Digital health and computational epidemiology

Lesson 2

Michele Tizzoni
michele.tizzoni@unitn.it
Dipartimento di Sociologia e Ricerca Sociale
Via Verdi 26, Trento
Ufficio 6, 3 piano

UNIVERSITÀ DI TRENTO
FONDAZIONE BRUNO KESSLER

C2S2 Center for Computational Social Science and Human Dynamics

Basic concepts of epidemiology

Epidemiology

- from the Greek “epi” (upon), and “demos” (the population): what is going on in the **population**
- **Epidemiology** is concerned of what happens in groups of individuals, not single individuals.
- We need to define what is the **group: the target population**

Sampling

- Once the target population is established, it is necessary to obtain a sample of it
- The sample must be **representative** (not biased).
- Consider the case “male obese smokers living in the US over the age of 65”. The full sample will be over 1 million.
- The issue of representativeness becomes crucial in digital epidemiology.

Questions

- **Who?** The target group
- **What?** The outcomes of interest
- **Why?** The causes of disease (causal methods)
- **When?** The temporal frame (days, years, centuries)
- **Where?** The location of interest (local, global, etc.)

Outcomes of interest

Morbidity

Suffering from a
non-lethal
reduction in health
through disease or injury

Mortality

Death

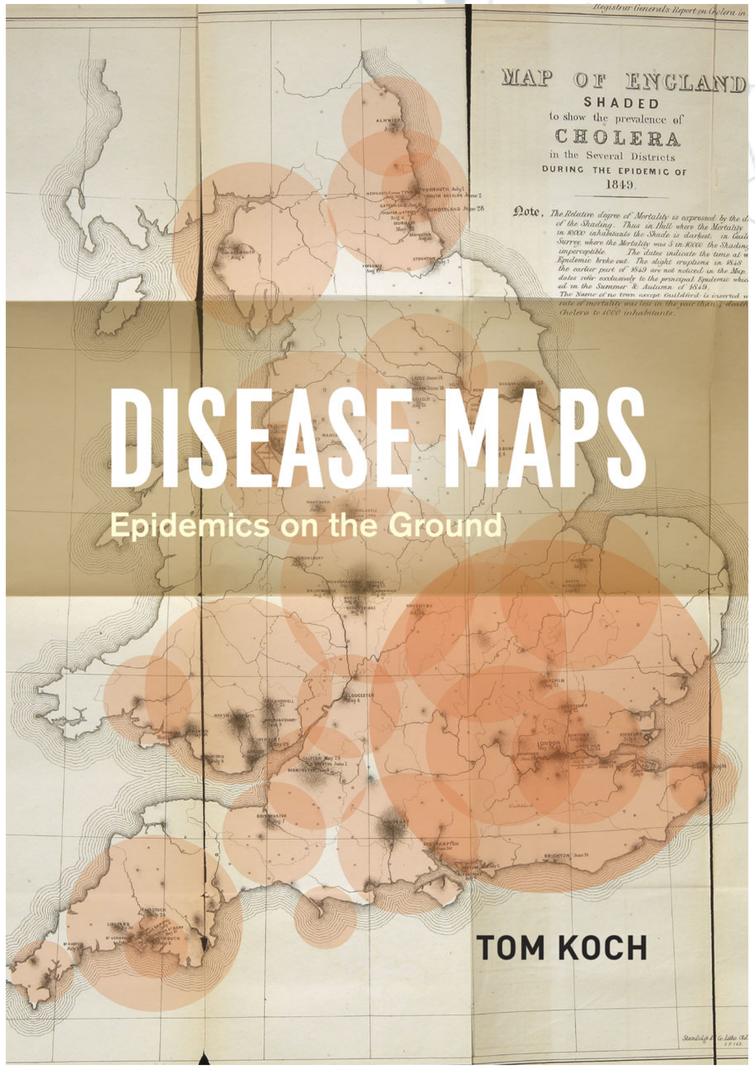
Why?

- The causes of disease and death, and their distributions, are at the core of epidemiology.
- Epidemiology has developed an arsenal of methods to identify causal paths of disease and deaths:
  - Case-control studies
  - Randomized control trials (RCTs)
  - Cohort studies
  - Mathematical modeling

When and where?

- When. It can be a time point in the life of individuals or the temporal evolution of an outbreak (when do people become symptomatic? When do they become infectious?)

- Where. It allows us to understand mechanisms of exposure and define interventions.

