Statistical Issues in Infectious Diseases

Laura F White

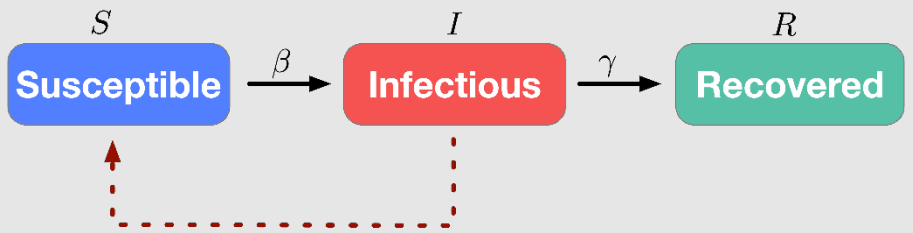
Boston University Department of Biostatistics

March 27, 2023

Outline

- What is a model?
- What is statistics?
- What is a statistical model?
- How is it different from math? Math models?
  - Uncertainty, data-driven
- What are areas of overlap between the two in ID?
  - ID models fit stochastically
  - Model calibration
- Other areas where stats is useful in ID modeling

What is a model?

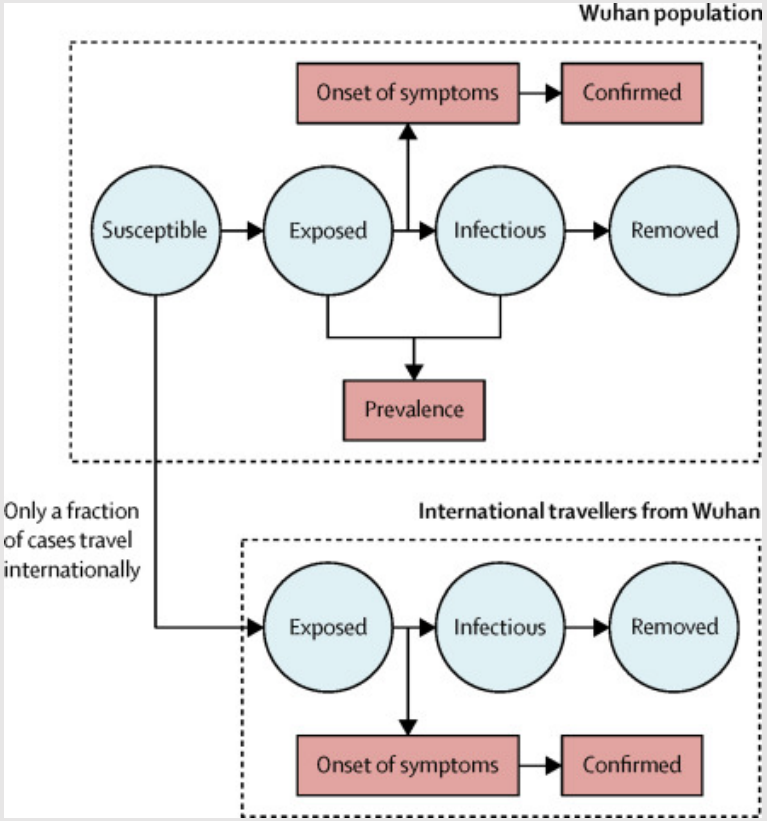


Mathematical models

Describes the relationship between different states in a system.

States are determined by domain knowledge.

Data can be used to inform transitions between the states.



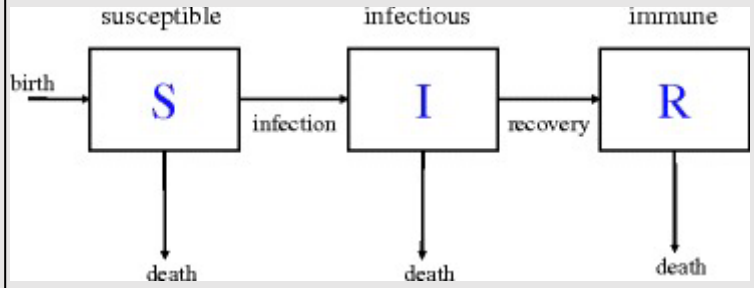
Source: Kucharski et al, 2020

Mathematical models

Inputs

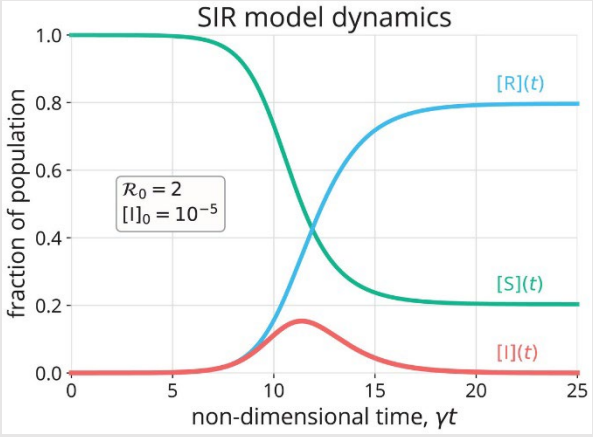
1. States/structure

2. Data to inform transitions between states (can integrate data from many sources)



Outputs

1. Projections
2. "What if" scenarios
3. Insights in dynamics



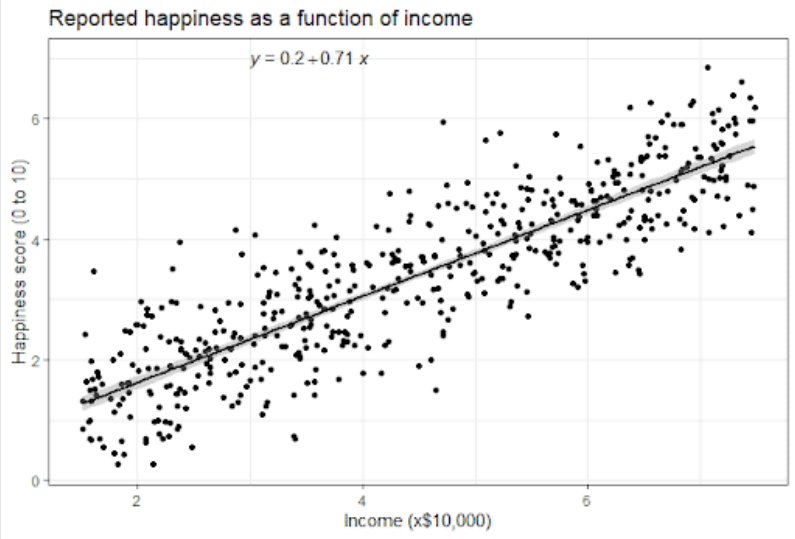
Statistical models

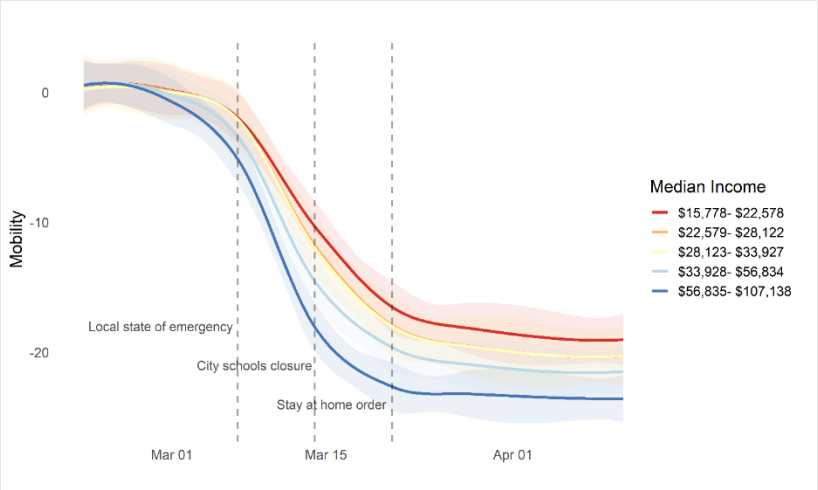
Start with data

Consider a probabilistic model that could lead to the data that is observed

Try to fit to that data

Quantify the errors in the model to describe our uncertainty about the model

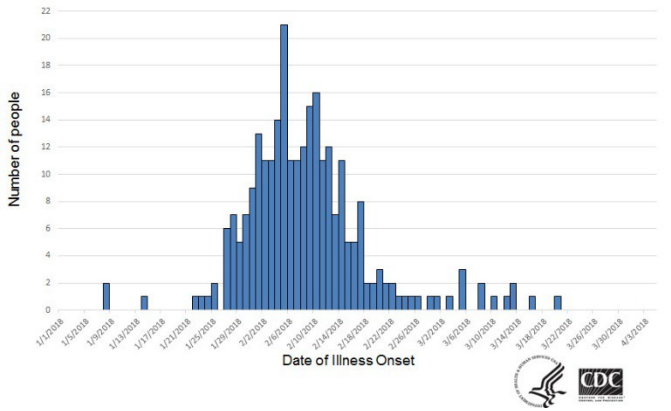




Statistical models

Inputs

1. Data
2. Assumptions about probabilistic process generating the data



$$L(R_i, p|N) = \prod_{i=1}^{T} \frac{\exp(\mu_i)\mu_i^{N_i}}{N_i!},$$

Outputs

1. Estimates of key parameters
2. Hypothesis testing
3. Error bounds

Collision!

Math models and statistical models are not so at odds!

