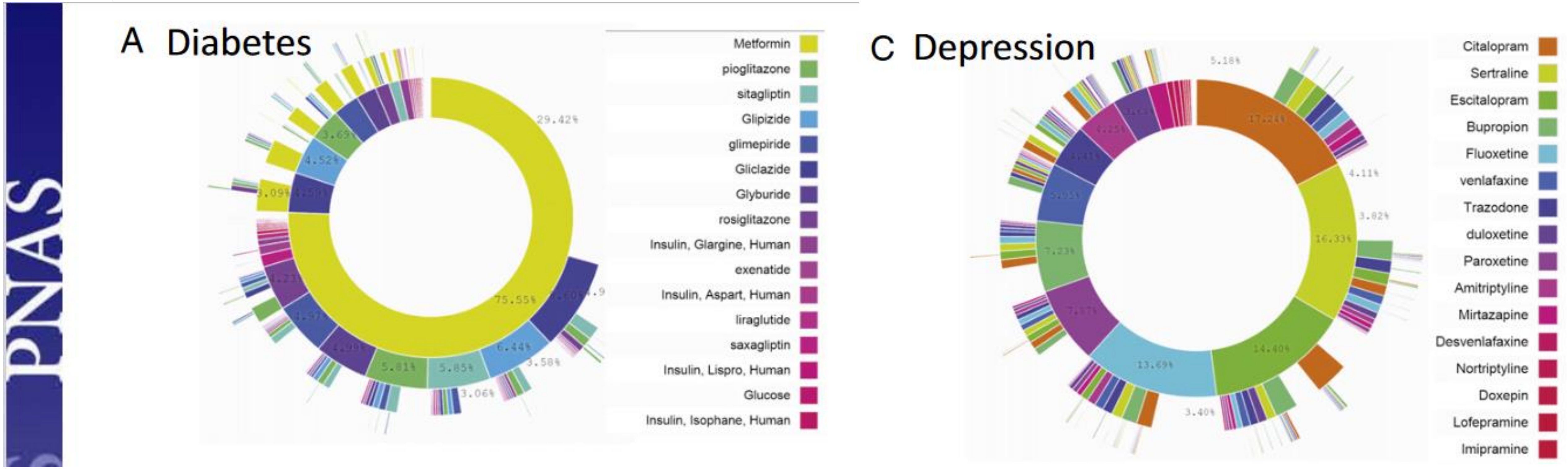


Fig. 2. Treatment pathways for all data sources. For each disease, diabetes (*A*), hypertension (*B*), and depression (*C*), and across all data sources, the inner circle shows the first relevant medication that the patient took, the second circle shows the second medication, and so forth. Only four levels are shown, but up to 20 medications were recorded. For example, 76% of diabetes patients started with metformin, and 29% took only metformin.

https://ehp.niehs.nih.gov/doi/10.1289/EHP12276

| Research Article | 30 August 2023

Geospatial Assessment of Racial/Ethnic Composition, Social Vulnerability, and Lead Water Service Lines in New York City

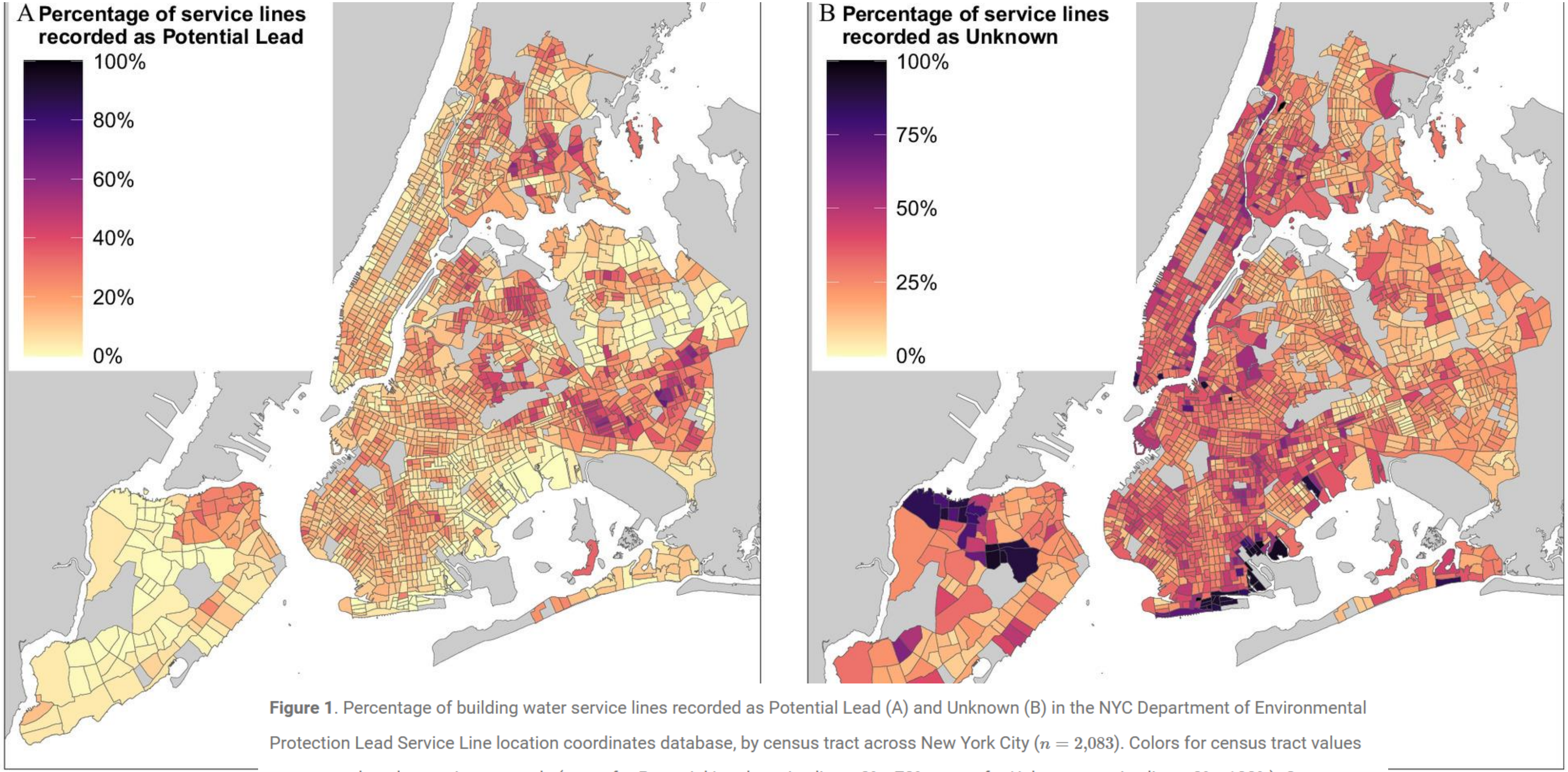


Figure 1. Percentage of building water service lines recorded as Potential Lead (A) and Unknown (B) in the NYC Department of Environmental Protection Lead Service Line location coordinates database, by census tract across New York City ($n = 2,083$). Colors for census tract values

https://www.tableau.com/blog/visualizing-covid-19-racial-disparities-to-create-path-toward-equity

Descriptive Epi for Identifying Disparities

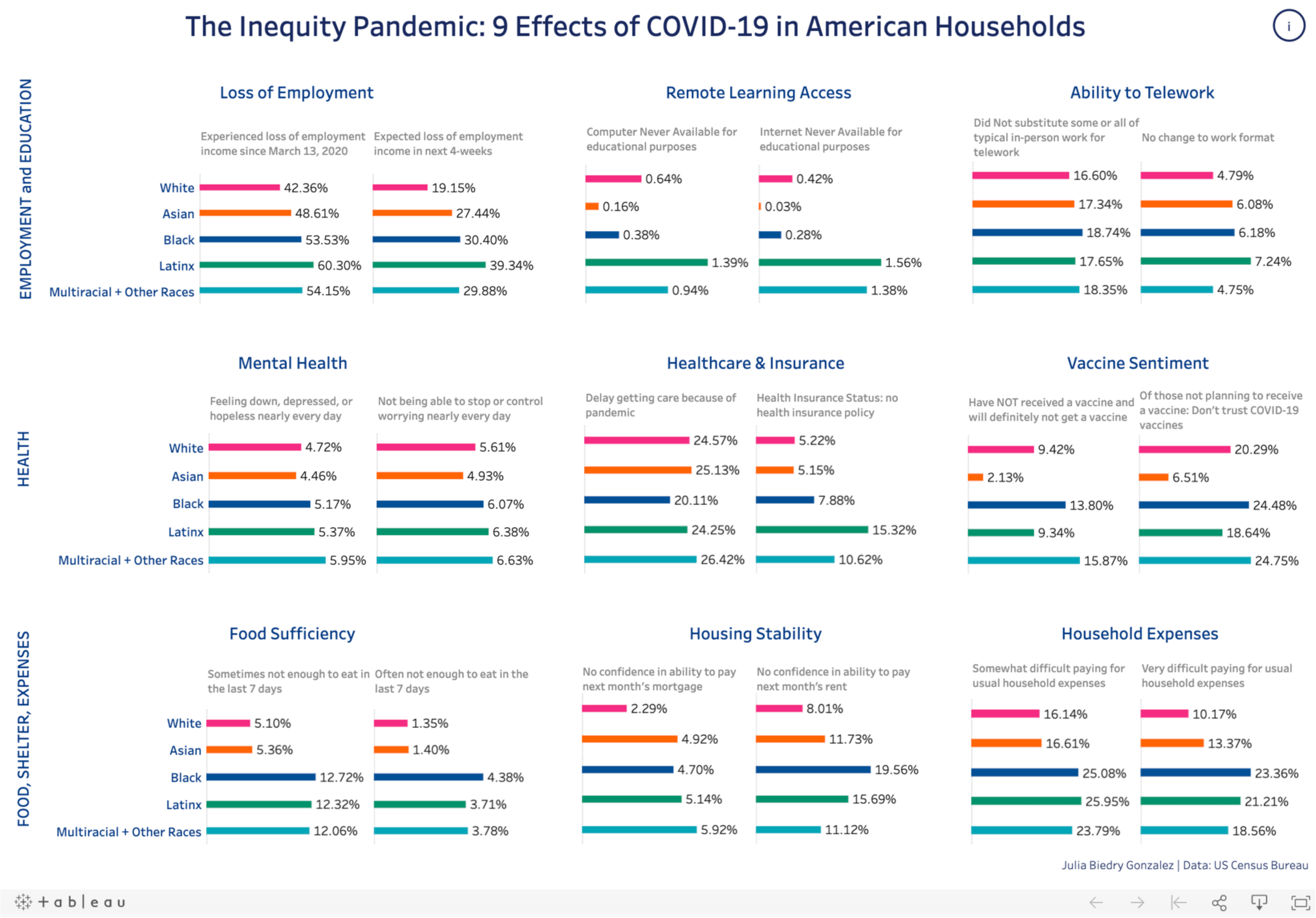


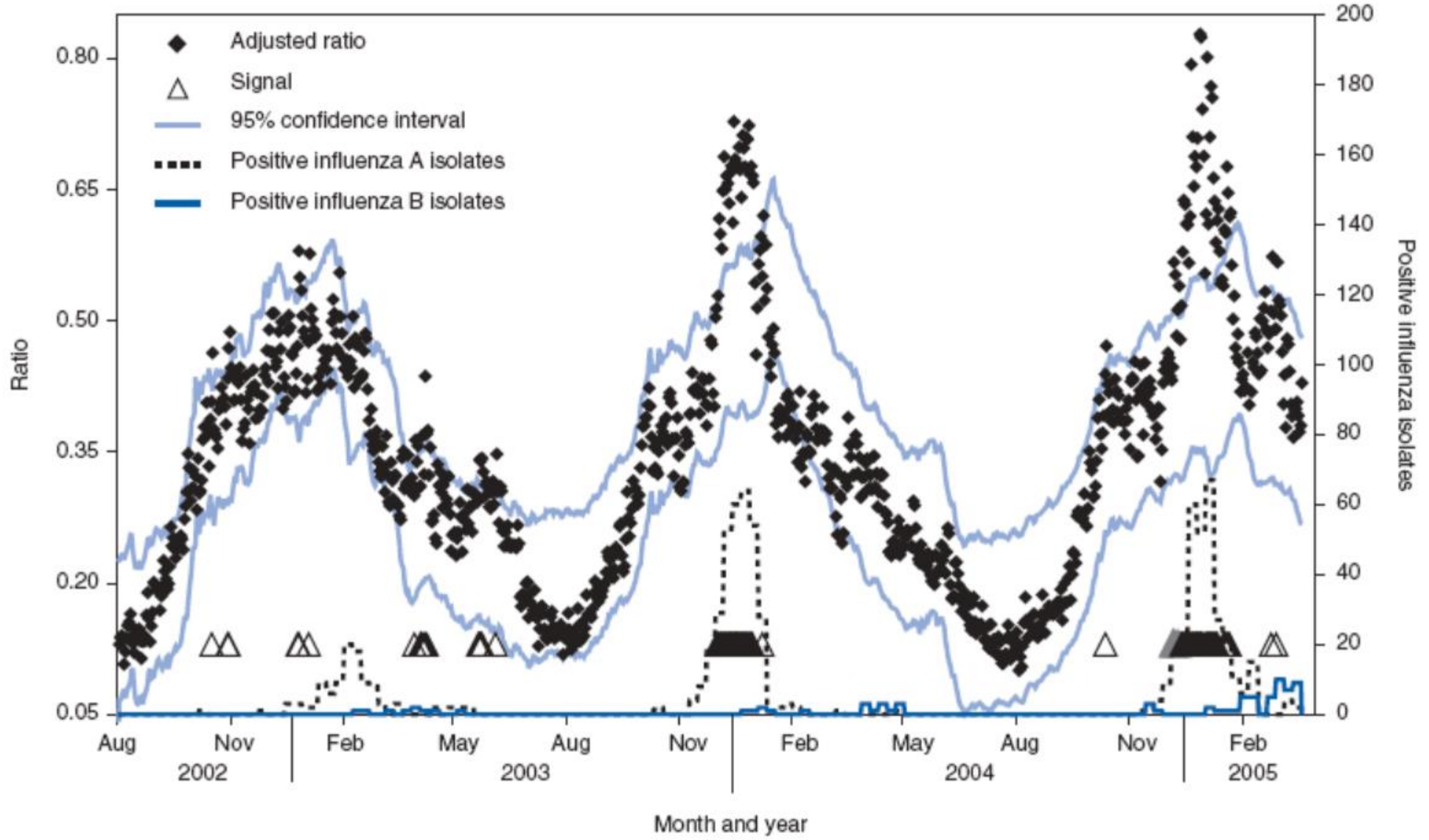
Table 2 Summary of emergency department syndromic surveillance systems (EDSyS) included in the review, by country/territory, with source and format of information used to define syndromic indicators and of areas of public health surveillance supported the EDSyS