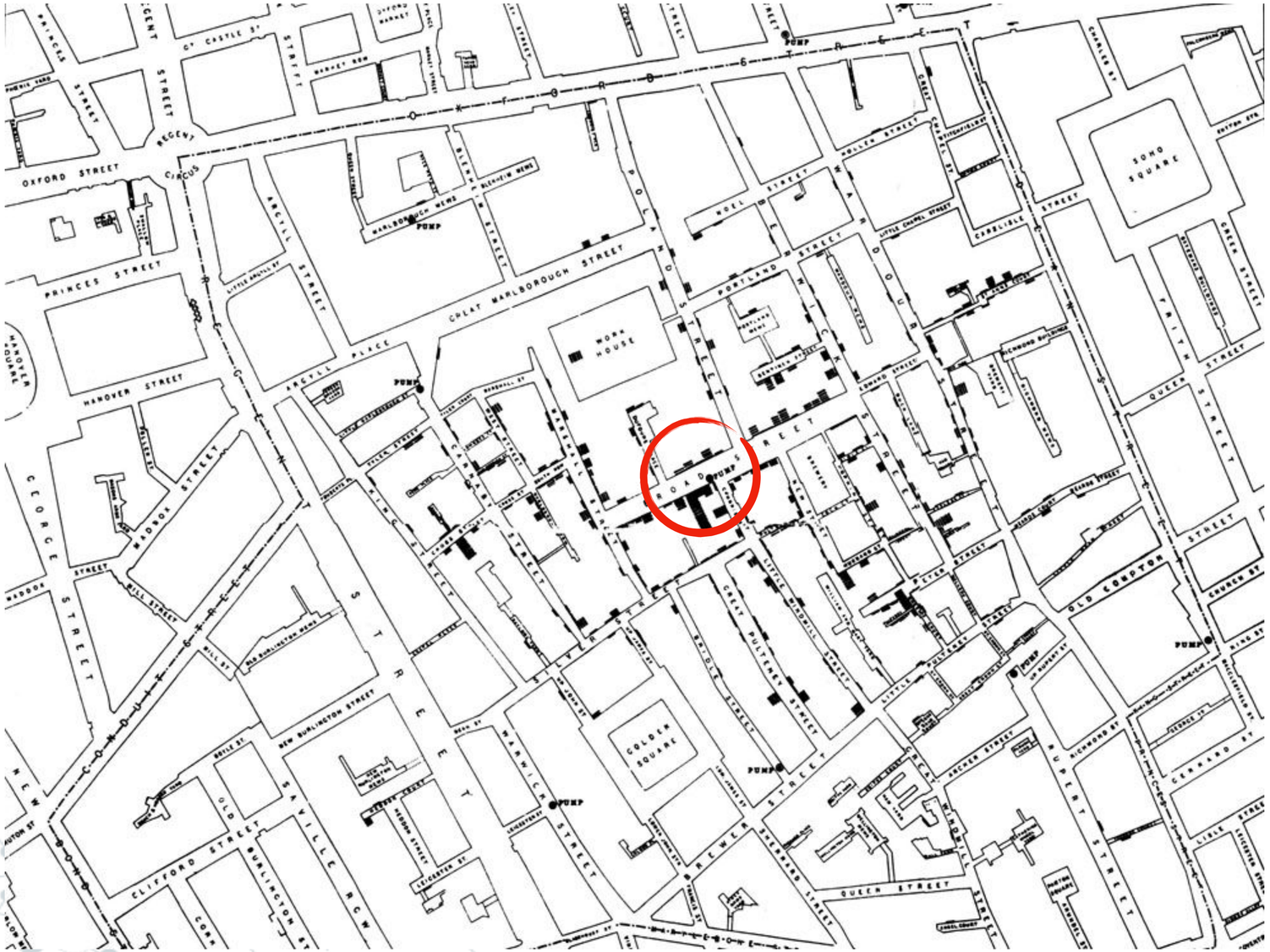


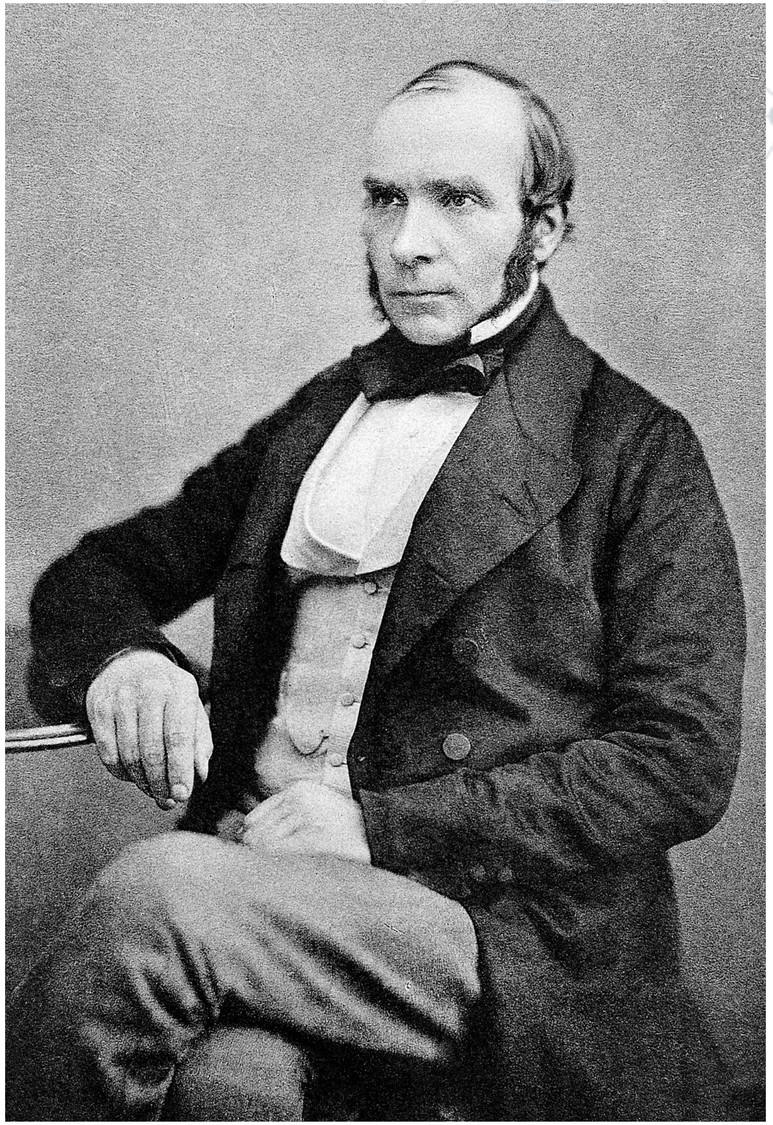
Looking for patterns

Looking for patterns

- Modern epidemiology was born with a map in the XIX century.
- **English physician John Snow** investigated the 1854 cholera outbreak in Soho, London.
- Although the causes of cholera were not known, he identified the source of the outbreak by drawing a map

Looking for patterns





Public health surveillance

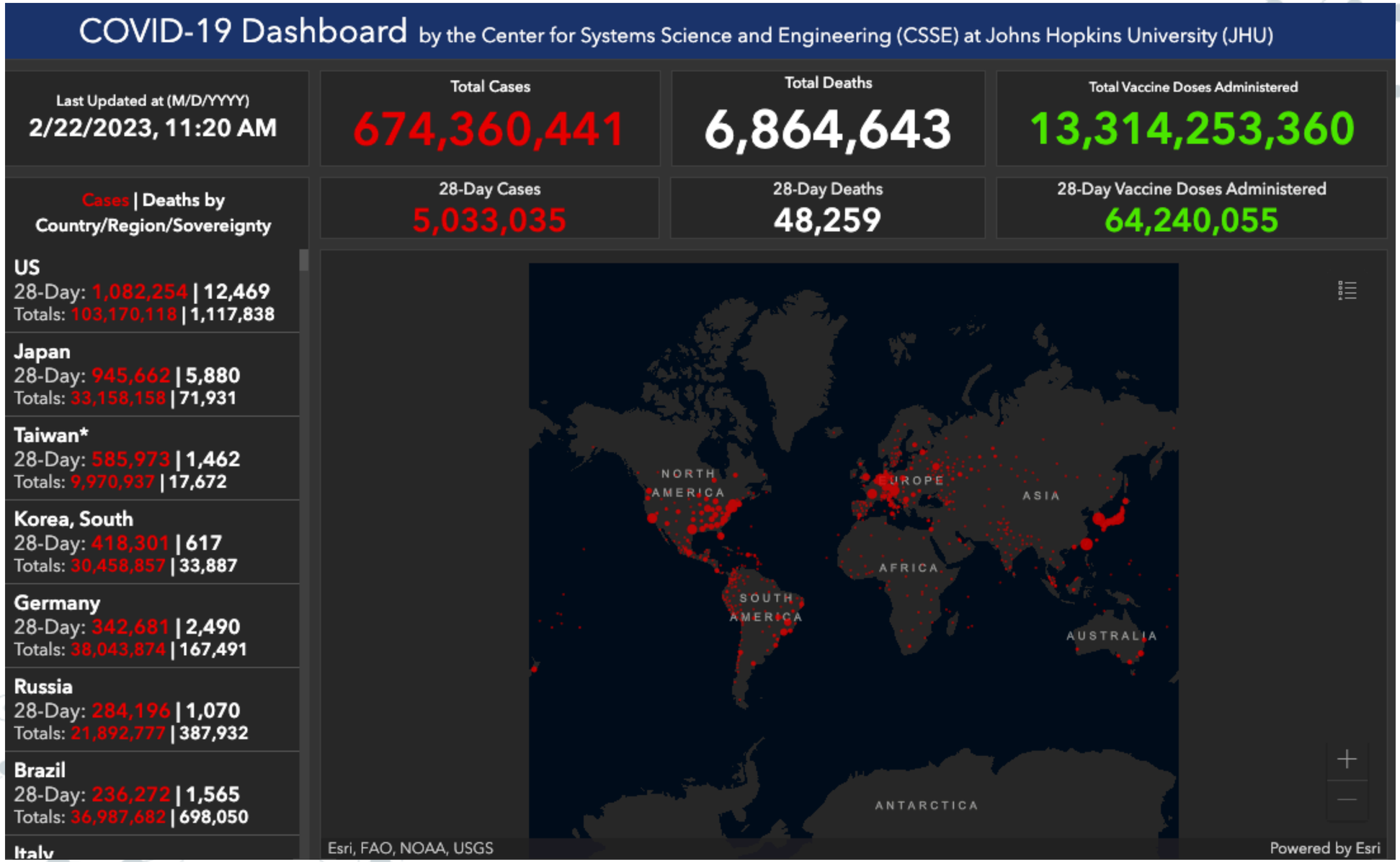
- The Soho outbreak demonstrates that epidemiological investigations can lead to correct conclusions, and propose useful, preventative courses of action even with an initially limited understanding of the biological details involved.

- The example of John Snow demonstrates something else: the importance of **good data**.

- Providing timely and accurate data on the extent of disease spread is core task of public health, generally referred to as **surveillance.**

- **Public health surveillance** has not universal definition but it can be generally described as “the continuous, systematic collection, analysis and interpretation of health-related data” (WHO)

COVID-19


COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)
Last Updated at (M/D/YYYY)
2/22/2023, 11:20 AM
Total Cases
674,360,441
Total Deaths
6,864,643
Total Vaccine Doses Administered
13,314,253,360
28-Day Cases
5,033,035
28-Day Deaths
48,259
28-Day Vaccine Doses Administered
64,240,055
Cases | Deaths by Country/Region/Sovereignty
US
28-Day: 1,082,254 | 12,469
Totals: 103,170,118 | 1,117,838
Japan
28-Day: 945,662 | 5,880
Totals: 33,158,158 | 71,931
Taiwan*
28-Day: 585,973 | 1,462
Totals: 9,970,937 | 17,672
Korea, South
28-Day: 418,301 | 617
Totals: 30,458,857 | 33,887
Germany
28-Day: 342,681 | 2,490
Totals: 38,043,874 | 167,491
Russia
28-Day: 284,196 | 1,070
Totals: 21,892,777 | 387,932
Brazil
28-Day: 236,272 | 1,565
Totals: 36,987,682 | 698,050
Italy
NORTH AMERICA
EUROPE
ASIA
AFRICA
SOUTH AMERICA
AUSTRALIA
ANTARCTICA
Esri, FAO, NOAA, USGS
Powered by Esri

Public health surveillance

- **Passive surveillance.** Data is reported on an ongoing basis by the medical system and information is collected by public health authorities (ISS in Italy).