- Model calibration
- Stochastic implementations of models

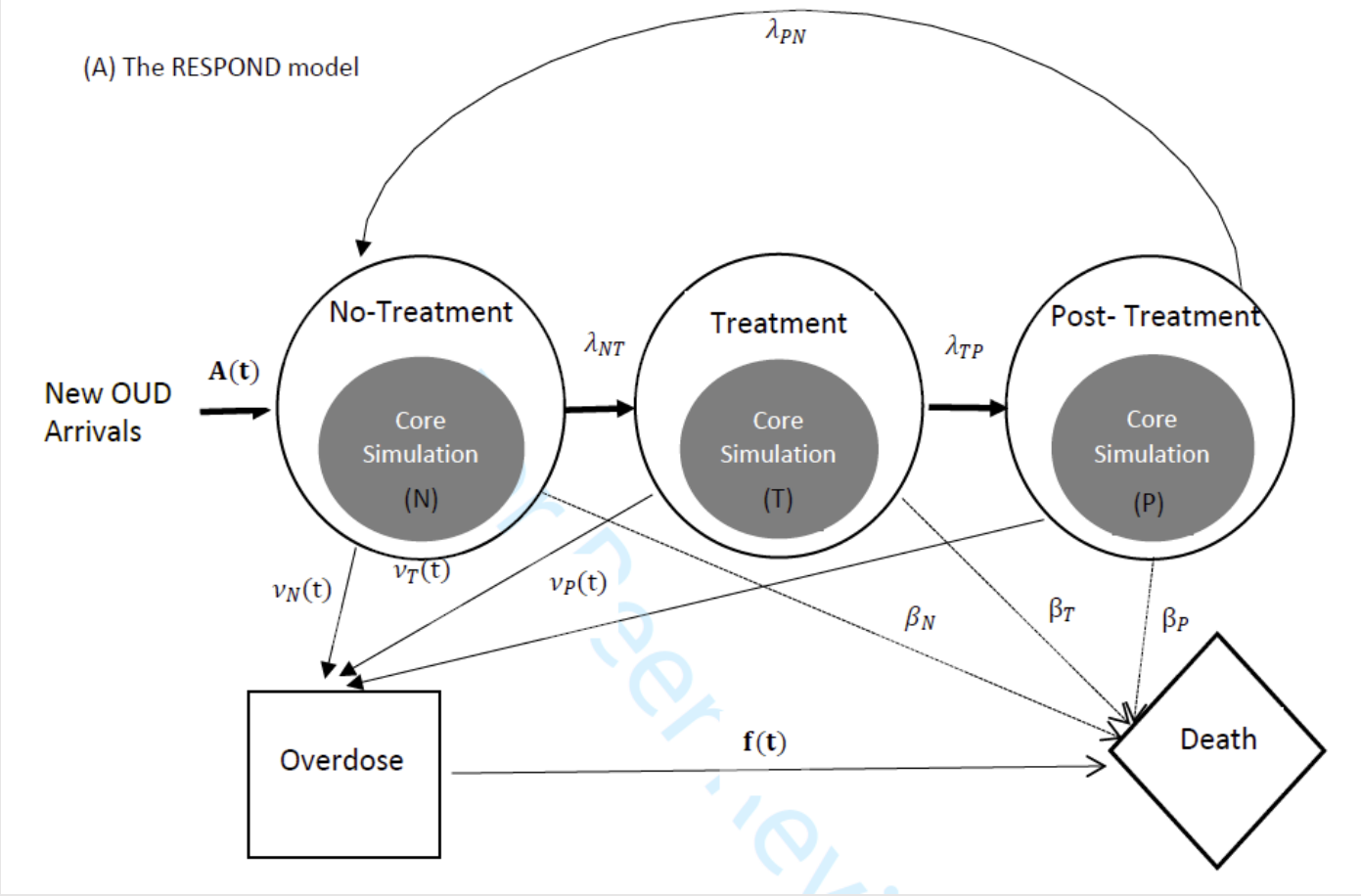
Model Calibration

Start with a math model structure

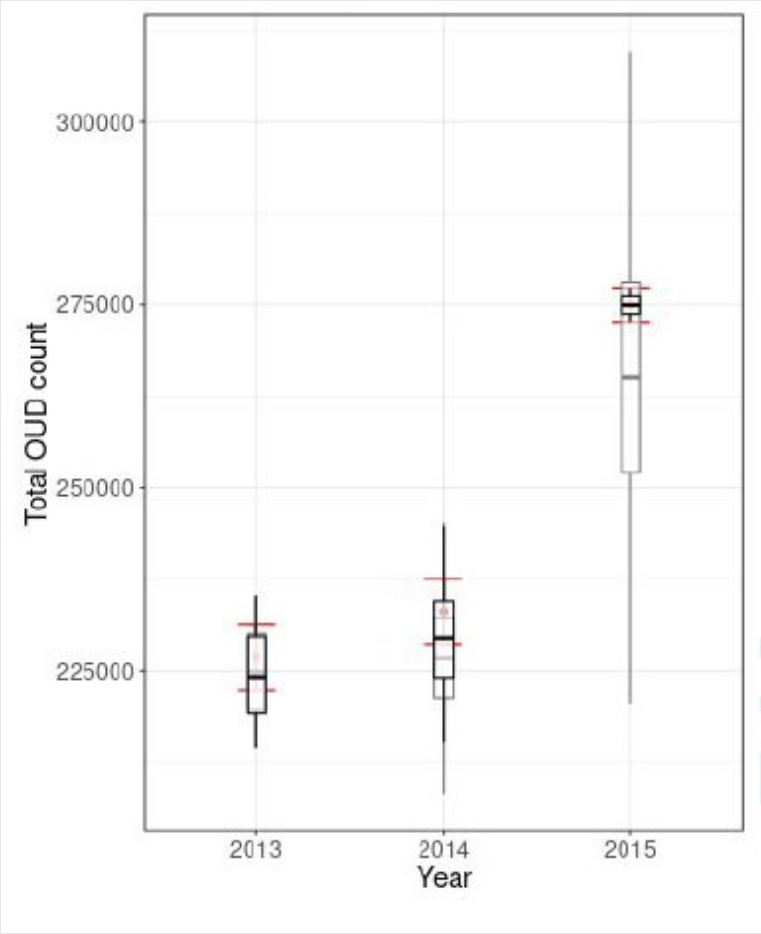
Mix in some data that you want your model to fit to (i.e. targets)

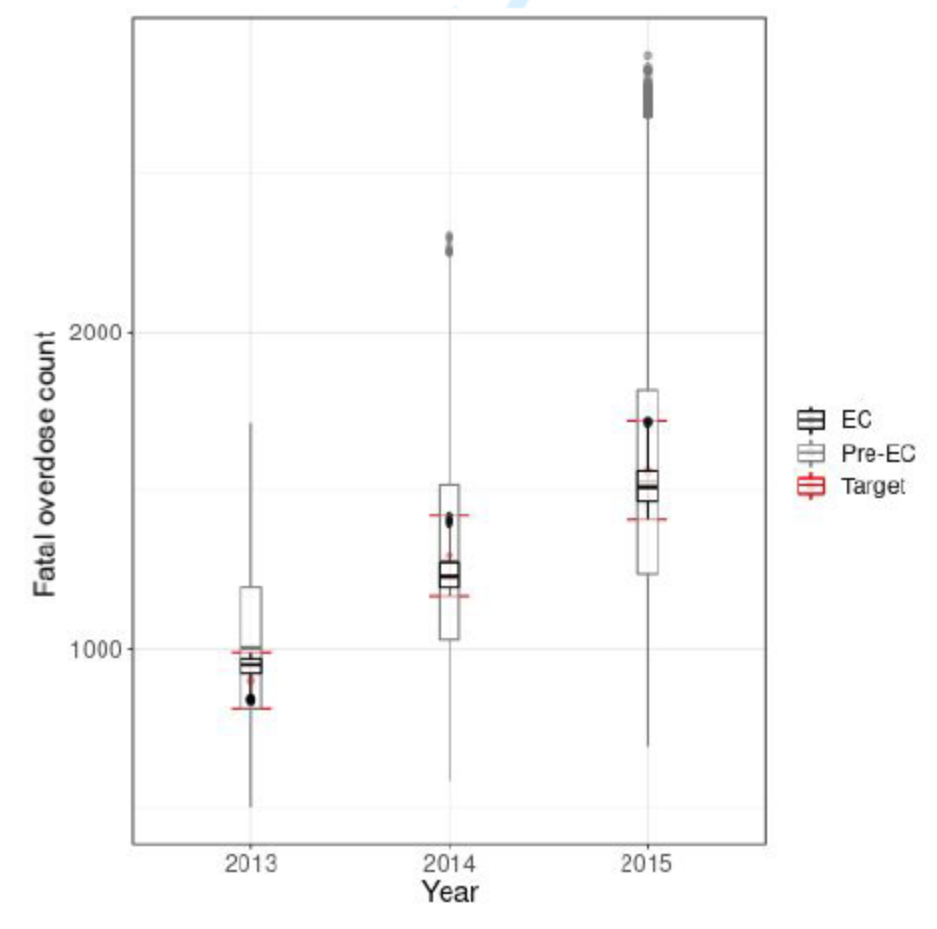
Figure out the input parameters that get you “close” to those targets.