Country/ territory	Syndromic indicator		Infectious diseases			Extreme weather		Other non-infectious		
	Source[a]	Format	Respiratory	Influenza	Gastrointestinal	Heat	Cold	Injury/trauma	alcohol	drug
Albania	diagnosis	coded	✓	✓	✓	–	–	–	–	–
Australia	diagnosis	coded	✓	✓	✓	✓	✓	✓	✓	✓
Canada	chief complaint	text	✓	✓	✓	✓	✓	✓	–	–
China	chief complaint	coded	✓	✓	–	–	–	–	–	–
France	diagnosis	coded	✓	✓	✓	✓	✓	✓	✓	–
Greece	chief complaint	pick list	✓	–	✓	–	–	–	–	–
Italy	chief complaint	text/coded	✓	✓	✓	–	–	–	–	–
Jamaica	"daily analysed data"		✓	–	✓	✓	–	–	–	–
Republic of Korea	diagnosis	coded	✓	✓	✓	–	–	–	–	–
New Zealand	diagnosis	coded	✓	✓	–	–	–	–	–	–
Singapore	unknown	coded	✓	✓	✓	–	–	–	–	–
Spain	chief complaint	coded	✓	✓	✓	–	–	–	–	–
Taiwan	chief complaint	text/ coded	✓	✓	✓	–	–	–	–	–
UK[b]	diagnosis	coded	✓	✓	✓	✓	✓	–	✓	–
USA	chief complaint	text	✓	✓	✓	✓	✓	✓	✓	✓

[a] EDSyS may collect more than one data item for syndromic indicators, but each reported a primary field used as standard
[b] UK: England & Northern Ireland ✓ relevant EDSyS indicators identified - no relevant EDSyS indicators identified

Descriptive Syndromic Surveillance

FIGURE 1. Citywide trends and signals in the adjusted ratio of influenza-like illness to analgesic over-the-counter (OTC) drug sales and positive isolates of influenza A and B — New York City, August 1, 2001–March 31, 2005*



*Temporal signals p<0.01 and 95% confidence intervals from daily linear regression. The OTC ratio is adjusted for day of week, major national winter holidays, and the day after these holidays (Thanksgiving, Christmas, New Years, and Martin Luther King observance day).

https://www.nejm.org/doi/full/10.1056/NEJMra2119215

Function	Examples
Early warning	• Natural-language processing of news sources to identify outbreaks (Freifeld et al., *JAMIA* 2008) • Unsupervised machine learning of social media data to detect unknown infections (Lim, Tucker, and Kumara, *J Biomed Inform* 2017)
Pathogen classification	• Convolutional neural network model for reading antibiograms (Pascucci et al., *Nat Commun* 2021) • Convolutional neural network model to automate malaria microscopy and diagnosis (Liang et al., *IEEE* 2016)
Risk assessment	• Reinforcement learning of Covid-19 positivity rates to target limited testing in Greece (Bastani et al., *Nature* 2021) • Machine-learning models including random forest and extreme gradient boosting to use syndromic surveillance for Covid-19 risk prediction (Dantas, *PLoS One* 2021)
Source identification	• Automated data mining of electronic medical records to uncover hidden routes of infection transmission (Sundermann et al., *Clin Infect Dis* 2021) • Supervised machine learning in combination with digital signal processing for genomic tracing of Covid-19 (Randhawa et al., *PLoS One* 2020)
Hotspot detection	• Neural computing engine to correlate sound from hospital waiting rooms with influenza spikes (Al Hossain et al., *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2020) • Multilayer perceptron artificial neural network model to detect spatial clustering of tuberculosis (Mollalo et al., *Int J Environ Res Public Health* 2019)
Tracking and forecasting	• Real-time stacking of multiple models to improve forecasts of seasonal influenza (Reich et al., *PLoS Comput Biol* 2019) • Machine learning to combine new data sources for monitoring Covid-19 (Liu et al., *J Med Internet Res* 2020)

Figure 1. Various Functions of Artificial Intelligence (AI) for Infectious-Disease Surveillance.

Shown is a nonexhaustive list of functions of AI-aided infectious-disease surveillance and representative examples from the published literature.[2-13] Each example includes the type of AI algorithm, a brief description of its purpose, and the associated citation. Covid-19 denotes coronavirus disease 2019.

Individual event	Example of signal-generating method	Algorithm category	Signal of possible infectious disease in a population	Surveillance output
Biosignals passively measured by smartwatch	Gradient-boosting decision tree	Supervised classification	A Change in biosignals	Early indication of possible outbreak
Method advantages: • Early warning can direct treatment and prevent spread • Continuously measured without requiring intervention			Method disadvantages: • Disease signal is nonspecific • Requires deployment of device before outbreak	
Cough detected by smart listening device	Regional proposal network	Artificial neural network	B Cough begins	Spike in persons whose symptoms are detected early
Method advantages: • Passively monitor with already adopted devices • Can be used in homes or larger settings (e.g., waiting rooms)			Method disadvantages: • Requires advanced privacy protection schemes • Symptomatic person (i.e., who coughed) may be unknown	
Internet search query for viral testing site	Support vector regression	Supervised classification	C Search query for testing	Hotspot of care-seeking behavior
Method advantages: • Can be inexpensive and centrally monitored • Captures behavior without requiring explicit participation			Method disadvantages: • Testing possibly unrelated to symptom status (e.g., for travel) • Searches may not lead to testing (e.g., resource constraints)	
Symptoms entered into website	Participatory surveillance	Human curated	D Enters symptoms online	Real-time prevalence of possible cases
Method advantages: • Information can be disseminated without bureaucratic delay • Captures mild cases that may not formally test across settings			Method disadvantages: • Participants skew toward persons with high health literacy • Relies on syndromic definitions that may describe many causes	
Test result positive for virus	Traditional public health surveillance	Human curated	E Positive test result returned	Official case counts
Method advantages: • Standard diagnostic accuracy • Mandatory reporting can capture rare and dangerous pathogens			Method disadvantages: • Verification can be slow and expensive • Requires resources that may not be available in certain settings	
Post on social media about diagnosis	Natural-language processing	Supervised classification	F Post diagnosis on social media	Real-time prevalence of confirmed cases
Method advantages: • Rapid collection and dissemination of results • Wide array of users who may be missed by most other systems			Method disadvantages: • Computationally expensive and difficult to parse signal from noise • Symptoms nonverified and can be vulnerable to Internet trolls	
Mask wearing captured by CCTV	Convolutional neural network	Artificial neural network	G Mask wearing starts	Nonpharmaceutical intervention levels

Done right, descriptive epidemiology saves lives

But then if done poorly?

Descriptive isn’t just for hypothesis generation

- We use descriptive epi to determine where disease is occurring, in whom and when
- This allows us to target resources
- During the COVID pandemic, where do we mobilize masks, vaccines, preventive services, implement distancing and lockdowns
- Can do with no assumptions of causation
- It can also be used for hypothesis generation, and this is important, but not necessary

BU

COVID Changed The Way We Look at Descriptive Epidemiology

We went from “so are you a skin doctor” to “when is it going to end”?

https://www.economist.com/graphic-detail/tracking-coronavirus-across-the-world

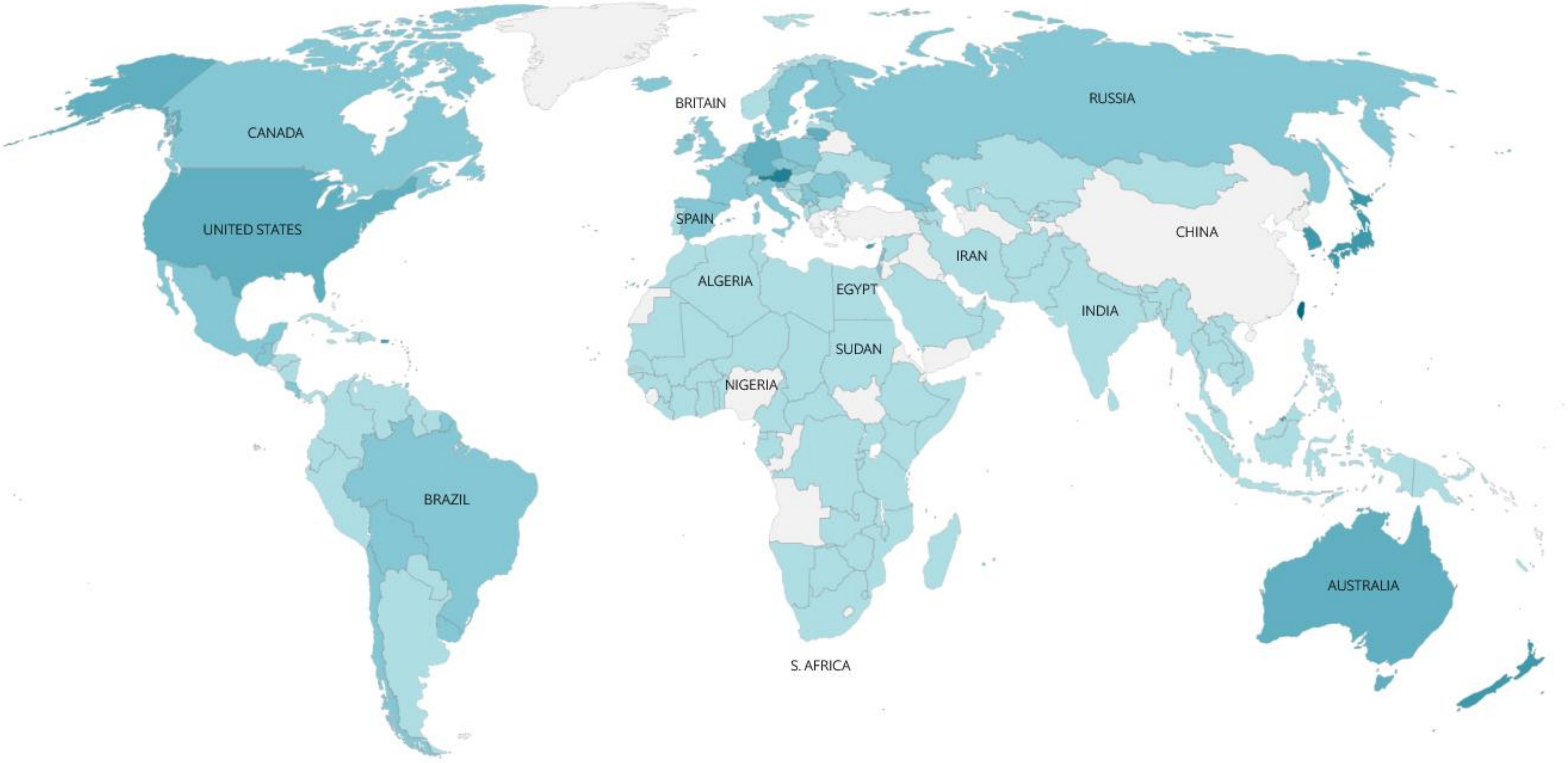
Menu | Weekly edition | The world in brief | Search | Subscribe | Log in