- **Active surveillance.** Public health officials proactively engage in search for cases to respond to a public health threat.

Active surveillance

Article

Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'

https://doi.org/10.1038/s41586-020-2488-1
Received: 2 April 2020
Accepted: 23 June 2020
Published online: 30 June 2020
Check for updates

Enrico Lavezzo[1,13], Elisa Franchin[1,13], Constanze Ciavarella[2], Gina Cuomo-Dannenburg[2], Luisa Barzon[1], Claudia Del Vecchio[1], Lucia Rossi[2], Riccardo Manganelli[1], Arianna Loregian[1], Nicolò Navarin[4,5], Davide Abate[1], Manuela Sciro[3], Stefano Merigliano[6], Ettore De Canale[3], Maria Cristina Vanuzzo[3], Valeria Besutti[3], Francesca Saluzzo[1], Francesco Onelia[1], Monia Pacenti[3], Saverio G. Parisi[1], Giovanni Carretta[3], Daniele Donato[3], Luciano Flor[3], Silvia Cocchio[7], Giulia Masi[1], Alessandro Sperduti[4,5], Lorenzo Cattarino[2], Renato Salvador[6], Michele Nicoletti[8], Federico Caldart[8], Gioele Castelli[8], Eleonora Nieddu[8], Beatrice Labella[8], Ludovico Fava[8], Matteo Drigo[8], Katy A. M. Gaythorpe[2], Imperial College COVID-19 Response Team[#], Alessandra R. Brazzale[9], Stefano Toppo[1,5], Marta Trevisan[1], Vincenzo Baldo[7], Christl A. Donnelly[2,10], Neil M. Ferguson[2], Ilaria Dorigatti[2]✉ & Andrea Crisanti[1,11]✉

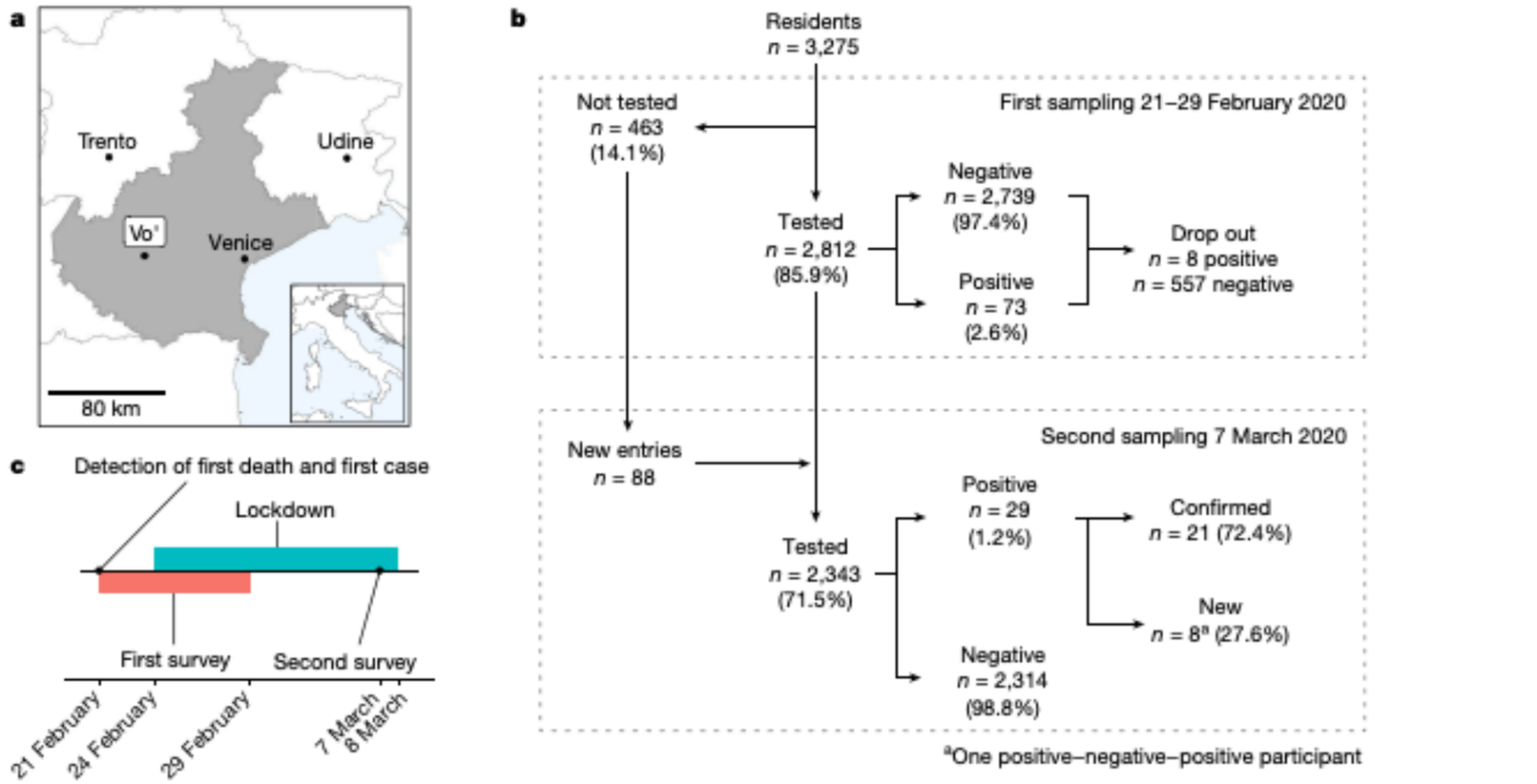


Fig. 1 | Study description. a, Map showing the location of Vo' and the Veneto region (grey area) within Italy, produced using shapefiles from GADM (https://gadm.org/) and Italian National Institute of Statistics (ISTAT; https://www.istat.it/it/archivio/222527 and https://www.istat.it/it/archivio/104317#accordions). **b**, Flow chart summarizing the key statistics on the two sequential nasopharyngeal swab surveys conducted in Vo' to assess the transmission of SARS-CoV-2 before and after the implementation of interventions. **c**, Summary of the key events in the study period.

Public health surveillance

- **Population-based surveillance.** The public health systems surveys everyone in the population of interest. This is very resource-intensive.
- **Sentinel surveillance.** *Sentinels* (health facility, medical doctors) form a sentinel network and report about cases. Influenza surveillance in Europe and the US is done through sentinel doctors (1-5% of total).
  - Biases: not everyone seeks care, highly skewed towards older age groups
  - **Digital surveillance**: collect data through Web-based platforms (more in the next lectures)