Example


(A) The RESPOND model
λ_{PN}
No-Treatment
Core Simulation (N)
Treatment
Core Simulation (T)
Post- Treatment
Core Simulation (P)
New OUD Arrivals
$\mathbf{A(t)}$
λ_{NT}
λ_{TP}
$\nu_N(t)$
$\nu_T(t)$
$\nu_P(t)$
β_N
β_T
β_P
Overdose
$\mathbf{f(t)}$
Death

Example

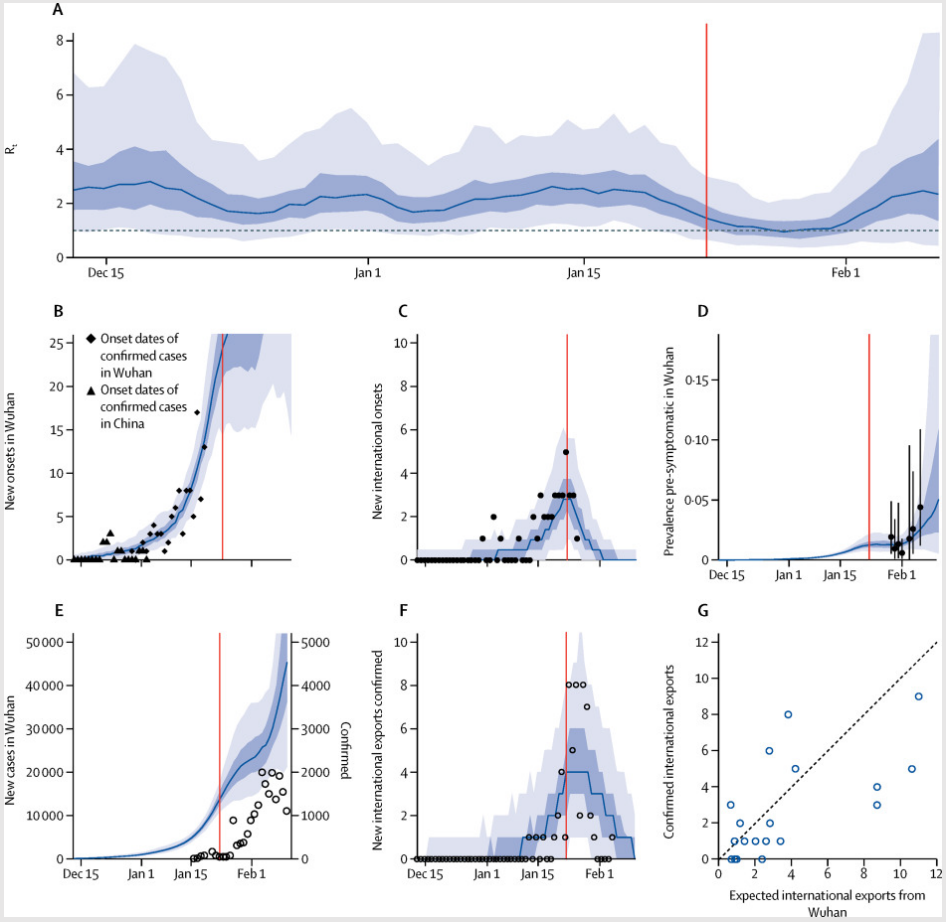

300000
275000
250000
225000
Total OUD count
2013
2014
2015
Year


2000
1000
Fatal overdose count
2013
2014
2015
Year
EC
Pre-EC
Target

Stochastic models

Start with a mathematical model structure

Incorporate uncertainty around the parameter inputs by running the model many times with varying input parameters



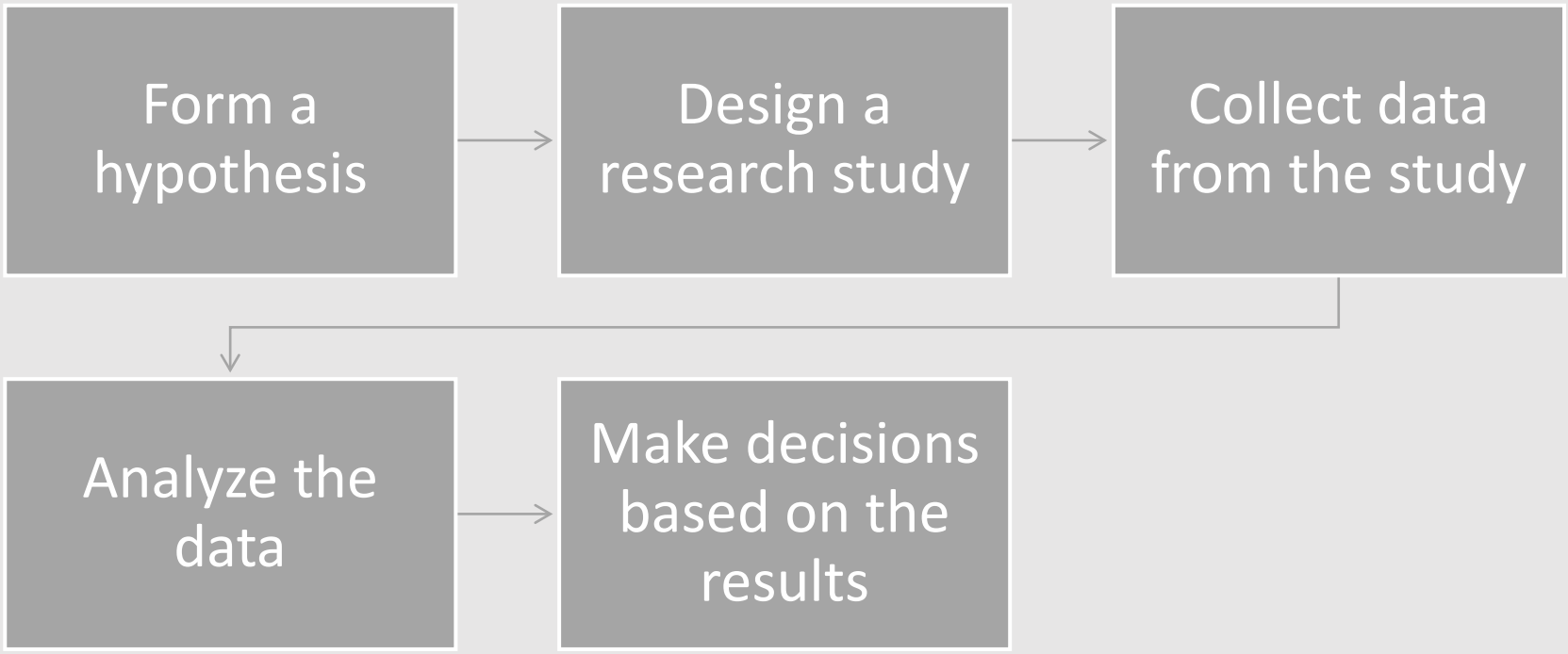
Kucharski et al, 2020

Other statistical topics in infectious diseases

- Efficacy of new therapies
- Vaccine studies
- Outbreak analysis
- Transmission patterns
- Estimates of severity (prevalence, incidence, case fatality ratio)

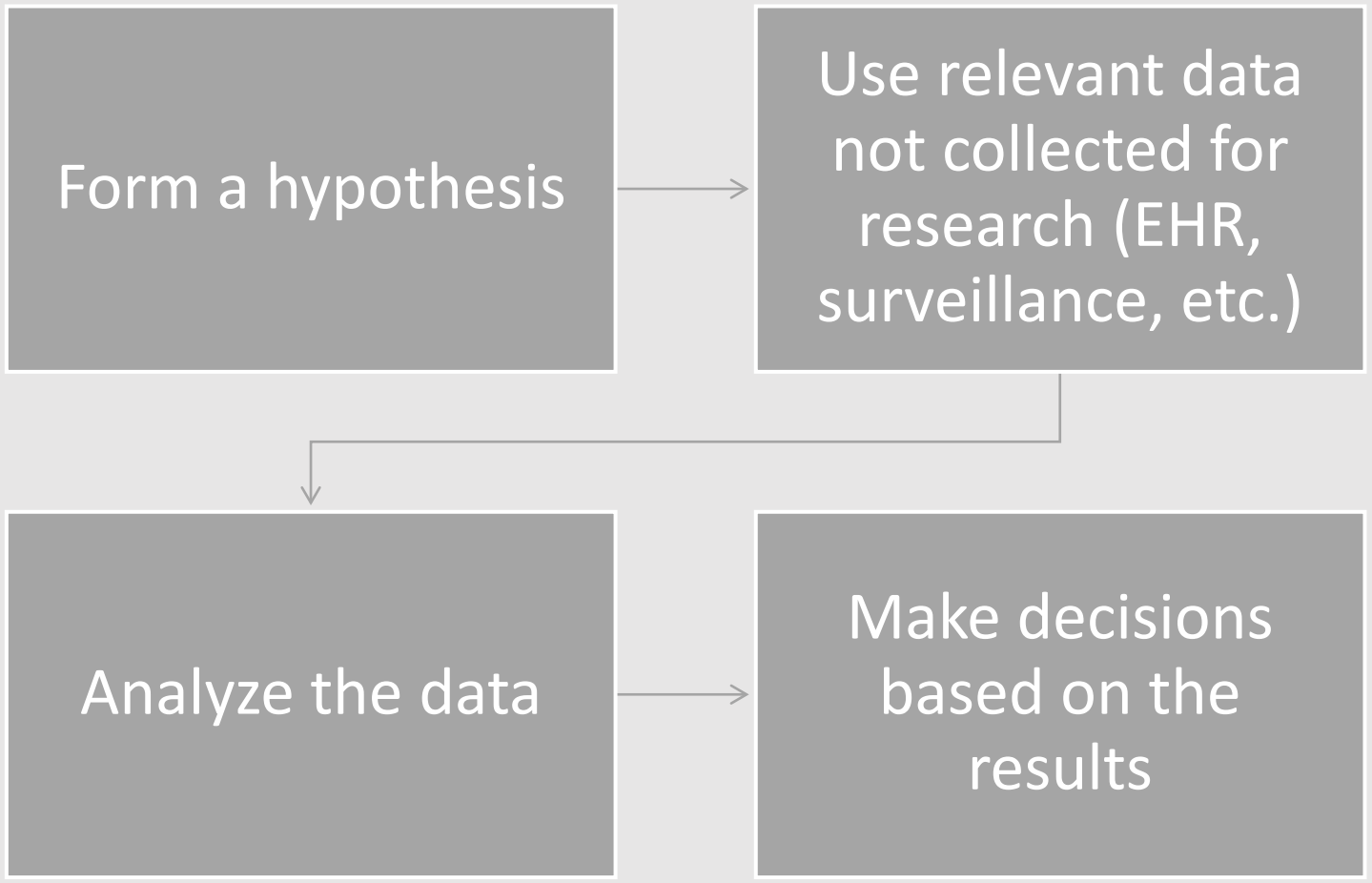
Much more…

A common statistical process (Prospective)