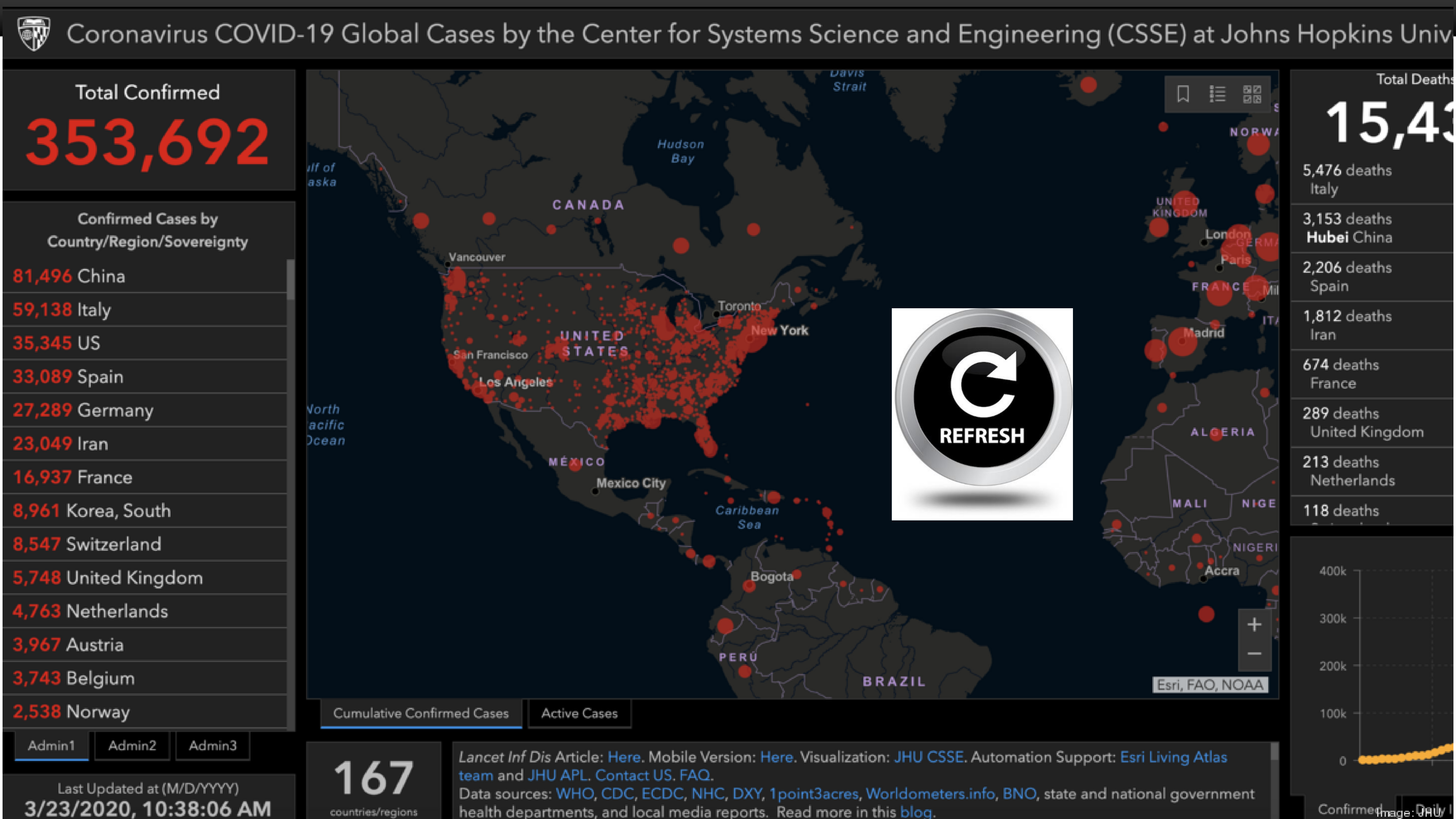
Tracking covid-19 across the world

Vaccines | New cases | New deaths

Cases per 100,000 people in the past 28 days

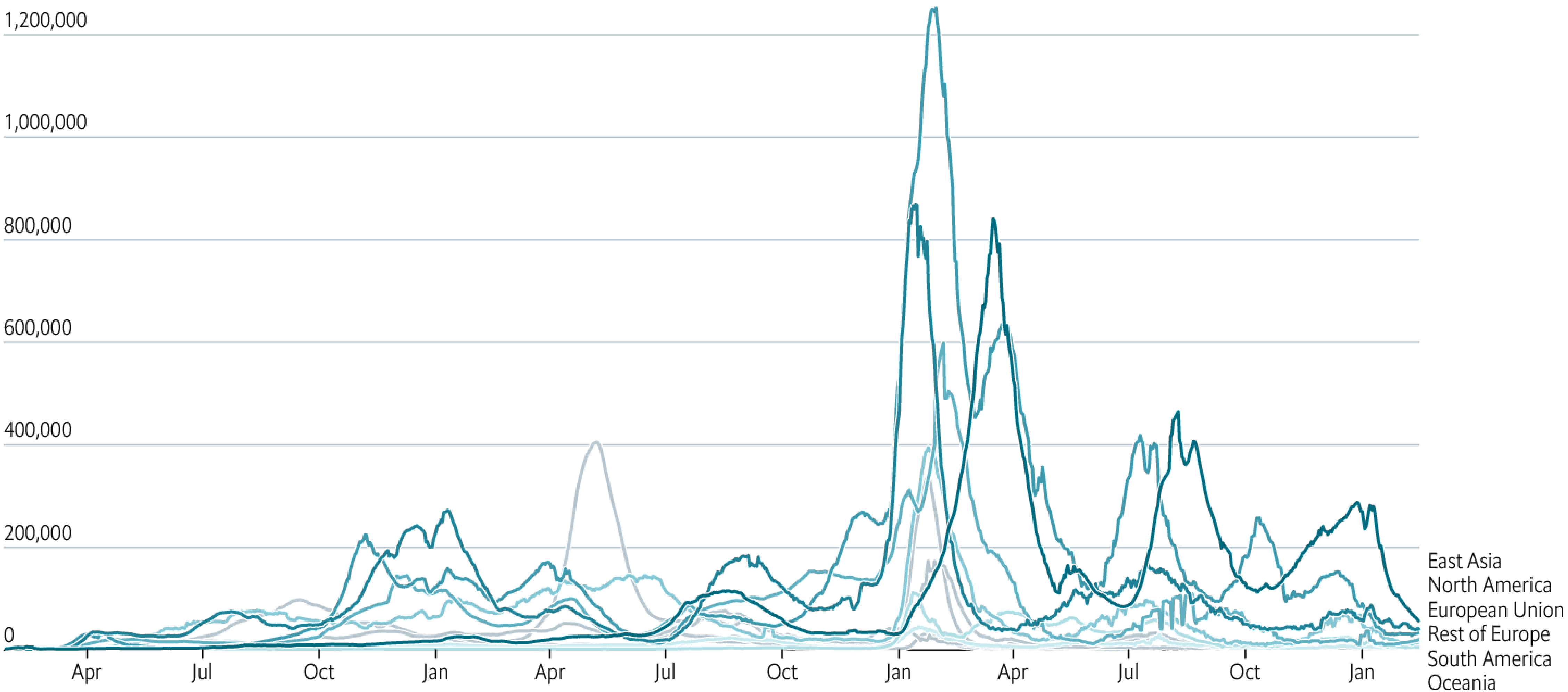
50 250 500 1,000 2,000 Few or no cases




Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins Univ
Total Confirmed
353,692
Confirmed Cases by Country/Region/Sovereignty
81,496 China
59,138 Italy
35,345 US
33,089 Spain
27,289 Germany
23,049 Iran
16,937 France
8,961 Korea, South
8,547 Switzerland
5,748 United Kingdom
4,763 Netherlands
3,967 Austria
3,743 Belgium
2,538 Norway
Admin1
Admin2
Admin3
Last Updated at (M/D/YYYY)
3/23/2020, 10:38:06 AM
CANADA
Hudson Bay
Vancouver
UNITED STATES
San Francisco
Los Angeles
Toronto
New York
MÉXICO
Mexico City
Caribbean Sea
Bogota
PERU
BRAZIL
UNITED KINGDOM
London
Paris
FRANCE
Madrid
ALGERIA
MALI
Accra
REFRESH
Esri, FAO, NOAA
Cumulative Confirmed Cases
Active Cases
167
countries/regions
Lancet Inf Dis Article: Here. Mobile Version: Here. Visualization: JHU CSSE. Automation Support: Esri Living Atlas team and JHU APL. Contact US. FAQ.
Data sources: WHO, CDC, ECDC, NHC, DXY, 1point3acres, Worldometers.info, BNO, state and national government health departments, and local media reports. Read more in this blog.
Total Deaths
5,476 deaths Italy
3,153 deaths Hubei China
2,206 deaths Spain
1,812 deaths Iran
674 deaths France
289 deaths United Kingdom
213 deaths Netherlands
118 deaths
400k
300k
200k
100k
0
Confirmed
Image: JHU

https://www.economist.com/graphic-detail/tracking-coronavirus-across-the-world

Tracking covid-19 across the world



New cases | Total cases

In times of crisis and novelty, descriptive epi is what people crave

Articles

COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study

Sarah Jefferies, Nigel French, Charlotte Gilkison, Giles Graham, Virginia Hope, Jonathan Marshall, Caroline McElnay, Andrea McNeill, Petra Muellner, Shevaun Paine, Namrata Prasad, Julia Scott, Jillian Sherwood, Liang Yang, Patricia Priest

oa

Summary

Background In early 2020, during the COVID-19 pandemic, New Zealand implemented graduated, risk-informed national COVID-19 suppression measures aimed at disease elimination. We investigated their impacts on the epidemiology of the first wave of COVID-19 in the country and response performance measures.

Methods We did a descriptive epidemiological study of all laboratory-confirmed and probable cases of COVID-19 and all patients tested for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in New Zealand from Feb 2 to May 13, 2020, after which time community transmission ceased. We extracted data from the national notifiable diseases database and the national SARS-CoV-2 test results repository. Demographic features and disease outcomes, transmission patterns (source of infection, outbreaks, household transmission), time-to-event intervals, and testing coverage were described over five phases of the response, capturing different levels of non-pharmaceutical interventions. Risk factors for severe outcomes (hospitalisation or death) were examined with multivariable logistic regression and time-to-event intervals were analysed by fitting parametric distributions using maximum likelihood estimation.

Lancet Public Health 2020; 5: e612–23

Published Online October 13, 2020 https://doi.org/10.1016/S2468-2667(20)30225-5

See **Comment** page e569

Institute of Environmental Science and Research, Porirua, New Zealand (S Jefferies MD, C Gilkison MPH, G Graham BSci, V Hope MPhil, A McNeill PhD, S Paine MAE, N Prasad MPH,

https://www.thelancet.com/pdfs/journals/lanpub/PIIS2468-2667(20)30225-5.pdf

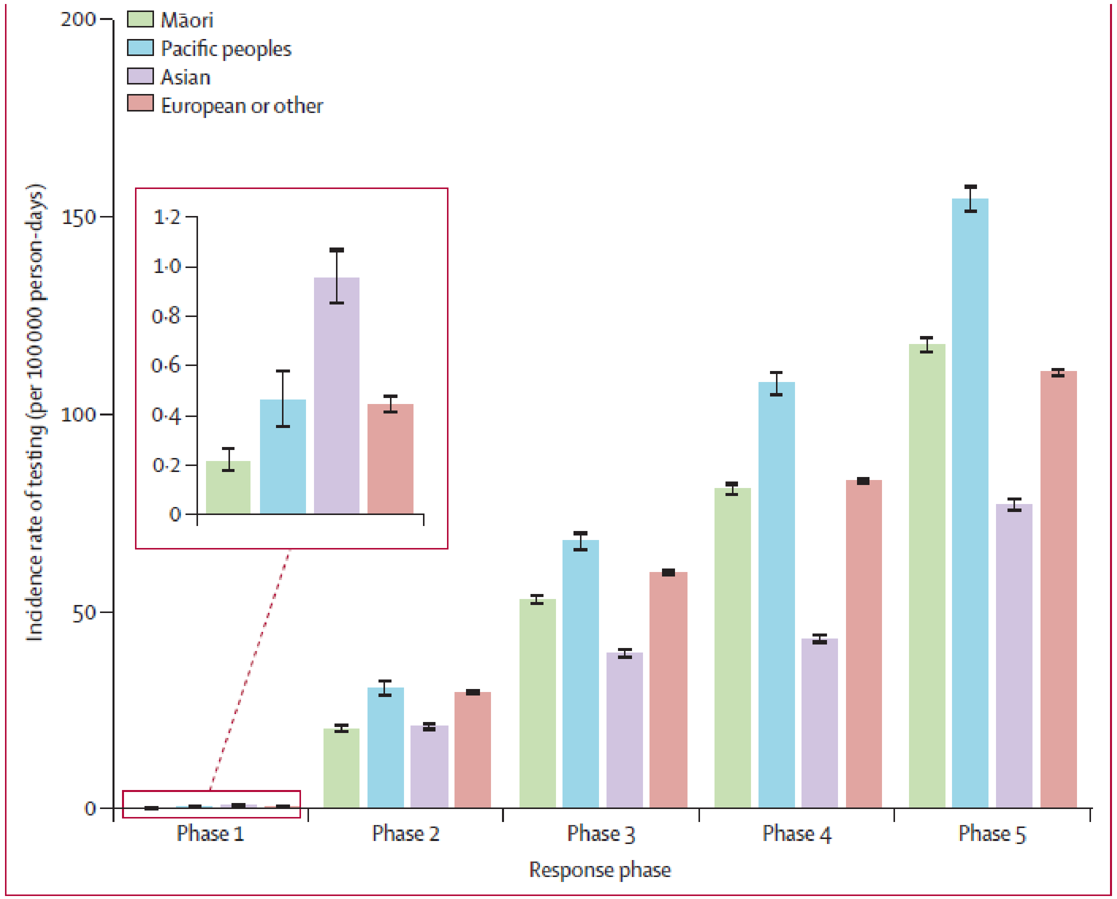


Figure 3: Incidence rates of SARS-CoV-2 testing by sex and response phase (A) and by ethnic group and response phase (B)