Public health surveillance

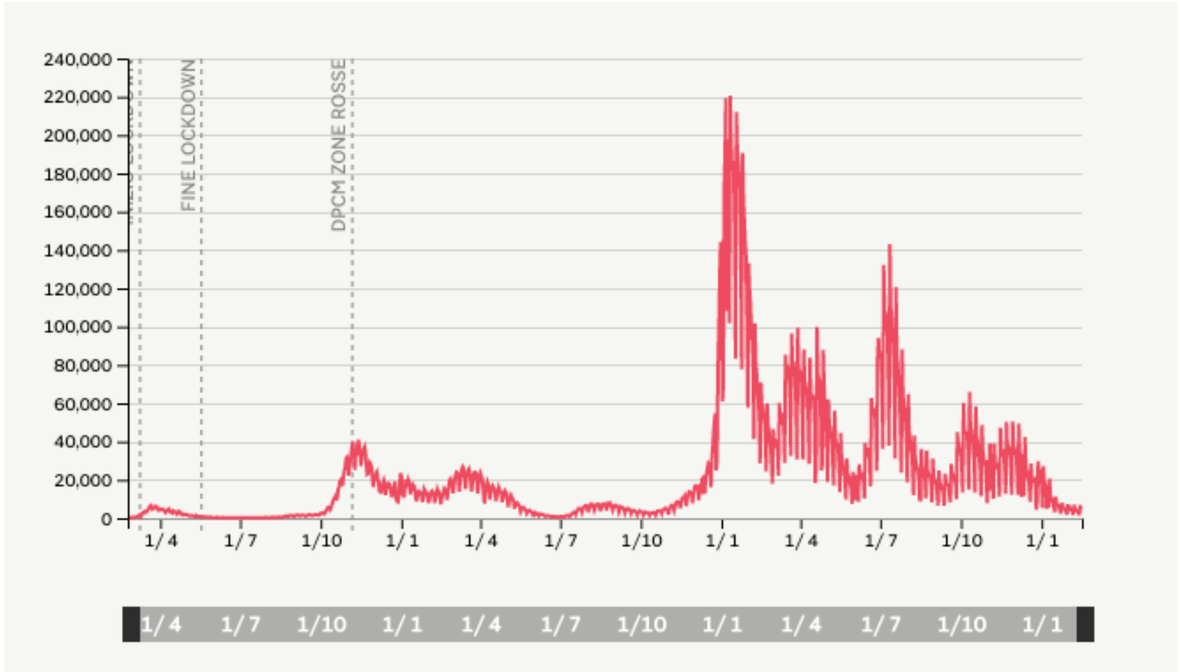
- **Syndromic surveillance.** A case is identified by the symptoms the individual expresses (this is a classic example for influenza, where diagnosis is based on symptoms: influenza-like-illness, ILI).
- **Laboratory-confirmed surveillance.** A case is defined by a lab test. For instance, COVID-19 cases were generally confirmed by RT-PCR or antigen tests.
- **Digital surveillance** is typically syndromic (until now).

Incidence and prevalence

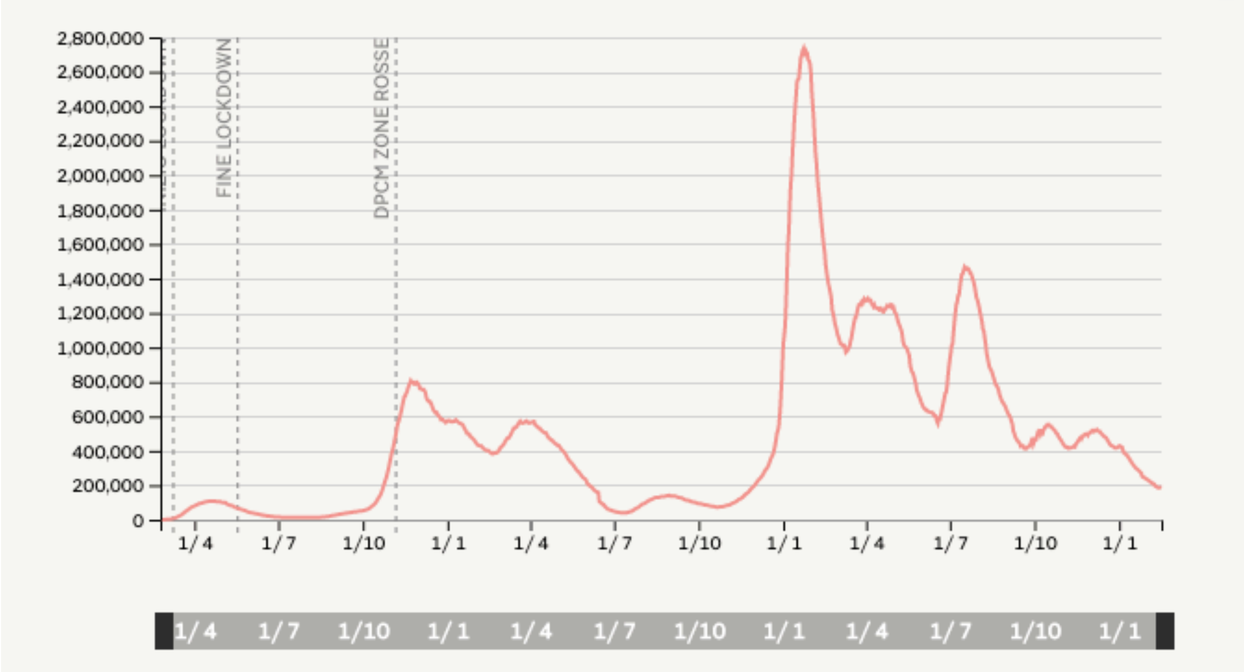
- **Incidence.** The number of **new cases** of a disease during a specified period of time. It is a measure of influx.
- **Prevalence.** The **total number of active cases** that are present during a specified period of time. It is a measure of stock.

Incidence and prevalence

COVID-19 incidence in Italy

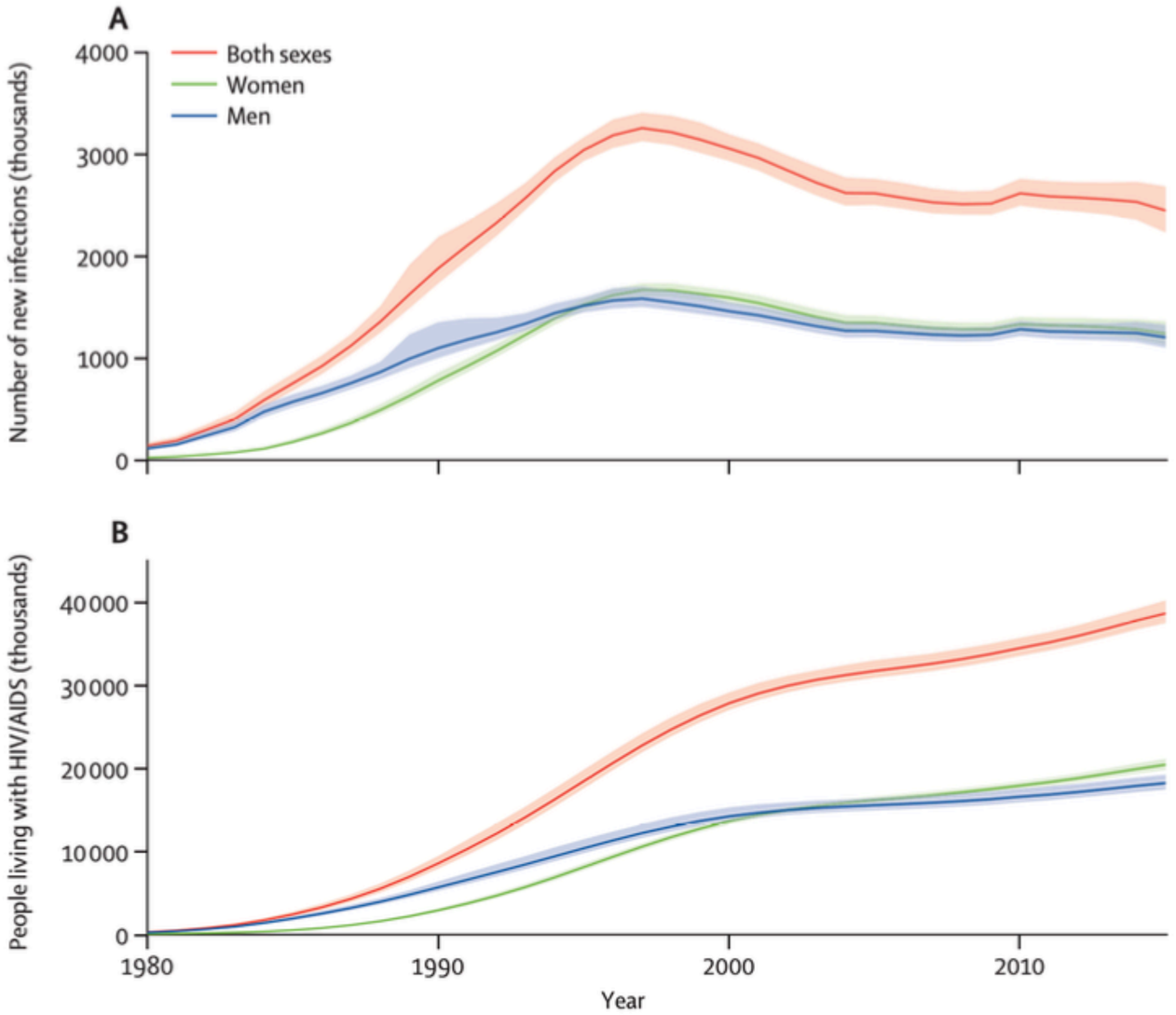


COVID-19 prevalence in Italy



A counterintuitive example

- Incidence and prevalence can provide very different pictures of a disease.
- A medical breakthrough that reduces lethality (people get sick but do not die as often) can increase the prevalence if incidence remains constant
- This was the effect of HIV antiretroviral therapies