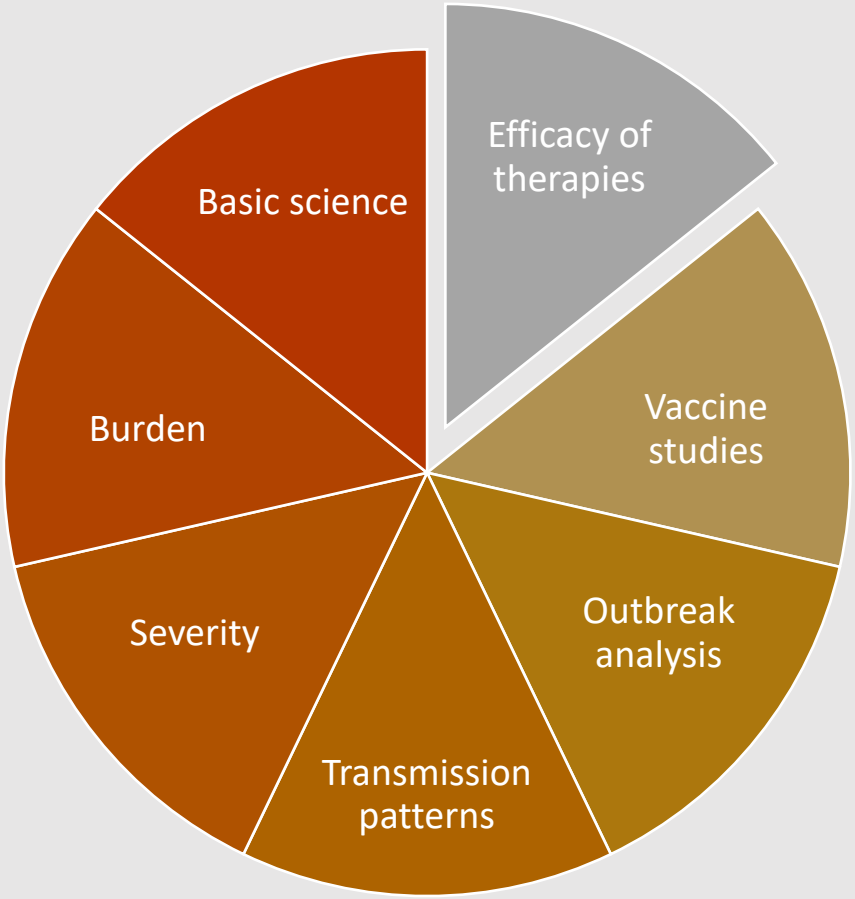
Statistics are involved in all these steps!

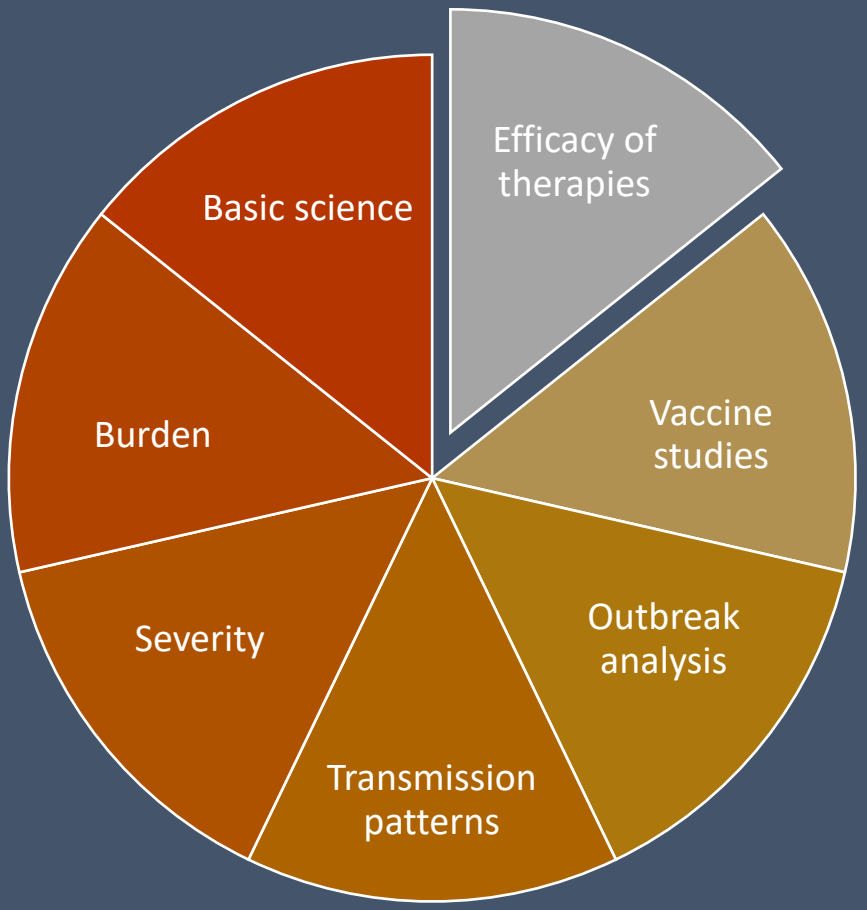
Another common statistical process (Retrospective)



Statistics are involved in all these steps!

Statistics and infectious diseases





Efficacy of new therapies

Infectious disease therapies

- Examples:
  - SARS COVID-19: multiple therapies proposed and approved
  - Ebola 2014: Zmapp received approval to deploy without full approval
  - HIV/AIDS: many drugs used without full testing due to urgency
  - Drug resistance is a chronic problem/challenge
- Some of this is not that different from other diseases, if the therapy is designed to cure disease or provide some other benefit.
- Population impacts are also possible and might be of interest to study.
  - Example: if the therapy will decrease transmission of the disease, what impact might that have at a population level?

Ivermectin to be Analyzed in UK Trial as Possible Covid-19 Treatment

June 26, 2021
Killian Meara

June 25, 2021
3:27 PM EDT
Last Updated 4 days ago

Healthcare & Pharmaceuticals

U.S. pauses distribution of Lilly's COVID-19 antibody combination therapy

June 16, 2021
12:58 PM EDT
Last Updated 13 days ago

Healthcare & Pharma

Regeneron therapy cuts deaths among some hospitalised COVID-19 patients -study

Assessing the efficacy and safety of hydroxychloroquine as outpatient treatment of COVID-19: a randomized controlled trial

Study finds no evidence for remdesivir's benefit in severe COVID-19 patients

Download PDF Copy

TIME SPOTLIGHT STC

By Dr. Liji Thomas, MD
Reviewed by Benedette Cuffari, M.Sc.

Jun 23 2021

HEALTH • COVID-19

Paxlovid Doesn't Work for Healthier Patients, Pfizer Says

Experimental drugs poised for use in Ebola outbreak

International health organizations are in discussions with the Democratic Republic of Congo about how and whether to deploy treatments in addition to a vaccine.

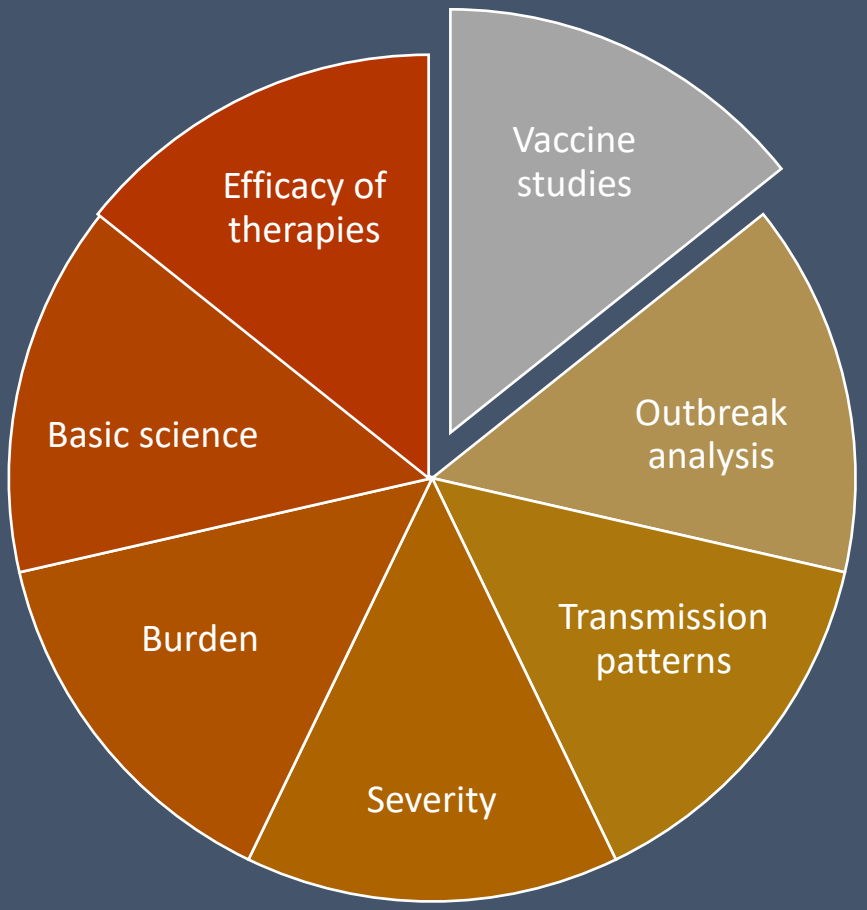
Aid workers responding to the Ebola virus outbreak in the Democratic Republic of Congo (DRC) are seeking approval to treat patients with experimental drugs. These include three potential treatments — ZMapp, favipiravir and GS-5734 — that were given to patients during the 2014–16 Ebola outbreak in West Africa.

The three drugs are being considered in addition to an existing plan to deploy an experimental vaccine; none of the treatments has been definitively proved to lower the risk of death from Ebola.