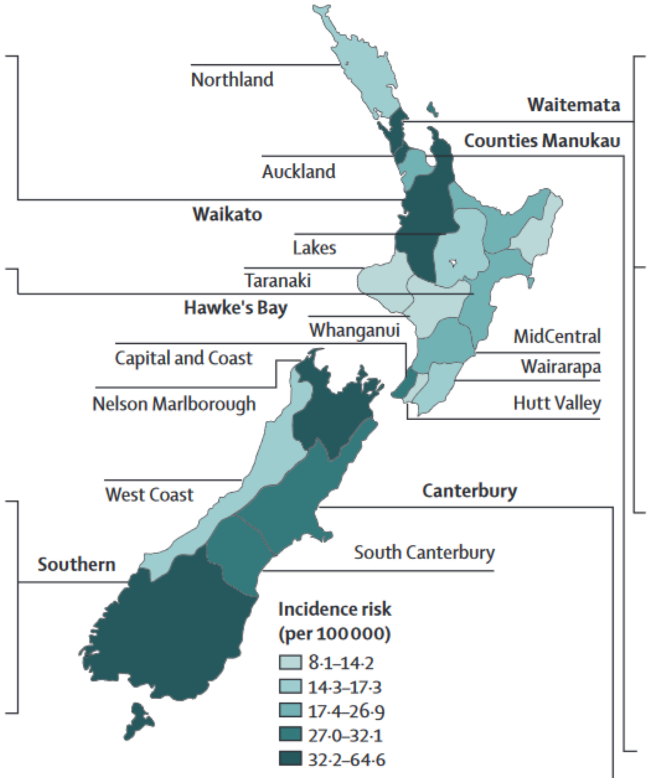
Import-related outbreaks

Hospitality venue: 77 cases
March 15–April 30, three DHBs
Female: 53%
Age: 39 years (27–44)
Māori: 14%, Pacific: 1·3%, Asian: 6·5%, European or other: 77%, unknown: 1·3%
Household transmission: 36%
High-risk workers: 6·5%

Cruise ship and aged residential care: 24 cases
March 16–April 5, four DHBs
Female: 54%
Age: 67 years (52–76)
Māori: 8·3%, Asian: 4·2%, European or other: 88%
Household transmission: 4·0%
High-risk workers: 17%

Wedding: 98 cases
March 19–April 18, seven DHBs
Female: 58%
Age: 49 years (29–57)
Māori: 31%, Asian: 3·1%, European or other: 66%
Household transmission: 26%
High-risk workers: 5·1%

International conference: 39 cases
March 11–April 2, eight DHBs
Female: 46%
Age: 49 years (24–59)
European or other: 100%
Household transmission: 21%
High-risk workers: 10%



Aged residential care: 19 cases
March 11–April 6, one DHB
Female: 79%
Age: 40 years (32–72)

Locally acquired outbreaks

Girls' school: 96 cases
March 12– April 16, four DHBs
Female: 73%
Age: 31 years (15–46)
Māori: 4·2%, Pacific: 18%, Asian: 34%, European or other: 44%
Household transmission: 34%
High-risk workers: 6·3%

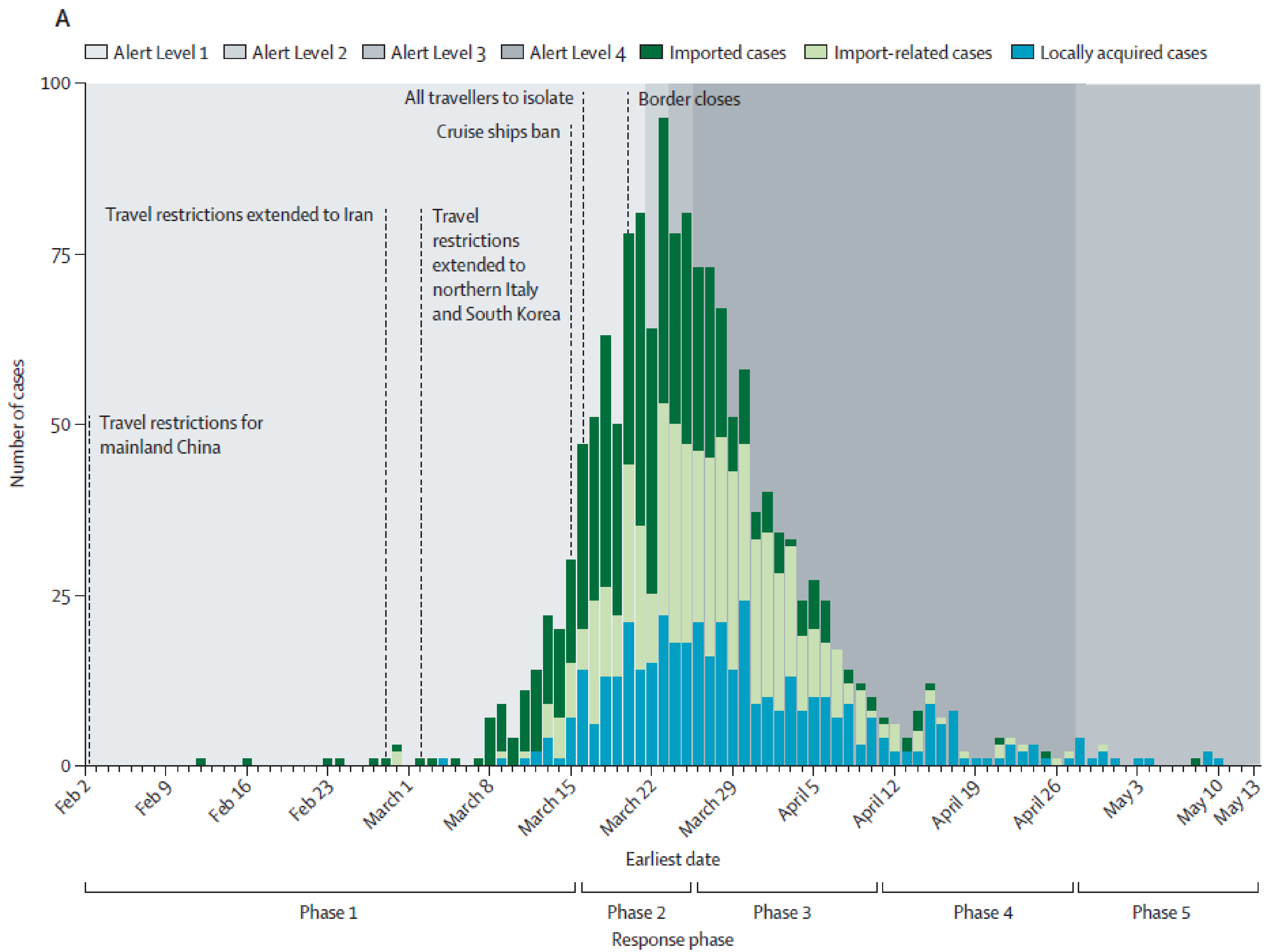
Aged residential care and hospital: 50 cases
March 28–May 9, three DHBs
Female: 74%
Age: 36 years (29–53)
Māori: 6·0%, Pacific: 32%, Asian: 48%, European or other: 14%
Household transmission: 18%
High-risk workers: 58%

Institution and residential facility: 30 cases
March 18–April 8, three DHBs
Female: 53%
Age: 47 years (28–58)
Māori: 6·7%, Pacific: 10%, Asian: 10%, European or other: 73%
Household transmission: 37%
High-risk workers: 30%

Private event: 40 cases
March 9–April 8, six DHBs
Female: 50%
Age: 31 years (24–36)
Māori: 7·5%, Pacific: 7·5%, Asian: 5·0%, European or other: 80%
Household transmission: 40%
High-risk workers: 5·0%

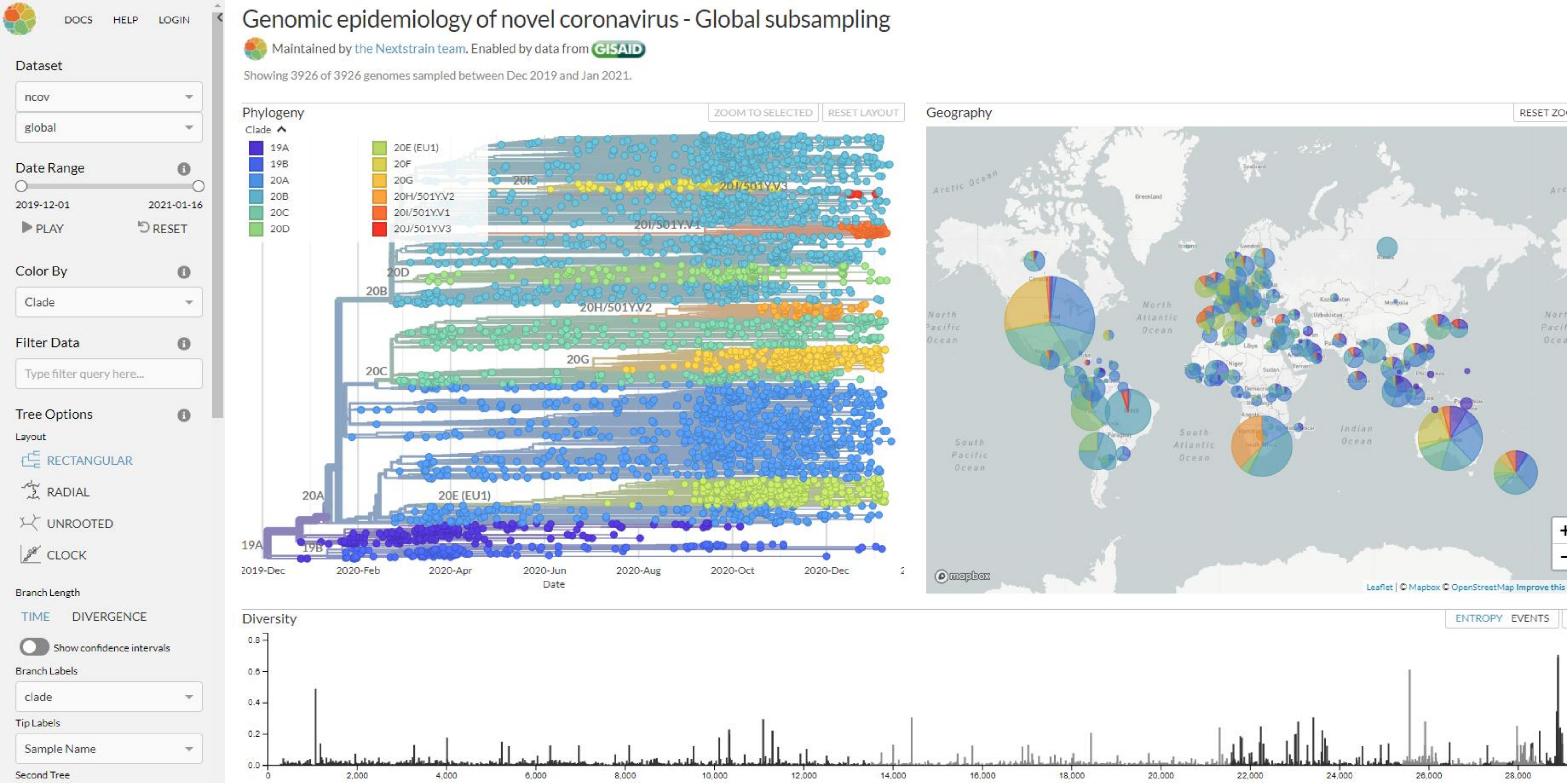
Aged residential care: 56 cases
March 26–May 10, one DHB
Female: 61%
Age: 47 years (26–75)

https://www.thelancet.com/pdfs/journals/lanpub/PIIS2468-2667(20)30225-5.pdf



BU

https://nextstrain.org/ncov/gisaid/global/6m



Descriptive Epi Has Biases Like All Studies

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9383568/

American Journal of Epidemiology

Vol. 191, No. 7
https://doi.org/10.1093/aje/kwac056
Advance Access publication:
March 22, 2022

Commentary

On the Need to Revitalize Descriptive Epidemiology

Matthew P. Fox*, **Eleanor J. Murray, Catherine R. Lesko, and Shawnita Sealy-Jefferson**

* Correspondence to Dr. Matthew Fox, Boston University School of Public Health, 801 Massachusetts Avenue, Room 390, Boston, MA 02118 (e-mail: mfox@bu.edu).

Initially submitted March 4, 2021; accepted for publication March 18, 2022.

Nearly every introductory epidemiology course begins with a focus on person, place, and time, the key components of descriptive epidemiology. And yet in our experience, introductory epidemiology courses were the last time we spent any significant amount of training time focused on descriptive epidemiology. This gave us the impression that descriptive epidemiology does not suffer from bias and is less impactful than causal epidemiology.

https://www.the-scientist.com/news-opinion/high-risk-of-bias-in-early-covid-19-studies-meta-analysis-68353

TheScientist

EXPLORING LIFE, INSPIRING INNOVATION

Search...