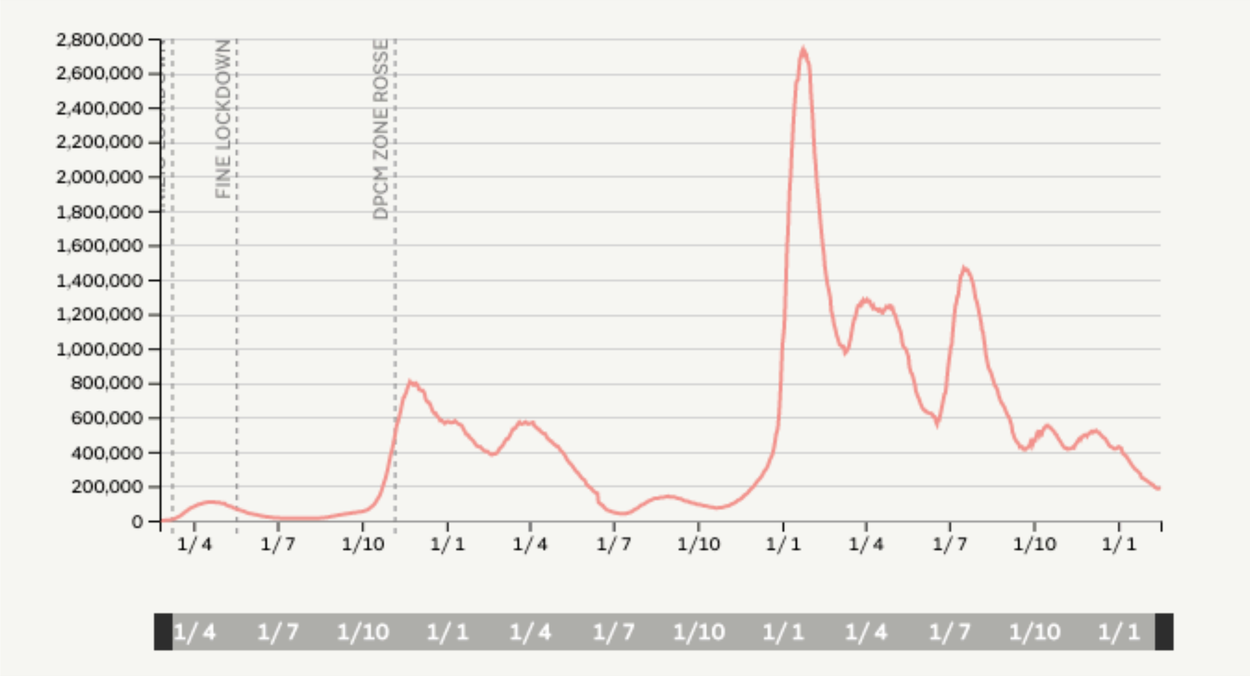
[The Lancet HIV 3.8 (2016): e361-e387.]

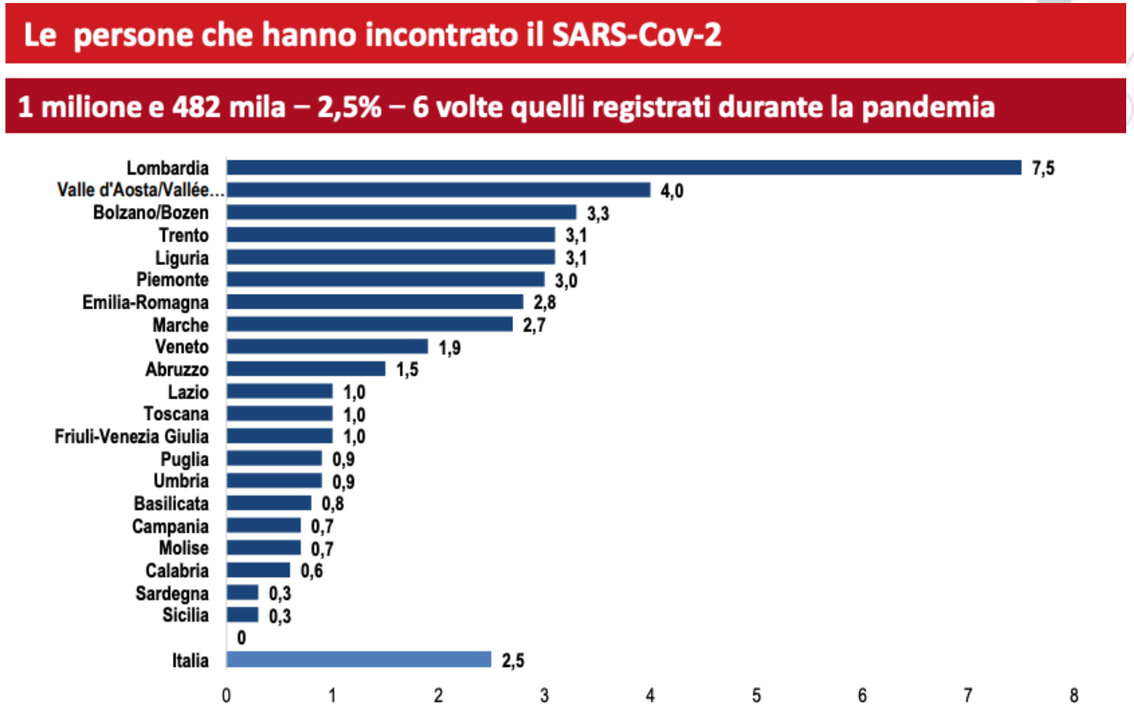
Seroprevalence

- **Seroprevalence** measures the number of infected people, those who have been exposed to the disease, based on antibodies in the blood serum.
- Seroprevalence studies are highly relevant to understand the true number of infected individuals in a population, and compare it with the reported numbers: cases are usually **under-reported.**
- Seroprevalence studies provide information on the level of immunity in the population.

Seroprevalence

Reported COVID-19 prevalence in Italy





Results of a seroprevalence study conducted by ISS in Italy between May-July 2020

Cases were underreported by a factor 6

Case definition

- **A case definition** is a set of criteria that defines a case for a given health condition.
- Case definitions may be adjusted locally and they may change over time.
- In general, given the limited local knowledge we may have about a case, case definitions define 3 sets of criteria for:
  - Suspected case
  - Probable case
  - Confirmed case
- The COVID-19 pandemic provides us with a perfect example. At the beginning of the outbreak case definition was quite unspecific. As we developed specific tests, case definitions became more precise.