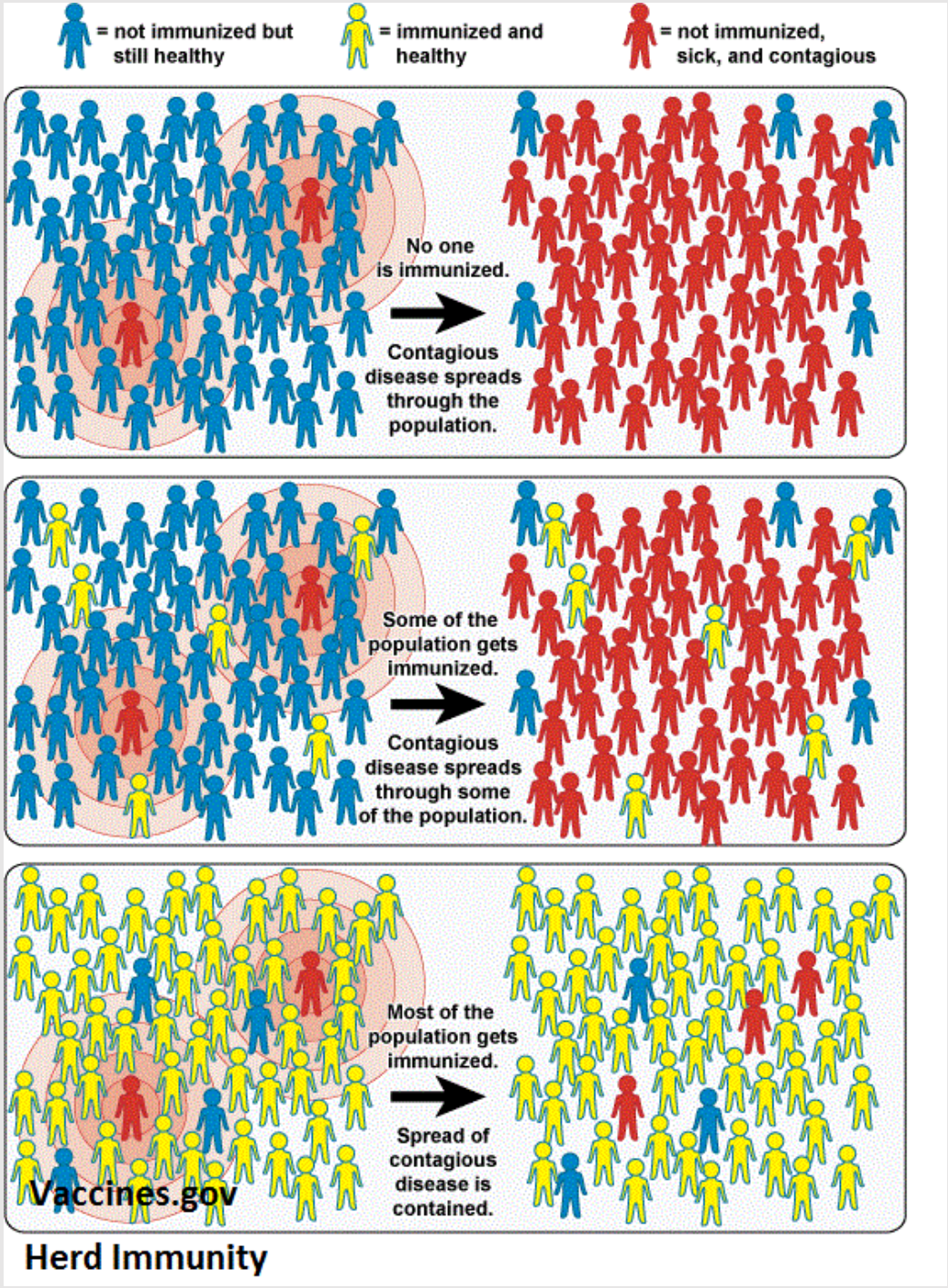
The move to test experimental drugs and vaccines early in the outbreak, which was confirmed on 8 May, is part of a push to start research as soon as possible after Ebola cases are detected in order to save lives. That's a change from the past, when doing research during an outbreak was seen as a distraction.

"In the past our major objective was containment," said Peter Salama, the World Health Organization (WHO) director-general for emergencies, at an 18 May press conference. "One of the paradigm shifts we're seeing in this response is to offer communities a lot more."

Source:
Nature, 18 May 2018



Vaccine studies



Herd Immunity

Vaccine studies

- How can we determine if a vaccine is effective in an ethical way?
- What are some of the challenges for this, compared to testing efficacy of a drug for a non-communicable disease?
- What kind of efficacy is most important for a vaccine?
  - Protection against serious illness, infection, transmission

Vaccine efficacy

There are many ways to measure vaccine efficacy:

$$\frac{\textit{Risk among unvaccinated group} - \textit{risk among vaccinated group}}{\textit{Risk among unvaccinated group}}$$

OR: *1 – risk ratio*

Risk of what?

Death, severe illness, symptomatic illness, infection

Largest CDC COVID-19 Vaccine Effectiveness Study in Health Workers Shows mRNA Vaccines 94% Effective

Geographically diverse population included across 33 sites in 25 U.S. states

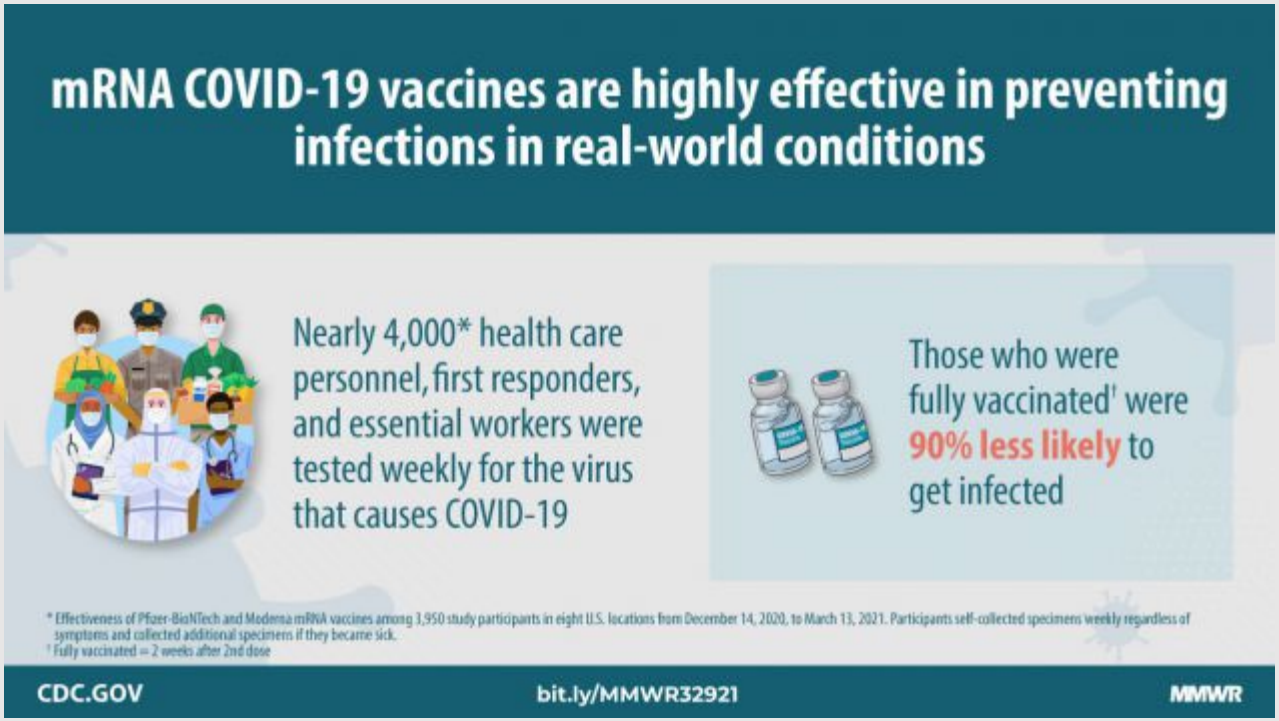
A new CDC study adds to the growing body of real-world evidence (outside of a clinical trial setting) showing that COVID-19 mRNA vaccines authorized by the Food and Drug Administration (FDA) protect health care personnel (HCP) against COVID-19. **mRNA vaccines (Pfizer-BioNTech and Moderna) reduced the risk of getting sick with COVID-19 by 94% among HCP** who were fully vaccinated. This assessment, conducted in a different study network with a larger sample size from across a broader geographic area than in the clinical trials, independently confirms U.S. vaccine effectiveness findings among health care workers that were first reported March 29.

https://www.cdc.gov/media/releases/2021/p0514-covid-19-vaccine-effectiveness.html

Interim Estimates of Vaccine Effectiveness of BNT162b2 and mRNA-1273 COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Health Care Personnel, First Responders, and Other Essential and Frontline Workers — Eight U.S. Locations, December 2020–March 2021

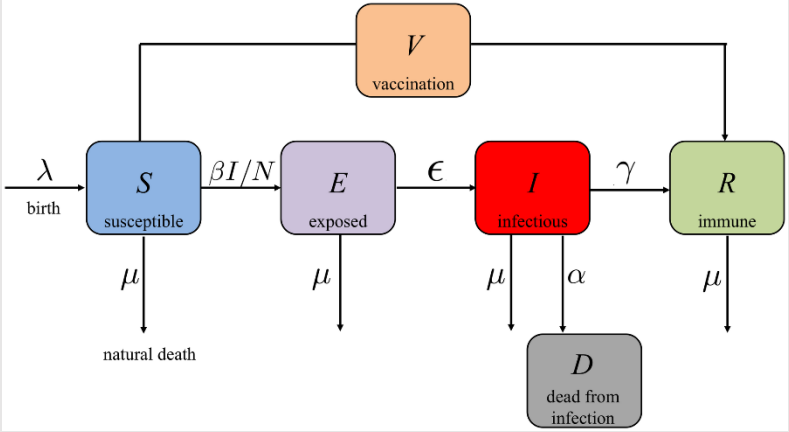
Weekly / April 2, 2021 / 70(13);495–500

On March 29, 2021, this report was posted online as an MMWR *Early Release.*



https://www.cdc.gov/mmwr/volumes/70/wr/mm7013e3.htm

What are the implications of imperfect vaccines?