NEWS & OPINION PUBLICATIONS CATEGORIES TS

High Risk of Bias in Early COVID-19 Studies: Meta-Analysis

Few peer-reviewed clinical papers on the pandemic contained original data, and many of those that did had poor experimental design.

Max Kozlov

Jan 14, 2021 | 5 min read

PDF VERSION

As scientists led initial investigations into the novel coronavirus last winter and spring, journal publishers saw an enormous surge in COVID-19 publications. A study published January 4 in *BMC Medical Research Methodology* reports that the majority of early clinical studies on the pandemic lacked original data, and those that did were rushed and did not include the appropriate measures to reduce bias.

ABOVE: © ISTOCK.COM, BRIANAJACKSON

Raynaud *et al. BMC Medical Research Methodology* (2021) 21:1
https://doi.org/10.1186/s12874-020-01190-w

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

COVID-19-related medical research: a meta-research and critical appraisal

Check for updates

Marc Raynaud[1†], Huanxi Zhang[2†], Kevin Louis[1†], Valentin Goutaudier[1,3†], Jiali Wang[2], Quentin Dubourg[4], Yongcheng Wei[2], Zeynep Demir[1,5], Charlotte Debiais[1], Olivier Aubert[1], Yassine Bouatou[1], Carmen Lefaucheur[6], Patricia Jabre[7], Longshan Liu[2], Changxi Wang[2], Xavier Jouven[1], Peter Reese[1,8], Jean-Philippe Empana[1] and Alexandre Loupy[1*]

Abstract

Background: Since the start of the COVID-19 outbreak, a large number of COVID-19-related papers have been published. However, concerns about the risk of expedited science have been raised. We aimed at reviewing and categorizing COVID-19-related medical research and to critically appraise peer-reviewed original articles.

Methods: The data sources were Pubmed, Cochrane COVID-19 register study, arXiv, medRxiv and bioRxiv, from 01/11/2019 to 01/05/2020. Peer-reviewed and preprints publications related to COVID-19 were included, written in English or Chinese. No limitations were placed on study design. Reviewers screened and categorized studies according to *i)* publication type, *ii)* country of publication, and *iii)* topics covered. Original articles were critically appraised using validated quality assessment tools.

Results: Among the 11,452 publications identified, 10,516 met the inclusion criteria, among which 7468 (71.0%) were peer-reviewed articles. Among these, 4190 publications (56.1%) did not include any data or analytics (comprising expert opinion pieces). Overall, the most represented topics were infectious disease (n = 2326, 22.1%), epidemiology (n = 1802, 17.1%), and global health (n = 1602, 15.2%). The top five publishing countries were China (25.8%), United States (22.3%), United Kingdom (8.8%), Italy (8.1%) and India (3.4%). The dynamic of publication showed that the exponential growth of COVID-19 peer-reviewed articles was mainly driven by publications without original data (mean 261.5 articles ± 51.1 per week) as compared with original articles (mean of 69.3 ± 22.3 articles per week). Original articles including patient data accounted for 713 (9.5%) of peer-reviewed studies. A total of 576 original articles (80.8%) showed intermediate to high risk of bias. Last, except for simulation studies that mainly used large-scale open data, the median number of patients enrolled was of 102 (IQR = 37–337).

Conclusions: Since the beginning of the COVID-19 pandemic, the majority of research is composed by publications without original data. Peer-reviewed original articles with data showed a high risk of bias and included a limited number of patients. Together, these findings underscore the urgent need to strike a balance between the velocity and quality of research, and to cautiously consider medical information and clinical applicability in a pressing, pandemic context.

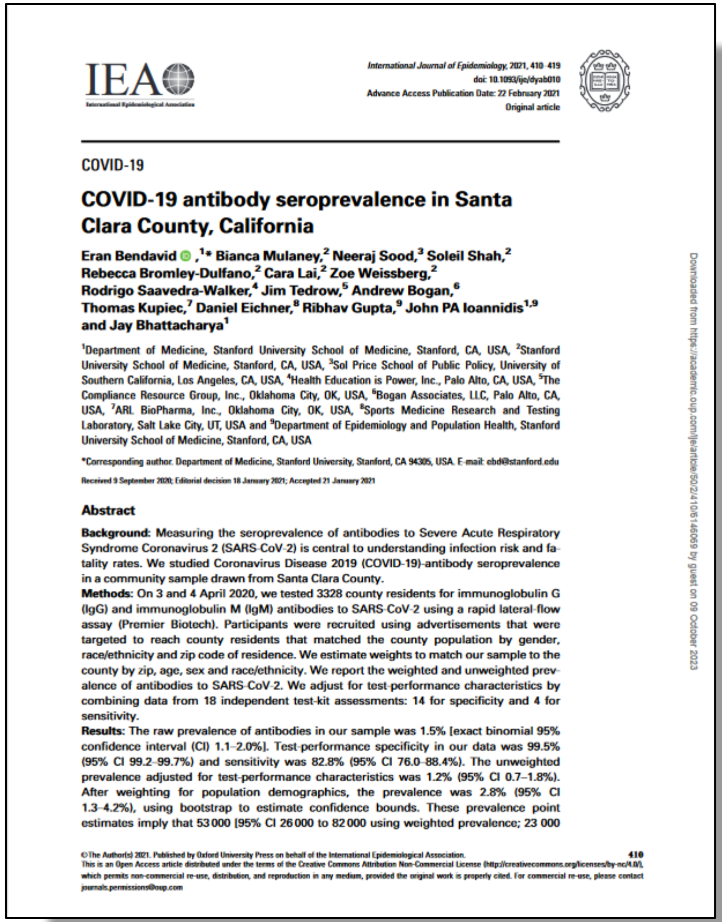
(Continued on next page)

* Correspondence: alexandre.loupy@inserm.fr
†Marc Raynaud, Huanxi Zhang, Kevin Louis, and Valentin Goutaudier contributed equally to the article as co-first author.
[1]Paris Translational Research Epidemiology and Biostatistics Department, Hôpital Necker, 149 rue de Sèvres, 75015 Paris, France
Full list of author information is available at the end of the article

BMC

Internal vs external validity in descriptive epi

- Unlike causal epidemiology, descriptive epidemiology does not suffer from confounding, but may still want to stratify, occasionally adjust
- But selection bias, surveillance bias, missing data and measurement error are still issues
- And samples needs to be externally valid
- Early COVID prevalence surveys suffered from selection bias and measurement error
- How well do we teach sampling?


IEA
International Epidemiological Association
International Journal of Epidemiology, 2021, 410–419
doi: 10.1093/ije/dyab010
Advance Access Publication Date: 22 February 2021
Original article

COVID-19

COVID-19 antibody seroprevalence in Santa Clara County, California

Eran Bendavid,[1,*] Bianca Mulaney,[2] Neeraj Sood,[3] Soleil Shah,[2] Rebecca Bromley-Dulfano,[2] Cara Lai,[2] Zoe Weissberg,[2] Rodrigo Saavedra-Walker,[4] Jim Tedrow,[5] Andrew Bogan,[6] Thomas Kupiec,[7] Daniel Eichner,[8] Ribhav Gupta,[9] John PA Ioannidis[1,9] and Jay Bhattacharya[1]

[1]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA, [2]Stanford University School of Medicine, Stanford, CA, USA, [3]Sol Price School of Public Policy, University of Southern California, Los Angeles, CA, USA, [4]Health Education is Power, Inc., Palo Alto, CA, USA, [5]The Compliance Resource Group, Inc., Oklahoma City, OK, USA, [6]Bogan Associates, LLC, Palo Alto, CA, USA, [7]ARL BioPharma, Inc., Oklahoma City, OK, USA, [8]Sports Medicine Research and Testing Laboratory, Salt Lake City, UT, USA and [9]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA

*Corresponding author. Department of Medicine, Stanford University, Stanford, CA 94305, USA. E-mail: ebd@stanford.edu

Received 9 September 2020; Editorial decision 18 January 2021; Accepted 21 January 2021

Abstract

Background: Measuring the seroprevalence of antibodies to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is central to understanding infection risk and fatality rates. We studied Coronavirus Disease 2019 (COVID-19)-antibody seroprevalence in a community sample drawn from Santa Clara County.
Methods: On 3 and 4 April 2020, we tested 3328 county residents for immunoglobulin G (IgG) and immunoglobulin M (IgM) antibodies to SARS-CoV-2 using a rapid lateral-flow assay (Premier Biotech). Participants were recruited using advertisements that were targeted to reach county residents that matched the county population by gender, race/ethnicity and zip code of residence. We estimate weights to match our sample to the county by zip, age, sex and race/ethnicity. We report the weighted and unweighted prevalence of antibodies to SARS-CoV-2. We adjust for test-performance characteristics by combining data from 18 independent test-kit assessments: 14 for specificity and 4 for sensitivity.
Results: The raw prevalence of antibodies in our sample was 1.5% [exact binomial 95% confidence interval (CI) 1.1–2.0%]. Test-performance specificity in our data was 99.5% (95% CI 99.2–99.7%) and sensitivity was 82.8% (95% CI 76.0–88.4%). The unweighted prevalence adjusted for test-performance characteristics was 1.2% (95% CI 0.7–1.8%). After weighting for population demographics, the prevalence was 2.8% (95% CI 1.3–4.2%), using bootstrap to estimate confidence bounds. These prevalence point estimates imply that 53 000 [95% CI 26 000 to 82 000 using weighted prevalence; 23 000

BU

https://academic.oup.com/aje/article/189/6/511/5811164

American Journal of Epidemiology

Vol. 189, No. 6
DOI: 10.1093/aje/kwaa030
Advance Access publication:
March 24, 2020

Commentary

Catherine R. Lesko*, Alexander P. Keil, and Jessie K. Edwards

The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job

A solid understanding of biases that plague single-sample estimation problems is made even more urgent as our modes of communication and transportation and our expectations of privacy evolve. Staple sampling methodologies such as random-digit dialing or using Department of Motor Vehicle registries are becoming less reliable. New sampling methods (e.g., Internet sampling, respondent-driven sampling) can reach hidden populations or return large samples quickly (17–21), but drawing population-level inference may be challenging.

BU

https://hrr.w.uib.no/files/2018/09/Rothman-2013-Why-representativene.pdf

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2013;**42**:1012–1014
doi:10.1093/ije/dys223

POINT COUNTERPOINT

Why representativeness should be avoided

Kenneth J Rothman,[1,2] John EJ Gallacher[3] and Elizabeth E Hatch[1]

[1]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, [2]RTI Health Solutions, RTI International, Research Triangle Park, NC, USA and [3]Institute of Primary Care and Public Health, Cardiff University, Cardiff, UK

bers from these other age-groups'. But we in fact acknowledged that there is a role for representativeness is certain circumstances, as when 'public-health professionals may rely on representative samples to describe the health status of specific populations'.[5]

The essence of ... wood in a cert... derived by gene... statement mean... always produce fire. The art of discovery is therefore the art of correct generalisation. What is irrelevant, such as the particular shape or size of the piece of wood used, is to be excluded from the generalisation; what is relevant, for example, the dryness of the wood, is to be included in it. The meaning of the term relevant can thus be defined: that is relevant which must be mentioned for the generalisation to be valid. The separation of relevant from irrelevant factors is the beginning of knowledge.

...s, in- ...eople. ...ocess ... way nature works. That process is uncertain, along with everything else in empirical science, but it is not an extrapolation from sample to target population. When Pasteur created the experiment that refuted the theory of spontaneous generation, he used a goose-neck flask to allow air to contact his cooling broth without letting organisms settle into the broth. His concern was to control the conditions in

Measurement Error

- Caused great problems in COVID pandemic
- Surveillance bias/missing data also parts of the measurement conundrum
- Is distinction between non-differential vs differential and dependent still important?