Case definition

Suspected case of SARS-CoV-2 infection (3 options)

A A person who meets the clinical **OR** epidemiological criteria:

Clinical criteria:

- acute onset of fever AND cough (ILI)

OR

- acute onset of **ANY THREE OR MORE** of the following signs or symptoms: fever, cough, general weakness/fatigue[1], headache, myalgia, sore throat, coryza, dyspnoea, nausea/diarrhoea/anorexia

OR

Epidemiological criteria [2] :

- contact of a probable or confirmed case, or linked to a **COVID-19 cluster.**[3]

B A patient with **severe acute respiratory illness**
(SARI: acute respiratory infection with history of fever or measured fever of ≥38 °C; and cough; with onset within the last 10 days; and requires hospitalization)

C A person
with no clinical signs or symptoms **OR** meeting epidemiologic criteria
with a **positive professional-use or self-test** SARS-CoV-2 Antigen-RDT.[4]

Case definition

Suspected case of SARS-CoV-2 infection (3 options)

A A person who meets the clinical **OR** epidemiological criteria:

Clinical criteria:

- acute onset of fever AND cough (ILI)

OR

- acute onset of **ANY THREE OR MORE** of the following signs or symptoms: fever, cough, general weakness/fatigue[1], headache, myalgia, sore throat, coryza, dyspnoea, nausea/diarrhoea/anorexia

OR

Epidemiological criteria [2] :

- contact of a probable or confirmed case, or linked to a **COVID-19 cluster.**[3]

B A patient with **severe acute respiratory illness**
(SARI: acute respiratory infection with history of fever or measured fever of ≥38 °C; and cough; with onset within the last 10 days; and requires hospitalization)

C A person
with no clinical signs or symptoms **OR** meeting epidemiologic criteria
with a **positive professional-use or self-test** SARS-CoV-2 Antigen-RDT.[4]

Probable case of SARS-CoV-2 infection (2 options)

A A patient who meets **clinical criteria AND** is a **contact of a probable or confirmed case**, or linked to a **COVID-19 cluster**[3]

B **Death**, not otherwise explained, in an adult with **respiratory distress** preceding death AND who **was a contact of a probable or confirmed case** or linked to a **COVID-19 cluster**[3]

Confirmed case of SARS-CoV-2 infection (2 options)

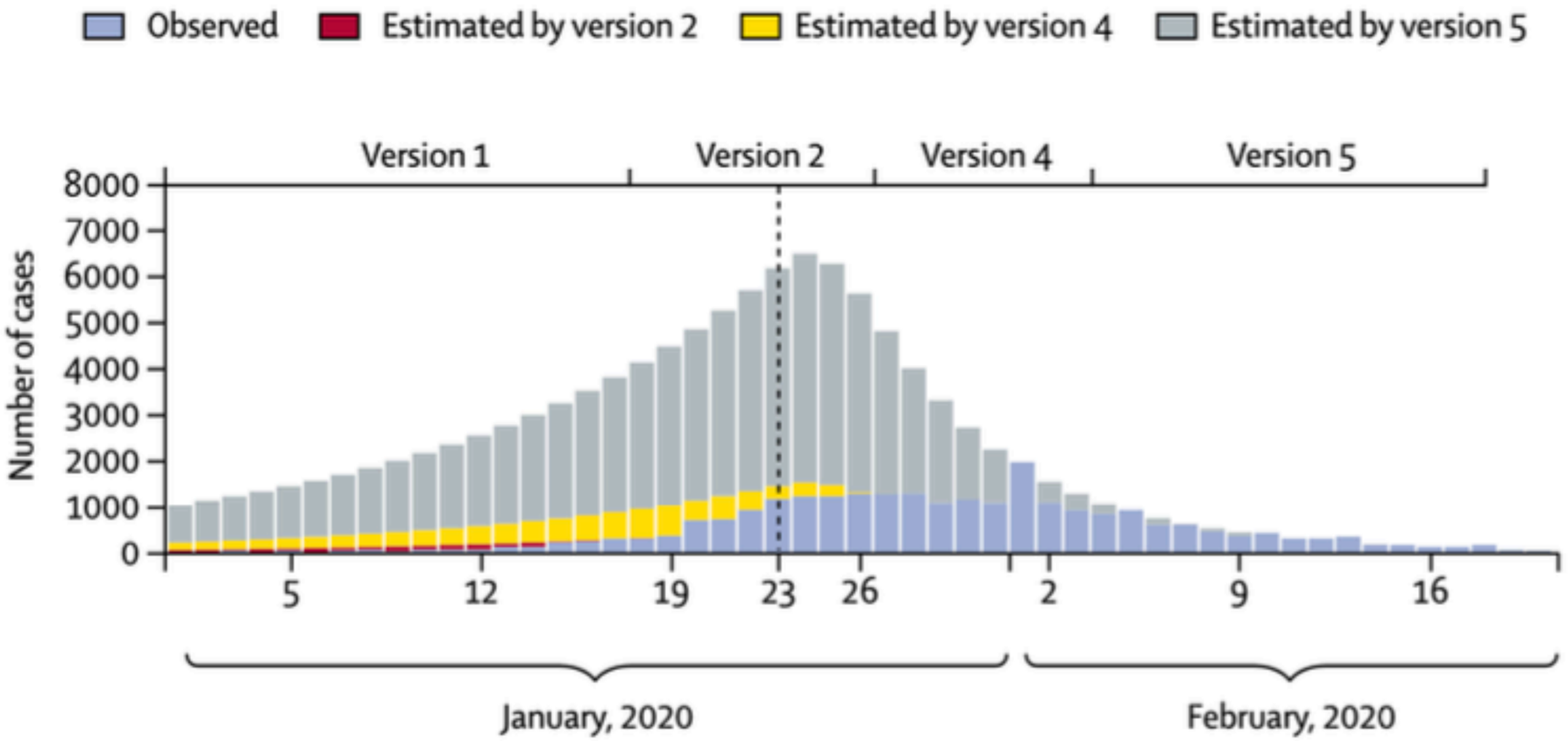
A A person with a positive **Nucleic Acid Amplification Test (NAAT), regardless of** clinical criteria **OR** epidemiological criteria

B A person
meeting clinical criteria **AND/OR** epidemiological criteria (suspect case A)
with a **positive professional-use or self- test** SARS-CoV-2 Antigen-RDT.[4]

Case definition

- In the winter of 2020, the National Health Commission in China issued seven versions of the case definition for COVID-19 within seven weeks.
- In each of the changes, the case definition broadened.
- China reported 55,508 cases by February 20, 2020, but had - hypothetically of course - the latest case definition been applied since the very beginning, the number of cases by February 20 would have been more than 4 times higher.

Case definition



Tsang, Tim K., et al. "Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study." The Lancet Public Health 5.5 (2020): e289-e296.

Mortality, lethality, fatality

- **How deadly is the disease?**
- **Mortality** measures how deadly a disease is in a population.
- **Lethality** measures how deadly a disease is for a single individual who has the disease.

Mortality rate

- **Mortality** is often quantified with a **mortality rate,** which is the number of people dying of a given disease, divided by the population of interest, in a specified period of time.

Mortality rate

- **Mortality** is often quantified with a **mortality rate,** which is the number of people dying of a given disease, divided by the population of interest, in a specified period of time.

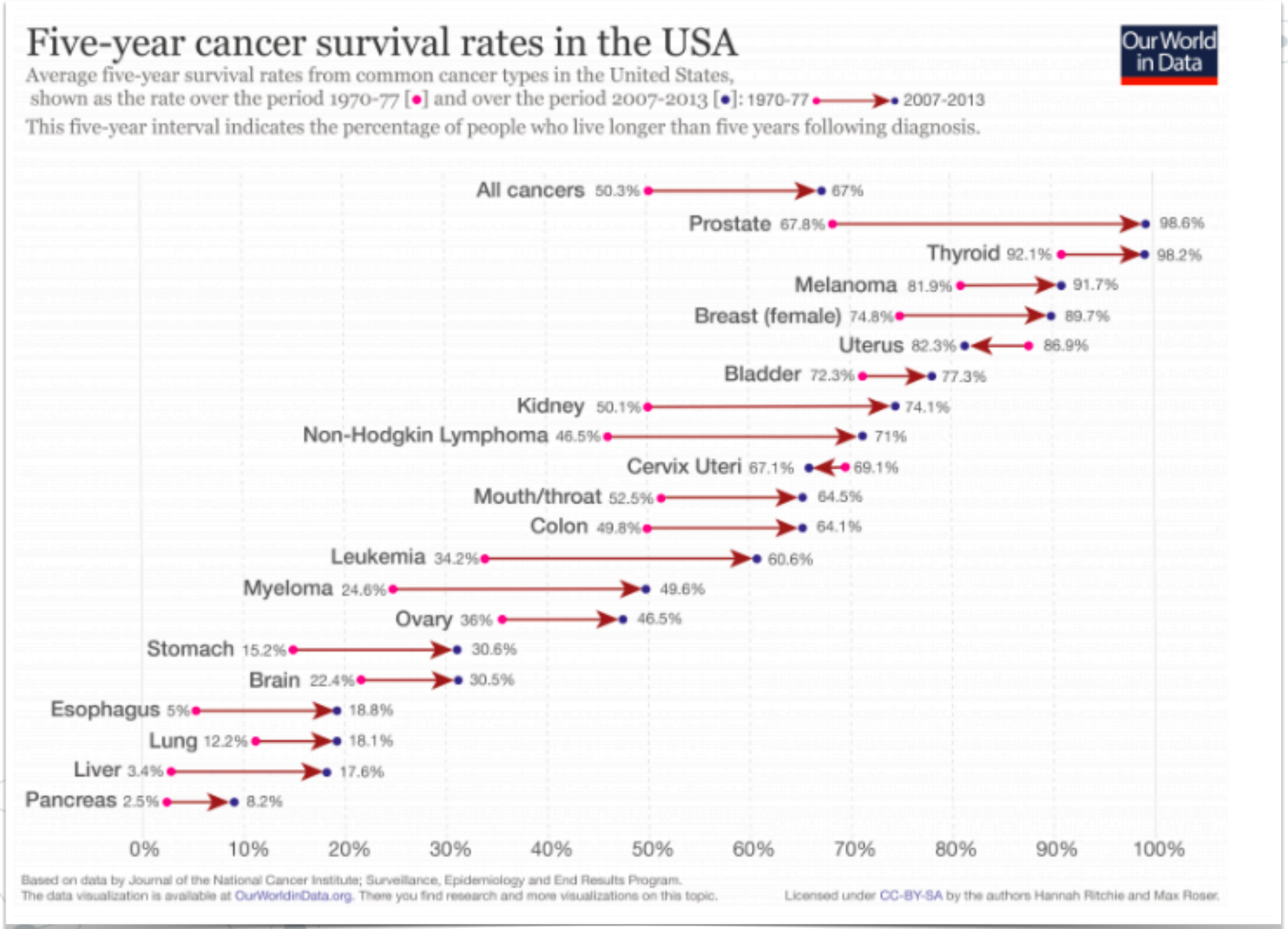
Example: what is the mortality rate of cancer?