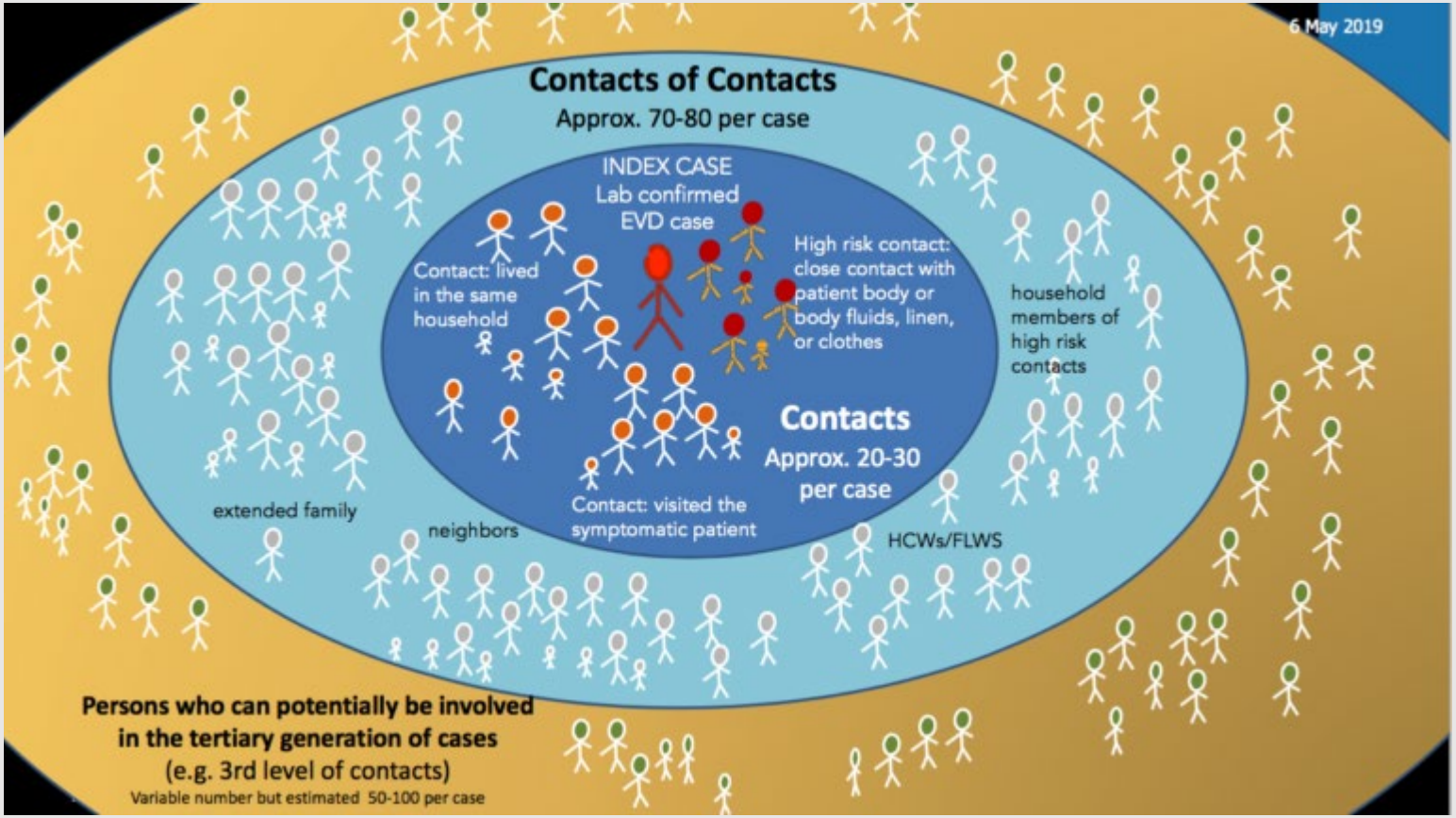
Xu et al, 2021

Important considerations:

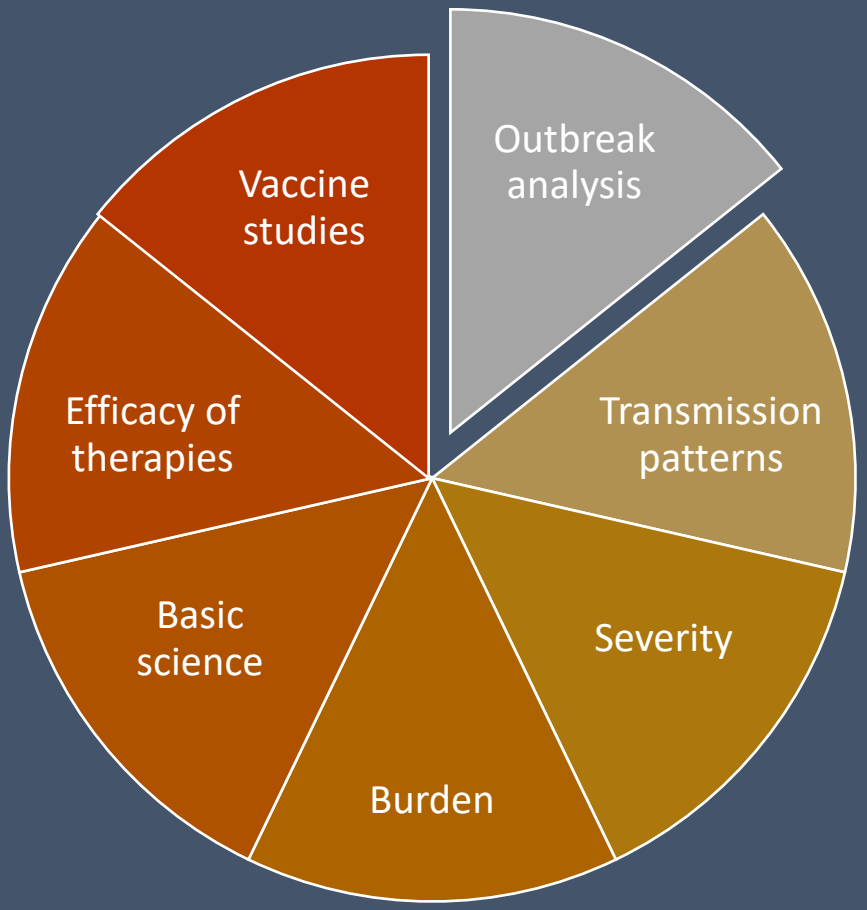
- Selection pressures on the virus from vaccines
- Potential immunity and protection against severe illness from vaccine or prior infection
- Long covid

Models can be helpful in exploring the long-term impact of these issues.

Ring vaccination: Ebola



Source: https://www.statnews.com/2019/05/07/who-broadens-eligibility-ebola-vaccine/



Outbreak analysis

Recent Examples of Outbreaks

- Monkeypox in non-endemic areas
- COVID-19
- DRC Ebola, June 2019-June 2020
- Zika virus, Jan 2016-present
- Ebola in West Africa, March 2014-Dec 2015
- Enterovirus D68, Aug 2014-Dec 2014
- Chikungunya in Americas, Jun 2014-present
- MERS
- Cholera in Haiti, 2010
- Pandemic Influenza H1N1, 2009
- SARS, 2003

https://emergency.cdc.gov/recentincidents/index.asp

What would we want to measure?

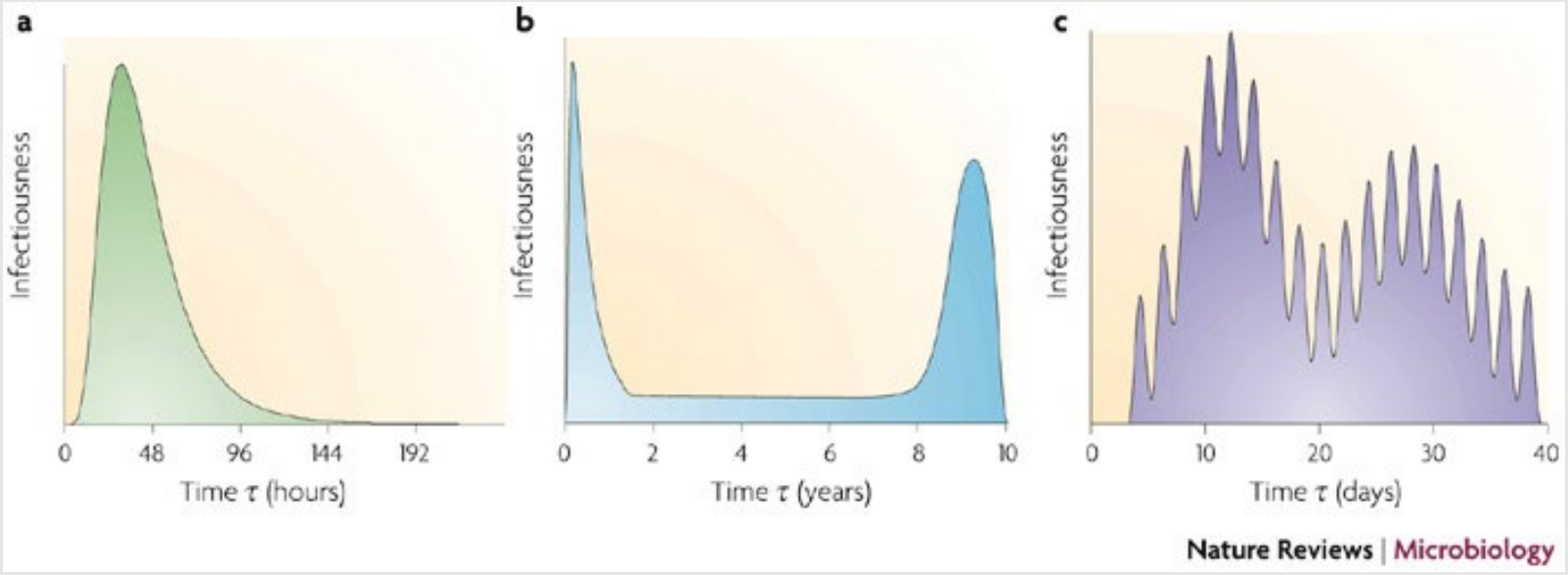
What are aspects of an outbreak that we would want to better understand during an outbreak?