American Journal of Epidemiology

Vol. 191, No. 7
https://doi.org/10.1093/aje/kwac056
Advance Access publication:
March 22, 2022

Commentary

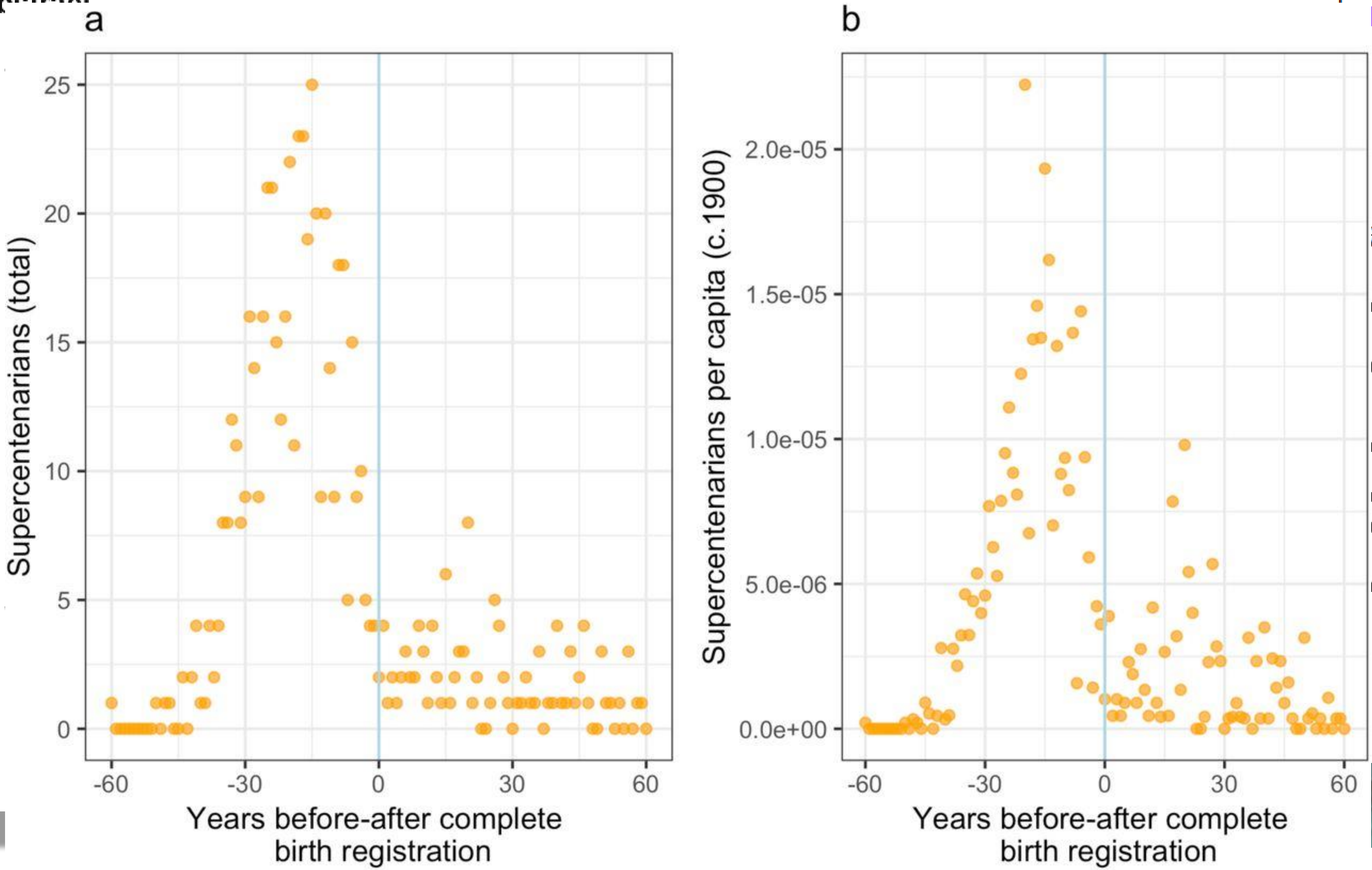
On the Need to Revitalize Descriptive Epidemiology

Matthew P. Fox*, Eleanor J. Murray, Catherine R. Lesko, and Shawnita Sealy-Jefferson

Early COVID-19 work provides an example of how a poorly conceived sampling strategy can lead to poor inference in descriptive epidemiology. Early serological surveys enrolled a sample of volunteers interested in knowing their SARS-CoV-2 antibody status (12). These surveys likely oversampled people who had experienced COVID-19-like symptoms during the Spring of 2020, and the resulting prevalence estimates were likely much higher than would have been obtained from a random population sample. Here, measurement error (as the tests used were not perfect at detecting COVID-19 antibodies and, due to low prevalence of COVID-19, false positives probably overwhelmed true positives) also likely biased results. However, in contrast to causal analyses, confounding bias is not an issue (indeed, confounding bias is not defined for this question).

BU

Ab...



erical errors

https://www.nature.com/articles/s41467-020-18272-4.pdf

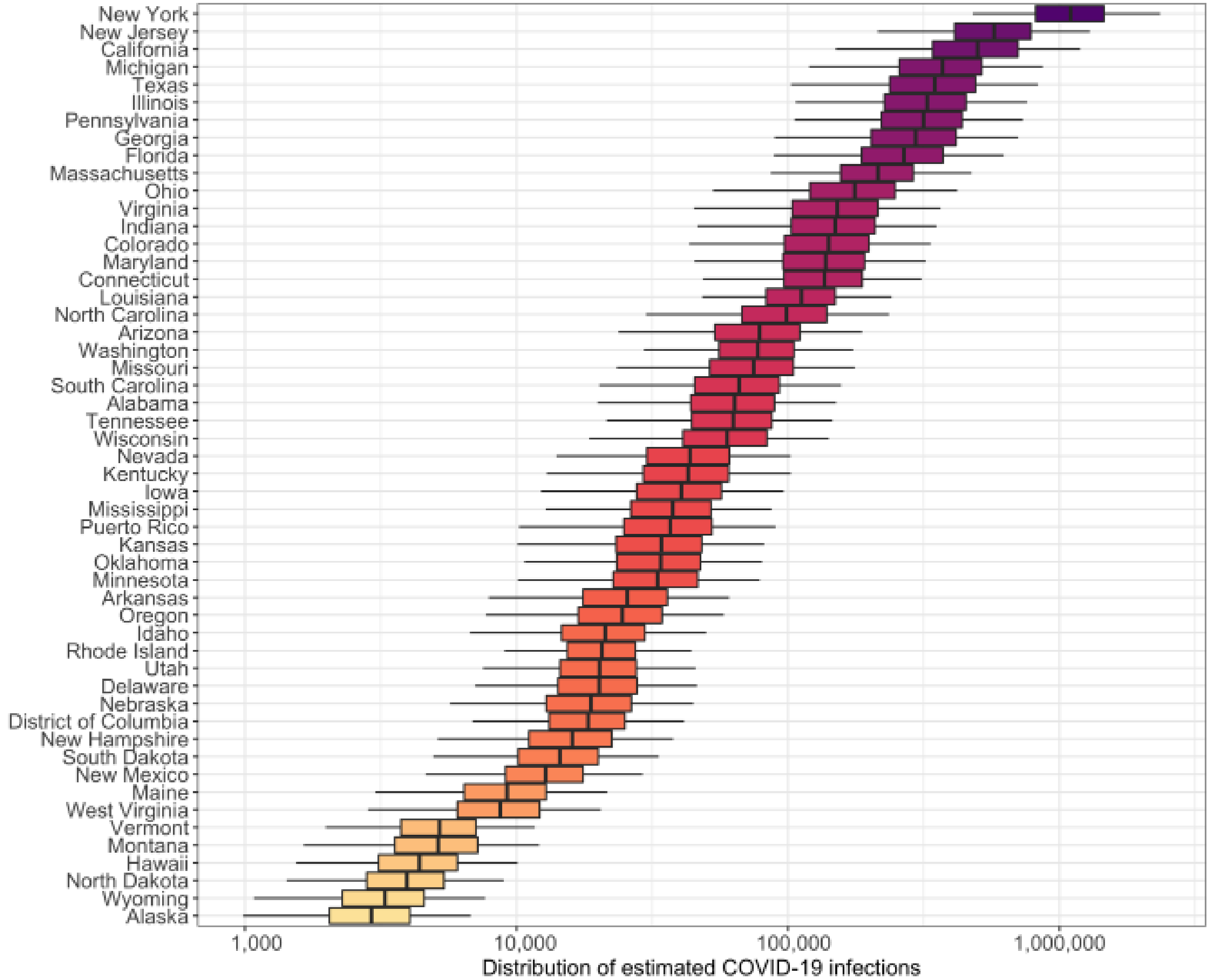
ARTICLE

https://doi.org/10.1038/s4

Substantia infection i

Sean L. Wu [1], Andre
Stephanie Djajadi [1],
Benjamin F. Arnold[6,7],

Supplementary Figure 1. Distribution of expected SARS-CoV-2 infections by state correcting for bias due to incomplete testing and imperfect test accuracy

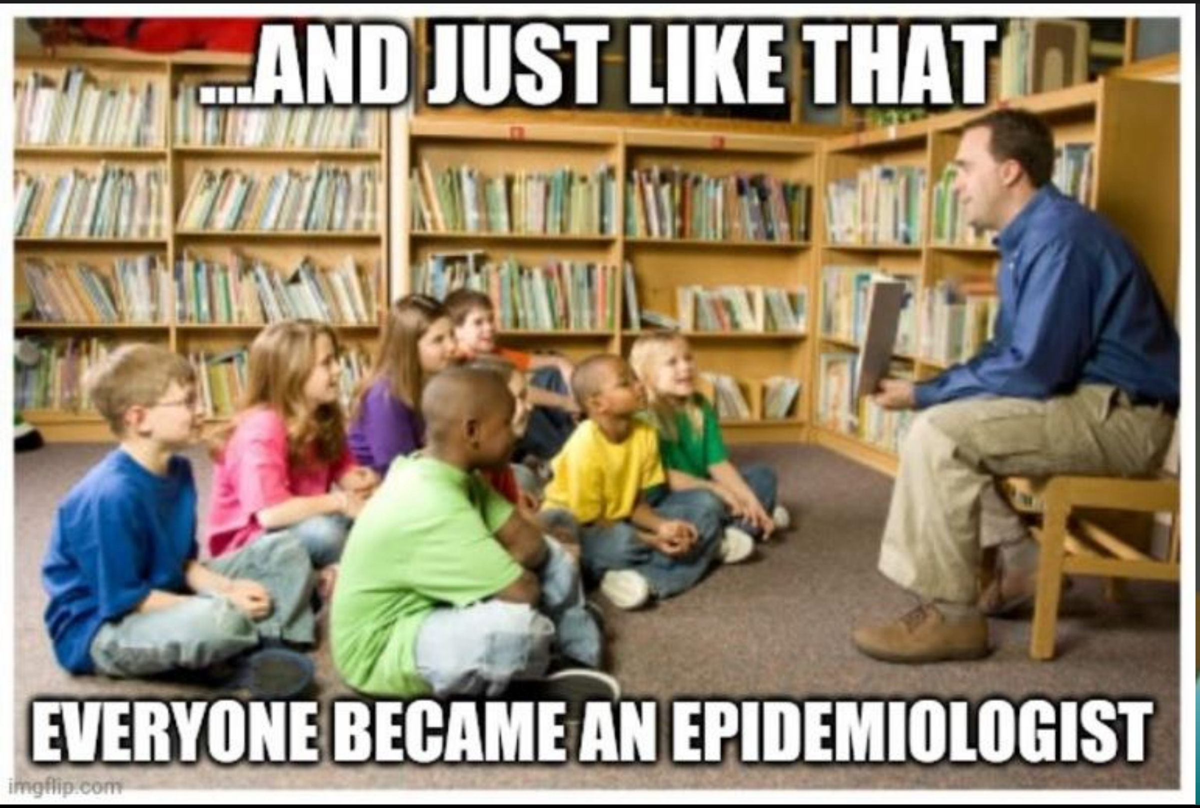


Accurate estimates of the
response. Confirmed COV
pandemic because testin
severe symptoms due to
bias analysis to account
estimate 6,454,951 cumu
0.2% of the population) in t
the number of infections du
confirmed cases. 86% (simulation interval: 64–99%) of this difference is due to incomplete

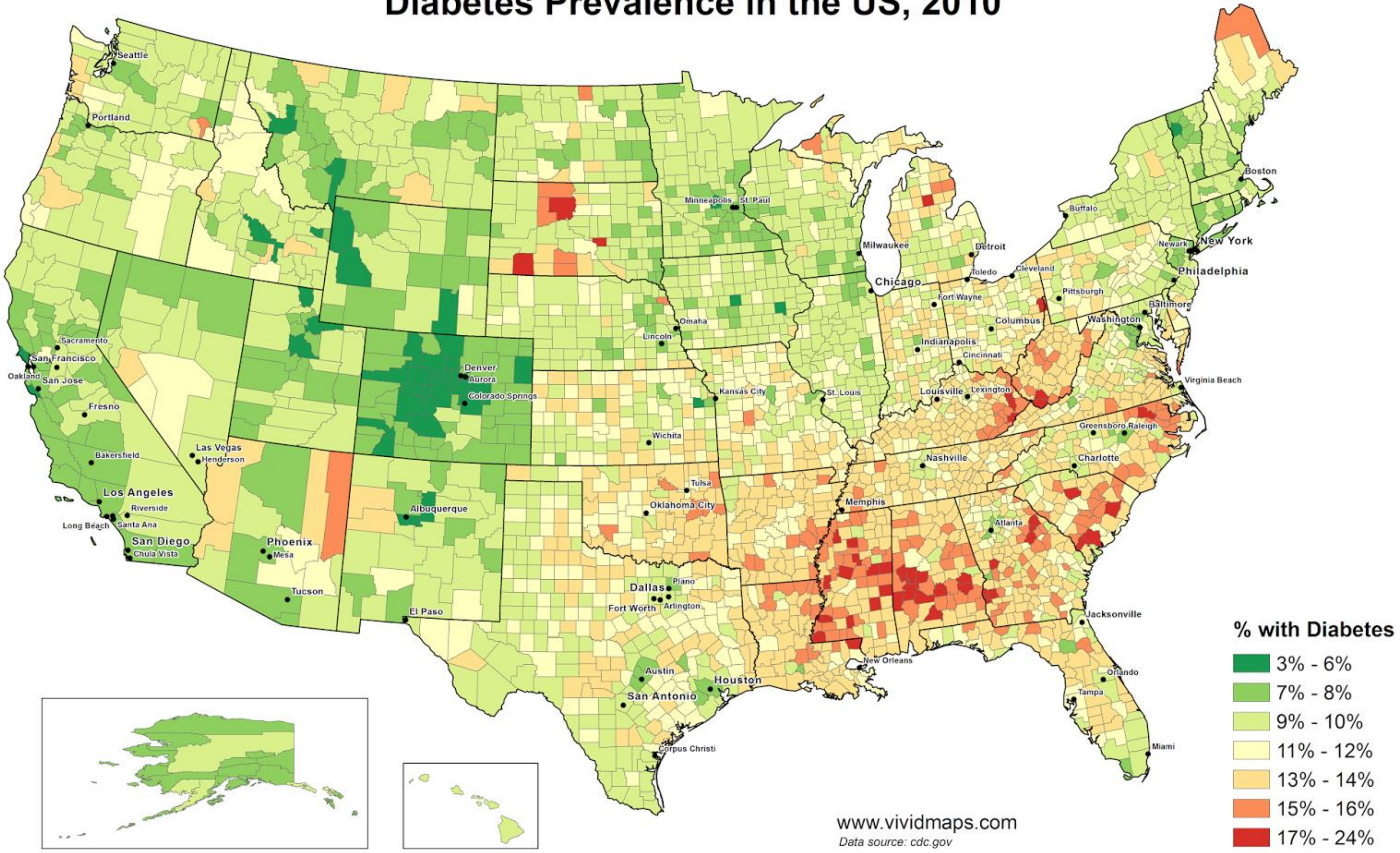
nique that attempts to correct for
out how the data are biased away
bias or misclassification). When
are treated as random variables
ure is known as probabilistic bias
an because it defines prior dis-
formal likelihood function to
d data[22]. In cases where a like-
limited information to update
fully Bayesian approaches lead to
certainty in several of the input
tment may remain partially
g priors. Since our goal was to
burden that removed bias due to
istics, we decided to use a
ons.
mplex mathematical relationship
ytic treatment of the induced
timates is often intractable, and the
mulation. In order to correct
bias (preferential testing of
moderate-severe cases) and imperfect diagnostic accuracy, we developed a simple