- In 2020, 602,350 people died of cancer in the USA.
- Population of the USA: 329.5 million
- Cancer mortality rate: 0.18%
- The general mortality rate (all causes): just above 1%
- Cancer is thus responsible of about 1/5 of all deaths
- Heart disease: 0.21%

Lethality

- **Lethality** asks how deadly a disease is once you have it.
- For **chronic, non-communicable, diseases** we typically use survival rates in a given time period (e.g., 5-years survival rate).
- For acute diseases, the most commons statistics is the **case fatality rate (CFR).**

Survival rates



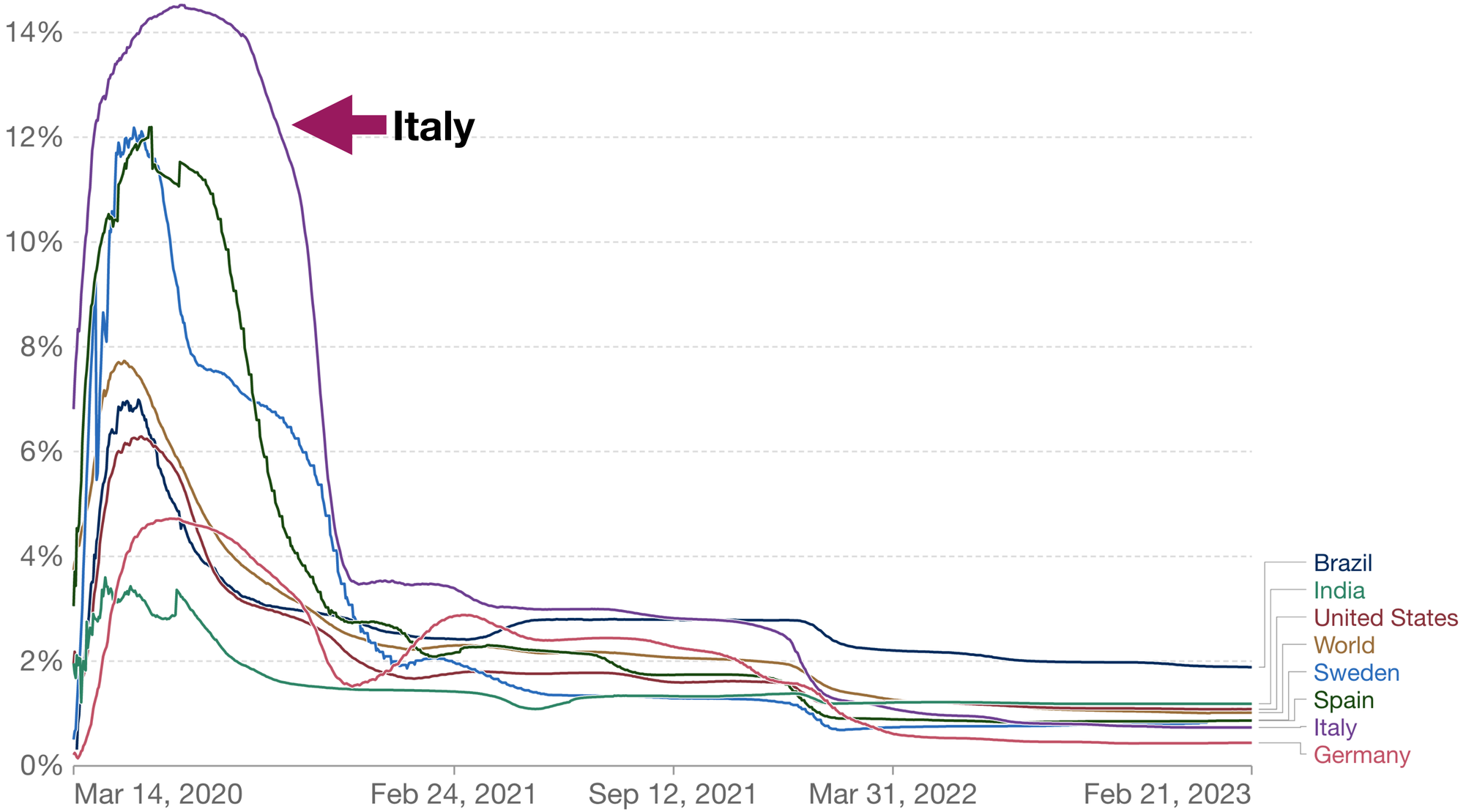
Fatality rates

- **Case fatality rate.** CFR denotes the number of deaths due to a disease, divided by the **confirmed number of cases** who have the disease.
- Issue: the confirmed number of cases can be highly biased.
- **Infection fatality rate.** IFR denotes the number of deaths due to a disease, divided by the **number of all cases** who have the disease.
- Issue: in general we don't know the true number of all cases, so we have to try and estimate it.

Case fatality rate of COVID-19

The case fatality rate (CFR) is the ratio between confirmed deaths and confirmed cases. The CFR can be a poor measure of the mortality risk of the disease. We explain this in detail at OurWorldInData.org/mortality-risk-covid