- Reproductive number
- Incubation period
- Infectious period
- Case fatality rate

Quantitative Questions/Issues

- How fast is the disease spreading?
  - Reproductive number
  - Serial interval (incubation distribution, infectious period)
- How severe is the disease?
  - Attack rate
  - Case fatality ratio
- Who is getting infected?
  - Household studies
  - Contact trace studies
- What control measures will help?
  - Modeling exercises
- What is coming next?
  - Forecasting

Biological Infectiousness through time



a : Influenza A. **b** : HIV-1. **c** : Malaria.

Source: Grassly and Fraser (2008)

SARS

Background of the outbreak

- Nov. 2002: First cases in Guangdong Province, China
- Feb 2003: Doctor travels to Hong Kong and infects a large number of people; he dies in hospital there
- Began to travel around the globe
- Mar 12: WHO issues global alert about new respiratory disease
- Mar 15: WHO issues heightened global alert
- April: alerts begin to be lifted in some countries
- July: WHO declares that SARS is contained

Impact

According to the WHO:

- 8098 cases documented worldwide
- 774 deaths documented (CFR ~ 9.6%)

Cases documented in 29 countries

Largest impact in Asia and Canada (Toronto)

Only 8 lab confirmed cases in the US

Epidemiological characteristics

Many super-spreading events

Reproductive number similar to Influenza

Long serial interval

No known asymptomatic transmissions

Impact

Many studies launched to try to understand SARS

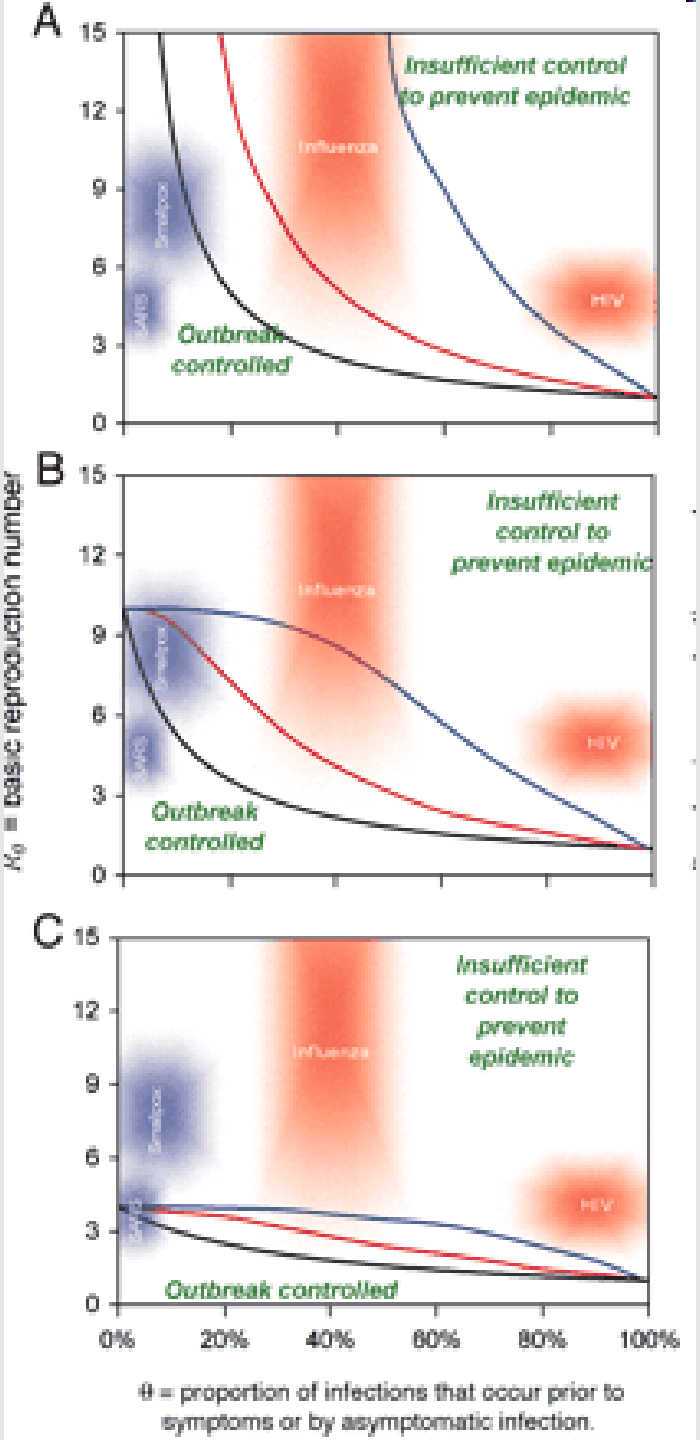
Wallinga and Teunis (AJE, 2004) publish a new technique for estimating the reproductive number

Spurs a lot of new research into new techniques for estimation of key epi parameters