But be careful of our biases…

But descriptive is not causal
And yet, the world thinks descriptive is easy



Diabetes Prevalence in the US, 2010

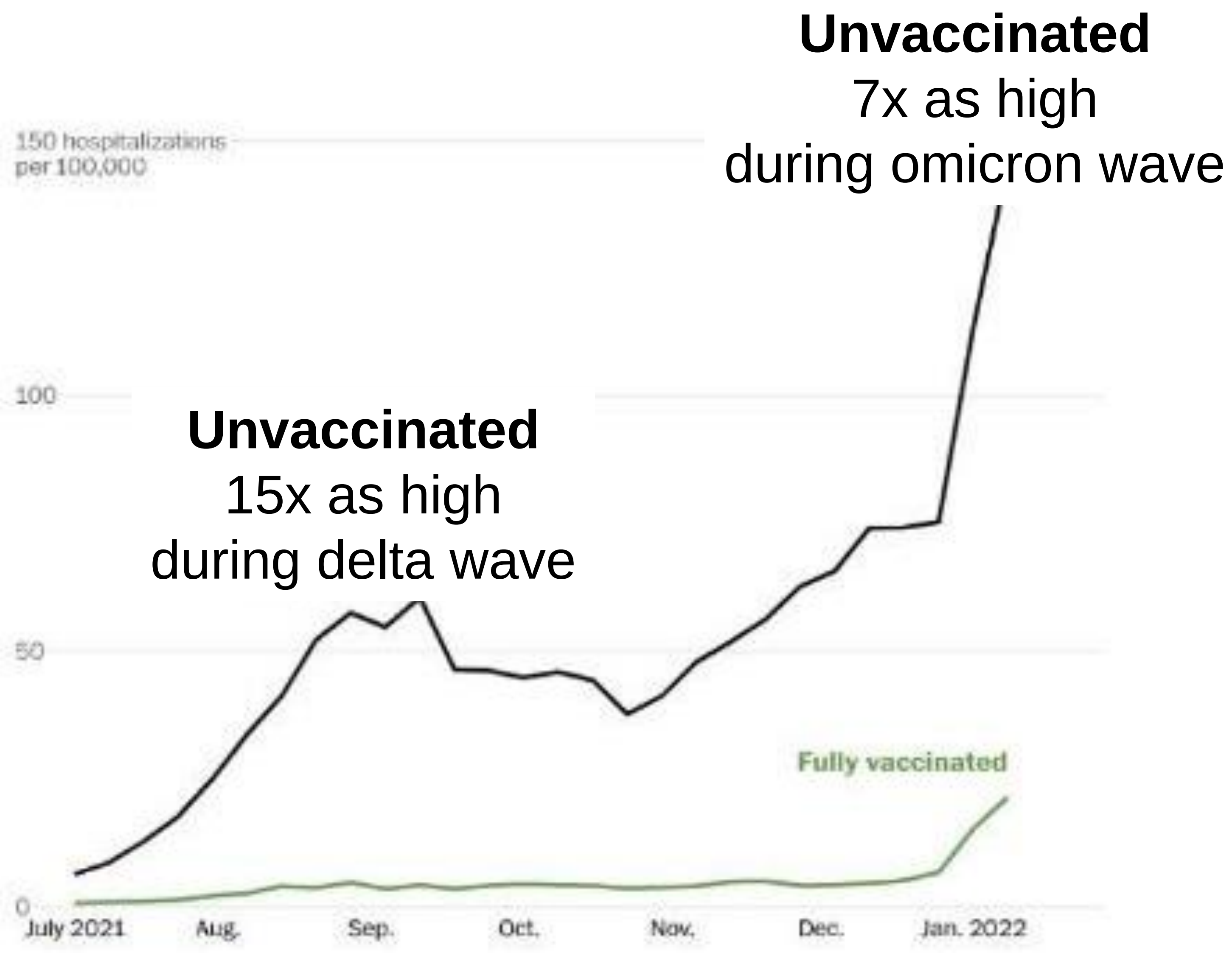


Correlation is not causation, but it sure does look like it…

https://rb.gy/b4bxy

The Washington Post
Democracy Dies in Darkness

Coronavirus vaccine protection was much weaker against omicron, data shows

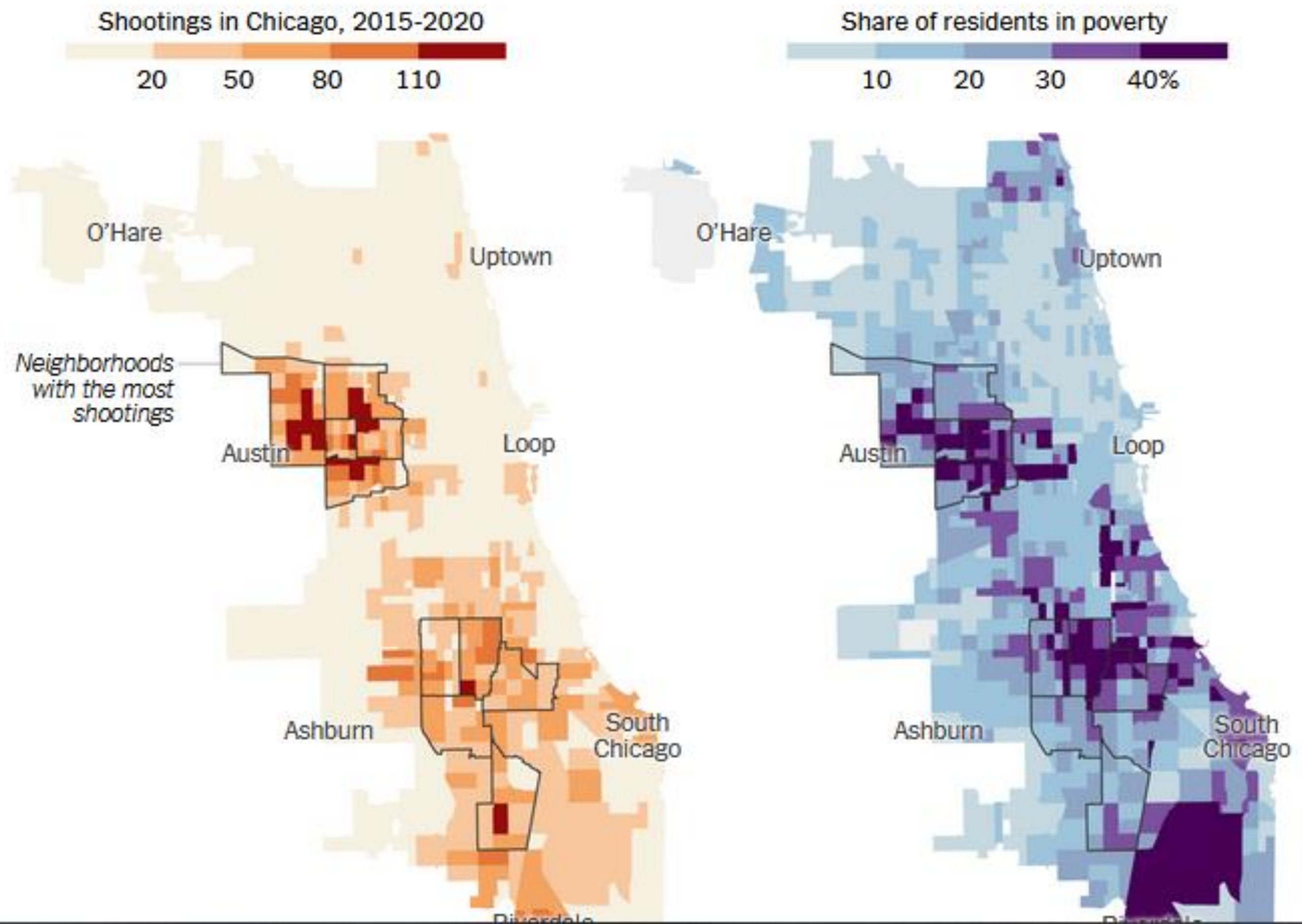


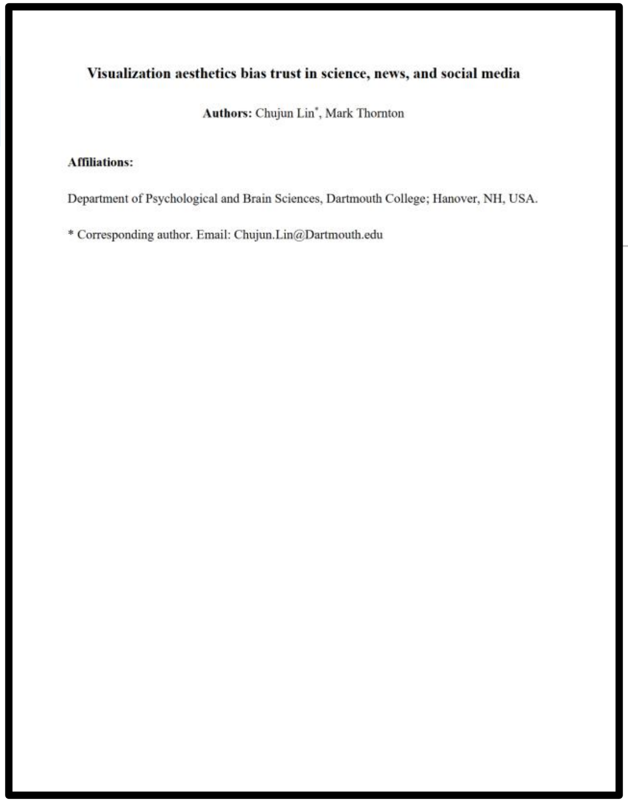
BU

Poverty and violence

There are several factors behind the concentration of violence. A major one is poverty.

In Chicago, violence and poverty closely overlap, as these maps demonstrate:





Visualization aesthetics bias trust in science, news, and social media

Authors: Chujun Lin*, Mark Thornton

Affiliations:

Department of Psychological and Brain Sciences, Dartmouth College; Hanover, NH, USA.

* Corresponding author. Email: Chujun.Lin@Dartmouth.edu

Visualization aesthetics bias trust in science, news, and social media

visual appeal, independent of data quality. Here we tested whether the beauty of a graph influences how much people trust it. Across three studies, we sampled graphs from social media, news reports, and scientific publications, and consistently found that graph beauty predicted trust. In a fourth study, we manipulated both the graph beauty and misleadingness. We found that beauty, but not actual misleadingness, causally affected trust. These findings reveal a source of

Conclusions

- Descriptive epidemiology is an essential function of our discipline and yet it doesn’t receive the prestige it should
- During the pandemic there was renewed interest in descriptive epi as we recognized its central role in responding to the pandemic
- But to be useful, descriptive epidemiology must be done with care and rigor and with attention to key sources of bias
- Renewed training in descriptive epi could really help improve out outputs

Thank you!

How much time did you spend on descriptive epidemiology in your epi training?