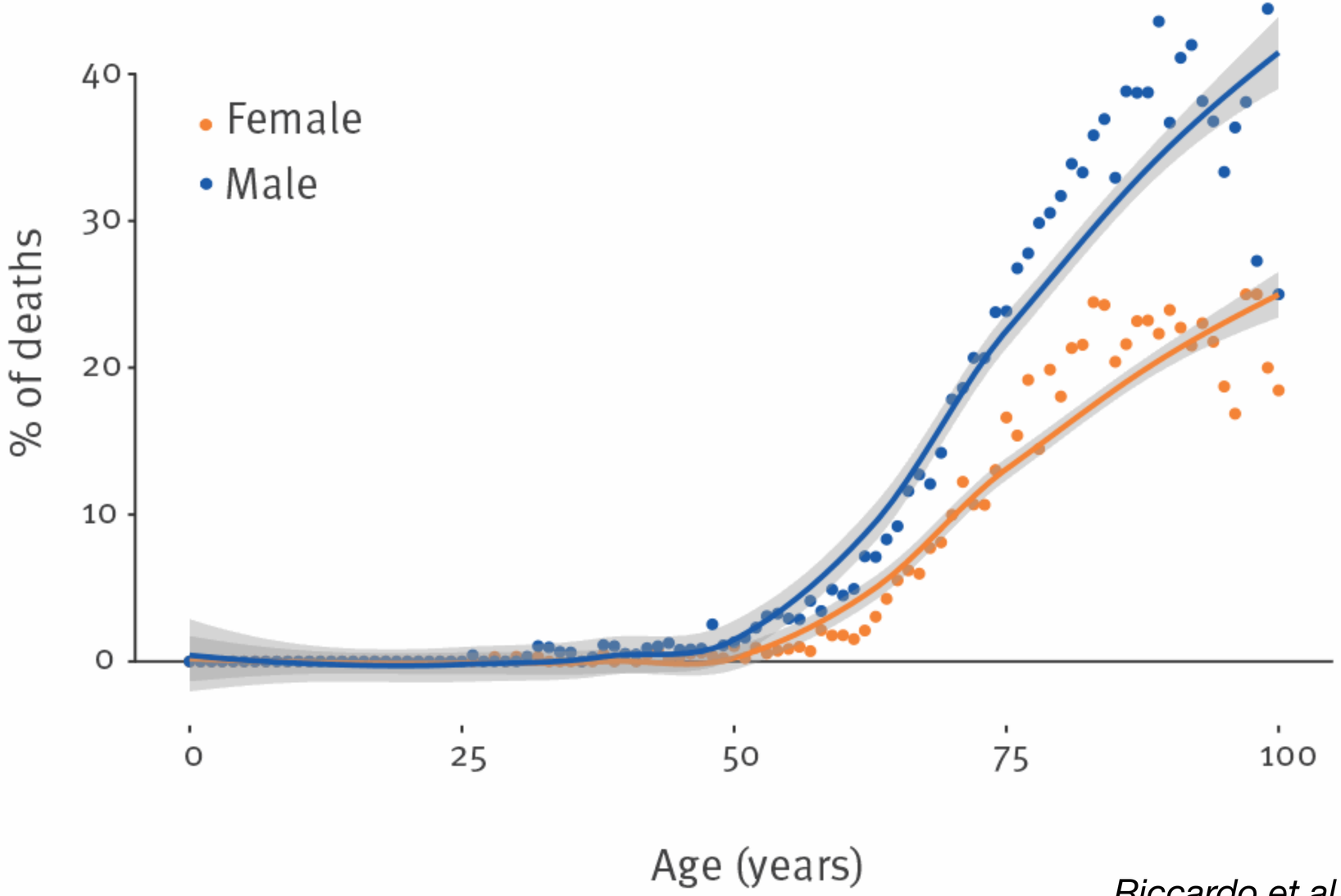
Our World in Data



Case fatality rate

- Due to its variability, the Case Fatality Rate does not reflect the risk of dying from COVID-19 in absolute terms.
- The **CFR reflects the situation in a particular context, at a particular time, in a particular population.**
- The CFR will depend on the available treatment, the patients' conditions, and their demographics (age, sex)
- For COVID-19, we know age plays a fundamental role.

COVID-19 CFR



Riccardo et al. Eurosurveillance 2020

Other diseases

Disease	Estimated case fatality rate (CFR)
SARS-CoV	10% Venkatesh and Memish (2004) Munster et al. (2020)
MERS-CoV	34% Munster et al. (2020)
Seasonal flu (US)	0.1 to 0.2% US CDC
Ebola	50% 40% in the 2013-16 outbreak WHO (2020) Shultz et al. (2016)

IFR

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

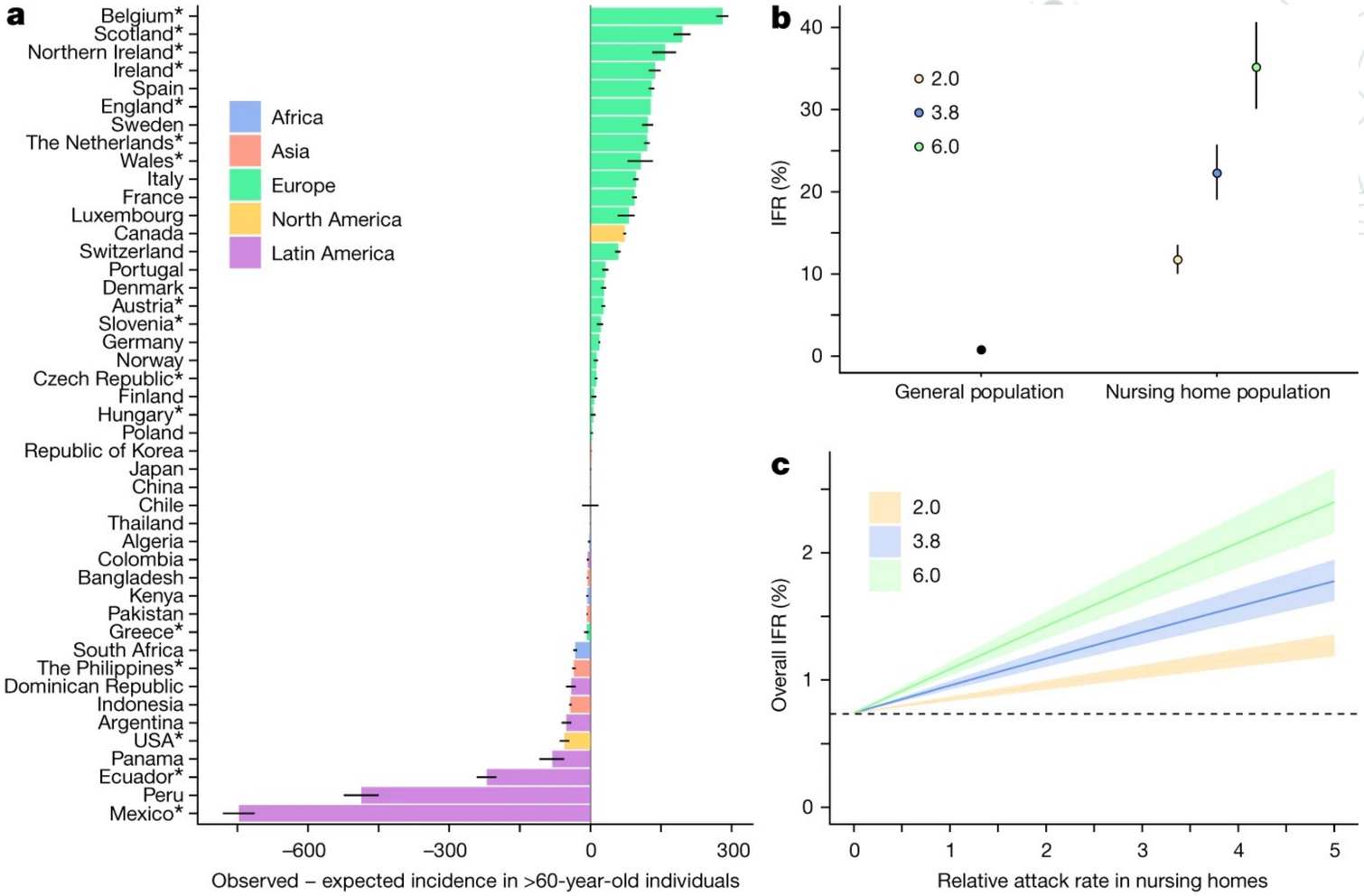
Article | Published: 02 November 2020

Age-specific mortality and immunity patterns of SARS-CoV-2

Megan O'Driscoll ✉, Gabriel Ribeiro Dos Santos, Lin Wang, Derek A. T. Cummings, Andrew S. Azman, Juliette Paireau, Arnaud Fontanet, Simon Cauchemez ✉ & Henrik Salje ✉

Nature **590**, 140–145 (2021) | Cite this article

355k Accesses | **486** Citations | **2722** Altmetric | Metrics



COVID-19 IFR

- Estimates of IFR for COVID-19 vary by country and setting
- According to Imperial College, the overall IFR in a typical low-income country, with a population structure skewed towards younger individuals, to be **0.23%** (0.14-0.42 95% prediction interval range).
- In contrast, in a typical high income country, with a greater concentration of elderly individuals, we estimate the **overall IFR to be 1.15%** (0.78-1.79 95% prediction interval range).
- These numbers may seem small but they are actually extremely high if compared to flu (IFR = 0.01%)
- Moreover, COVID-19 is highly contagious. In a naive population, this results in millions of deaths.

Next… Testing & diagnostics