BOSTON UNIVERSITY

https://academic.oup.com/aje/article/189/6/511/5811164

American Journal of Epidemiology

Vol. 189, No. 6
DOI: 10.1093/aje/kwaa030
Advance Access publication:
March 24, 2020

Commentary

Catherine R. Lesko*, Alexander P. Keil, and Jessie K. Edwards

The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job

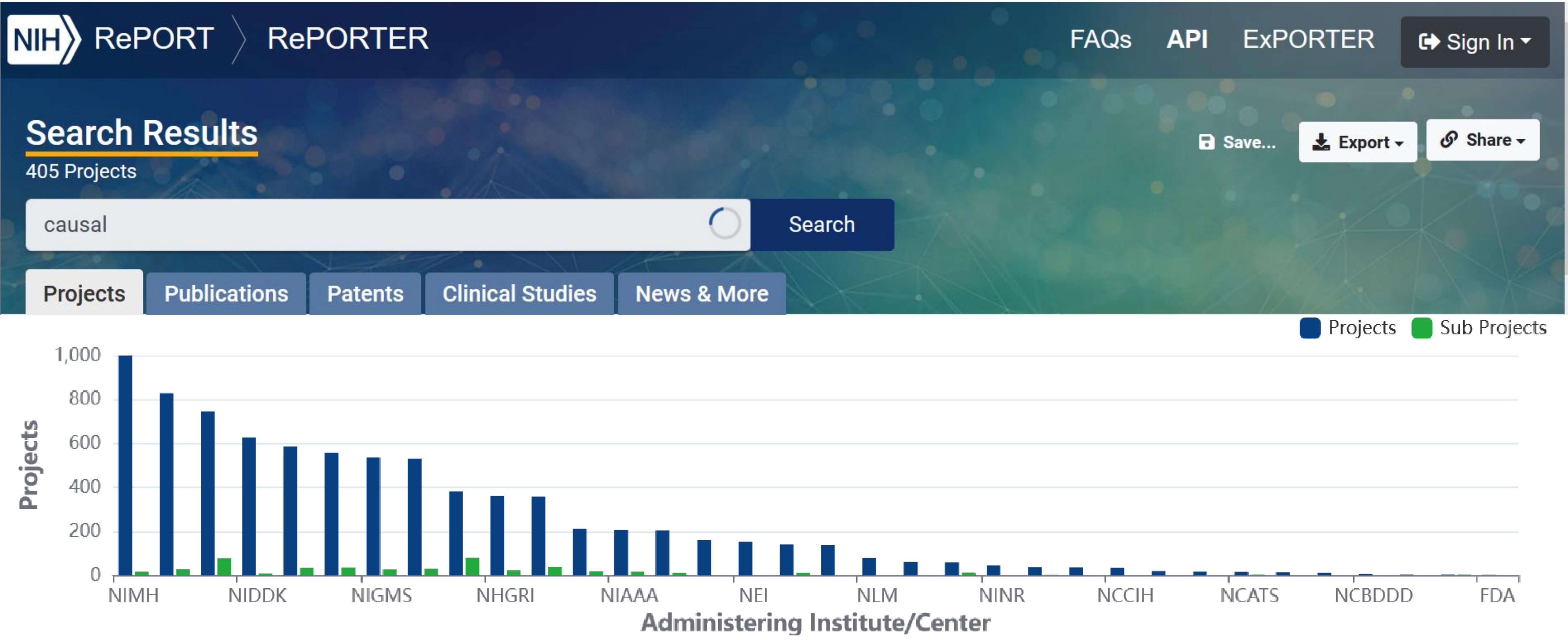
it is evident that 1) epidemiologic principles and methods are applicable to many questions beyond causal effect estimation, and 2) epidemiologic curricula and journals have prioritized analytic epidemiology and questions related to identifying (causal) determinants of disease over descriptive epidemiology and questions related to accurately characterizing the health of populations (14, 15). Descriptive epidemiologic studies are frequently excluded from peer-reviewed journals for not being generalizable enough. We

Do we value descriptive epidemiology?

- How many of you work primarily on descriptive epidemiologic studies?
- How many of you would feel you had the skills to properly evaluate as a peer reviewer a descriptive study?
- How many of you have more than cursory training in descriptive epidemiology?
- How many of you would be comfortable with your students writing a dissertation that is entirely descriptive epidemiology?

Descriptive Research at NIH

Causal Research at NIH



https://academic.oup.com/aje/article/189/6/511/5811164

American Journal of Epidemiology

Vol. 189, No. 6
DOI: 10.1093/aje/kwaa030
Advance Access publication:
March 24, 2020

Commentary

Catherine R. Lesko*, Alexander P. Keil, and Jessie K. Edwards

The Epidemiologic Toolbox: Identifying, Honing, and Using the Right Tools for the Job

tive courses. An epidemiology curriculum that emphasized descriptive epidemiology might spend the entirety of the first term on single-sample estimation problems and describing the natural course of disease (i.e., the course of disease in the absence of any interventions). This could be framed in terms of designing a target study, or an idealized study that would accurately estimate the descriptive parameter of interest in the absence of real-world constraints like missing data and measurement error (foreshadowing introduction of the target trial as a heuristic for study design for causal effect estimation but encompassing a broader set of questions).

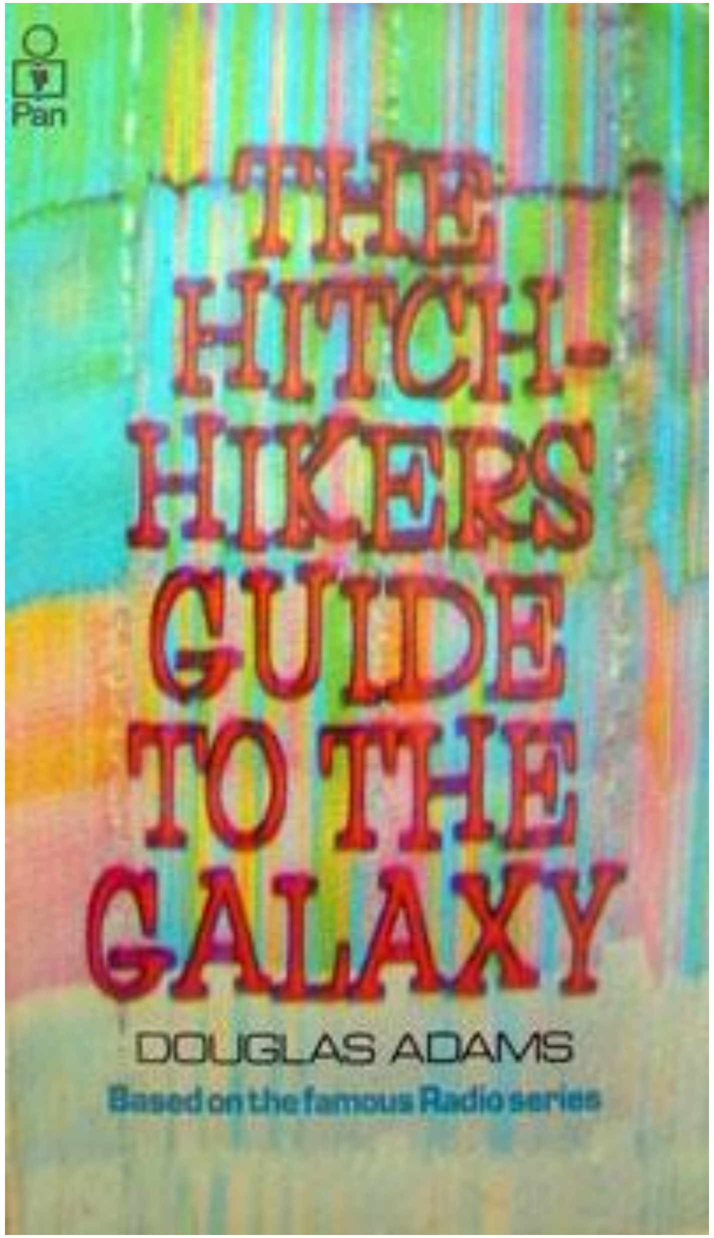
What is the goal of your study?

BOSTON UNIVERSITY

The number 42 is, in *The Hitchhiker's Guide to the Galaxy* by Douglas Adams, the "Answer to the Ultimate Question of Life, the Universe, and Everything", calculated by an enormous supercomputer named Deep Thought over a period of 7.5 million years. Unfortunately, no one knows what the question is. Thus, to calculate the Ultimate

The Answer to the Ultimate Question of Life,
The Universe, and Everything

BU



What’s the question?

Journal of Orthopaedic & Sports Physical Therapy   Current Issue   Just Accepted   Archive

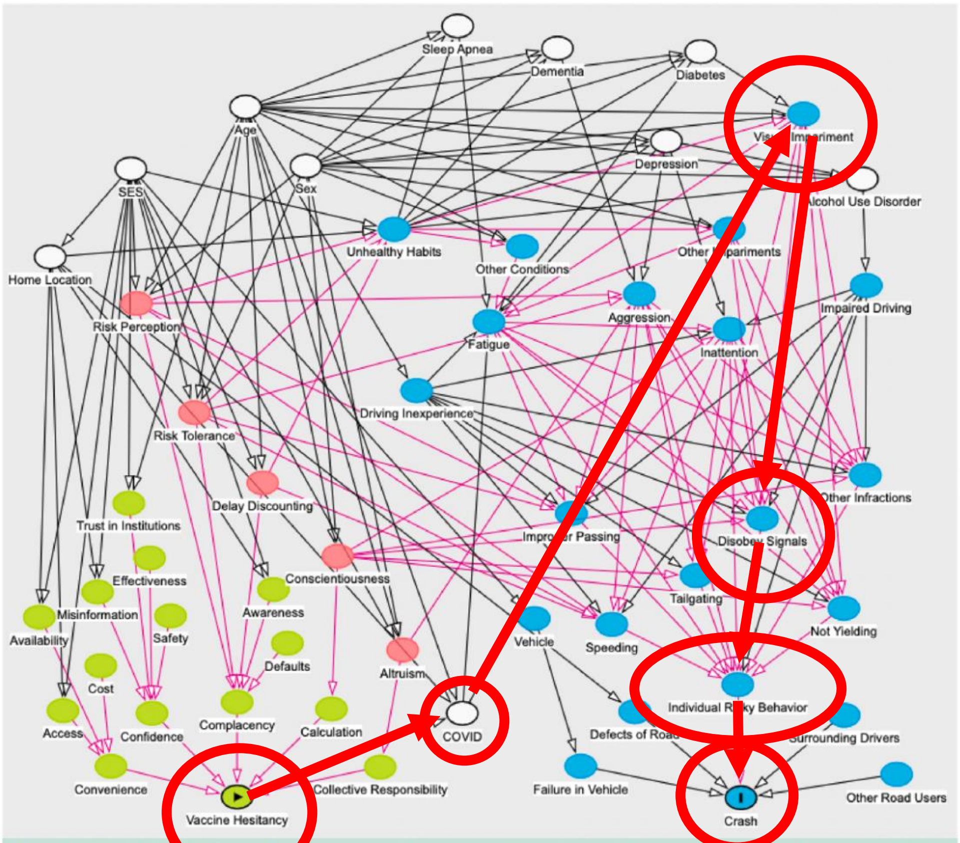
A previous Evidence in Practice article explained why a specific and answerable research question is important for clinicians and researchers. Determining whether a study aims to answer a descriptive, predictive, or causal question should be one of the first things a reader does when reading an article. Any type of question can be relevant and useful to support evidence-based practice, but only if the question is well defined, matched to the right study design, and reported correctly. *J Orthop Sports Phys Ther 2020;50(8):468–469. doi:10.2519/jospt.2020.0703*

- **Likely the biggest problem we have is asking a question that can be translated into statistical notation that can be answered**
- **Often it isn’t clear if interested in description, prediction or causation**
- **Often simpler with descriptive epi, but the problem doesn’t go away**

We should know what the question is…

- **Causation, description or prediction?**
- **Seems at first like causation**
  - “COVID vaccine hesitancy is a reflection of psychology that might also contribute to traffic safety. We tested whether COVID vaccination was **associated with** the risks of a traffic crash.”
  - Vax wouldn’t help here, so not causation.
- **Last bullet seems like description**
  - Find them so we can give them an intervention for traffic safety, not vax to reduce traffic accident/
  - But then why adjust?
- **DAG suggest COVID causes impairments that cause accidents**

§2 Directed Acyclic Graph



Footnote: Directerd Acyclic Graph of possible causal pathways relevant to vaccine hesitancy and traffic risks. The diagram displays measured factors (white), unmeasured ancestors of vaccine hesitancy (green), unmeasured ancestors of traffic risks (blue), and unmeasured ancestors to both vaccine hesitancy and traffic risks (pink). Causal pathways denoted as closed (black lines) or open (magenta lines). Specific causal pathways based on literature review, direct clinical experience (Canada’s largest trauma center), and expert consultation (International Traffic Medicine Association).

Description or Association?

- Well, I guess this is just meant to be descriptive as we can’t say anything causal for sure, so I just want to see if things are associated…
- So why did you adjust?
- If description is the goal, adjustment is usually not necessary or helpful (there are cases where it is)?
- But the opposite can also be true, it can be so tempting to read causation from description…

https://academic.oup.com/aje/article/191/12/2071/6675231

American Journal of Epidemiology

Vol. 191, No. 12
https://doi.org/10.1093/aje/kwac153
Advance Access publication:
August 25, 2022

Invited Commentary

Invited Commentary: The Importance of Descriptive Epidemiology

Robert W. Platt*

* Correspondence to Robert W. Platt, Department of Epidemiology, Biostatistics, and Occupational Health McGill University 2001 McGill College, 12th floor, Montreal QC H3G 1A1, Canada (e-mail: robert.platt@mcgill.ca).