Black lines: Isolate symptomatic individuals

Red: Contact tracing

Blue: Isolation of contacts

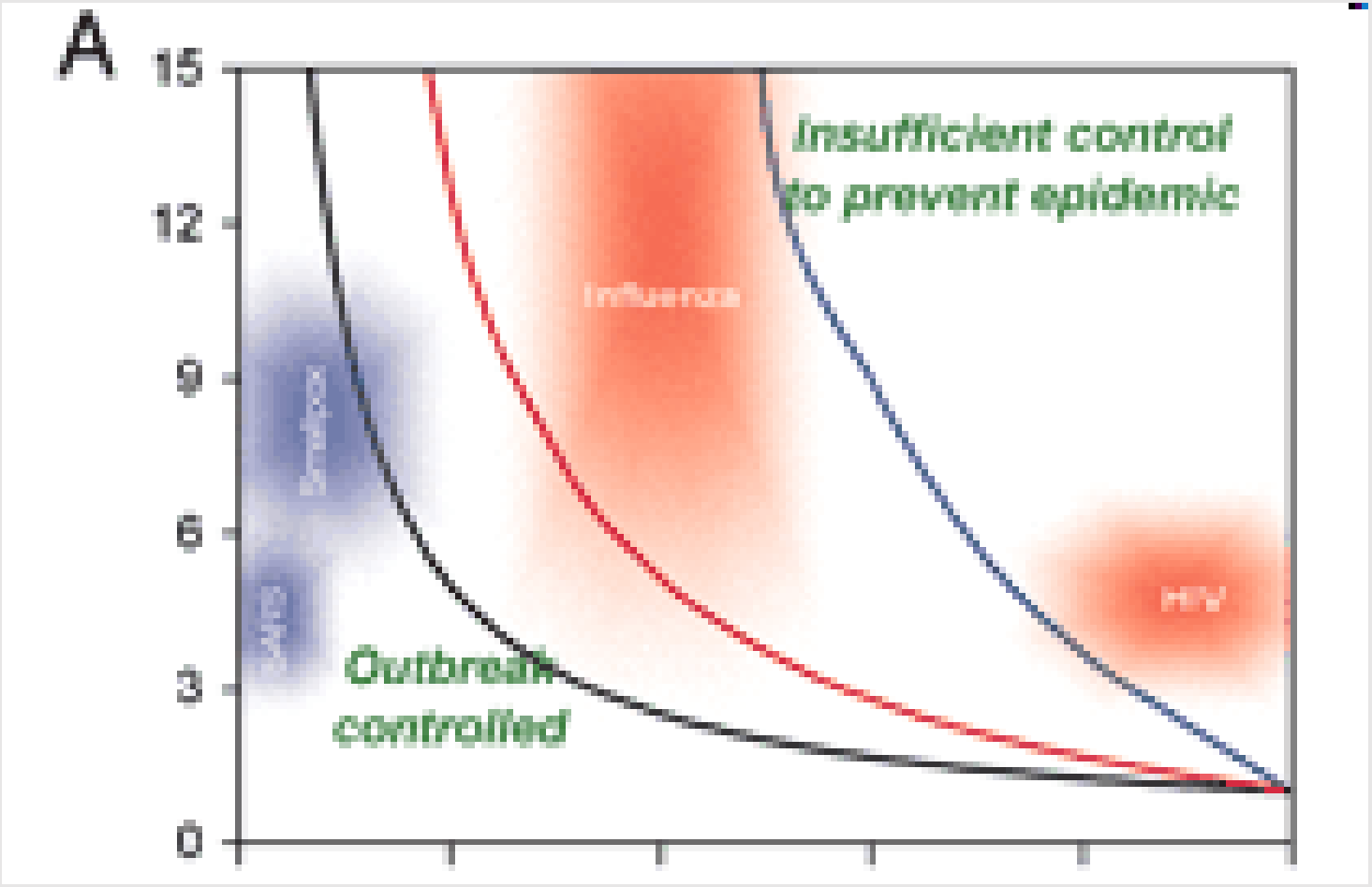


Isolation 100% effective

Isolation 90% effective

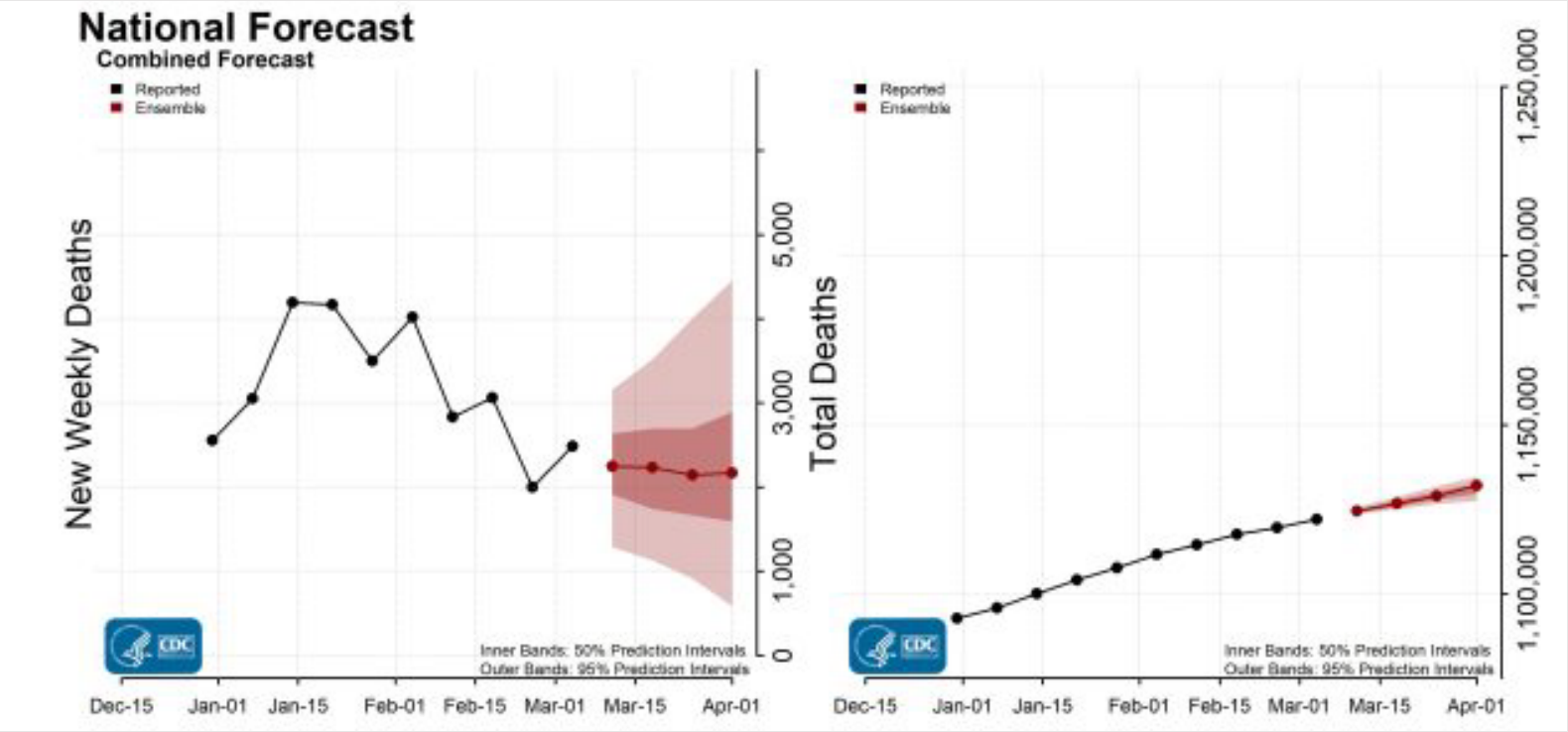
Isolation 75% effective

Source: Fraser (PNAS, 2004)

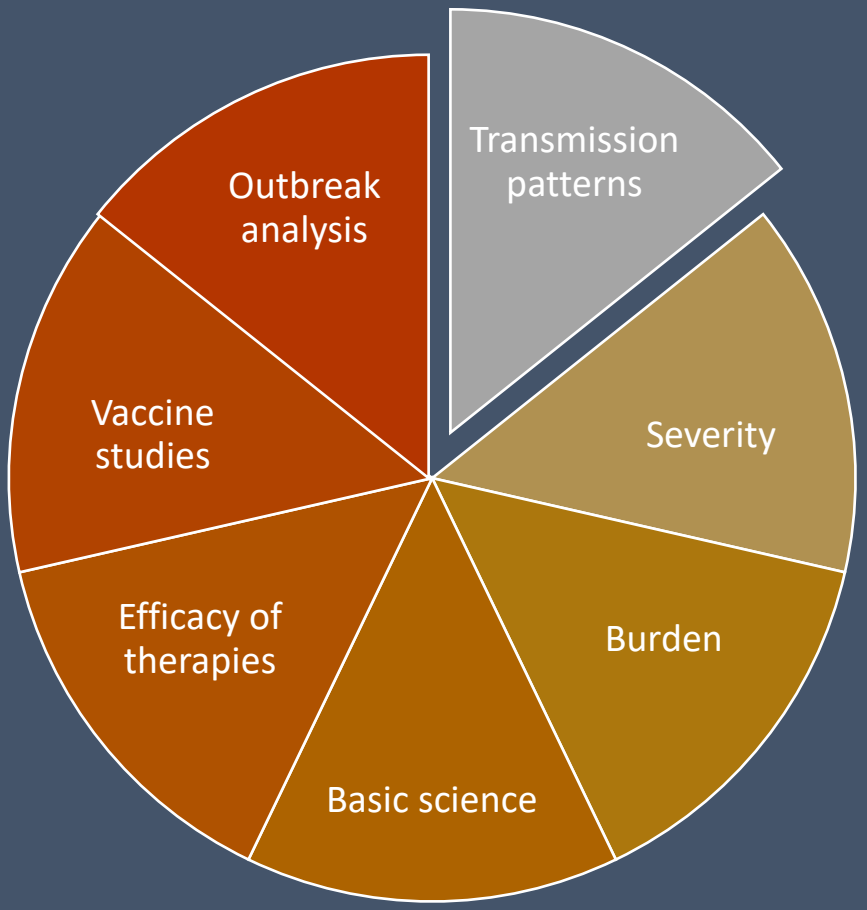

A
15
12
9
6
3
0
Insufficient control
to prevent epidemic
Influenza
Smallpox
SARS
Outbreak
controlled
HIV

Forecasting

Increase in ensemble models to forecast what is coming next.



https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/forecasting-us.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcovid-data%2Fforecasting-us.html



Transmission patterns

Important Intervals

Serial Interval: interval of time between *symptom onset* in primary and secondary cases.

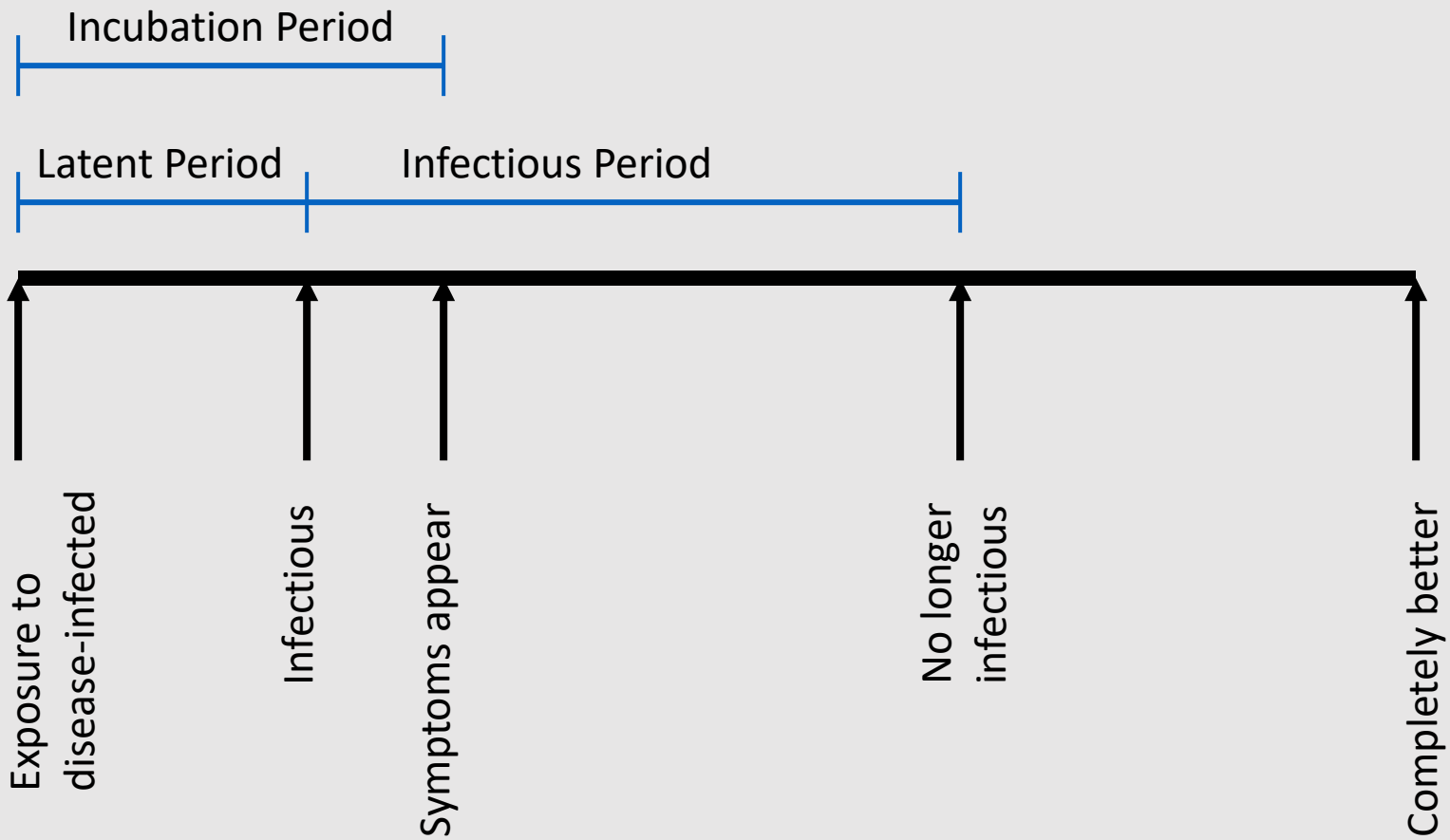
Generation Interval: interval of time between *infection* of primary and secondary cases.

Latent Interval: Time between infection and becoming infectious

Incubation Period: Time between infection and development of symptoms

Infectious Interval: Period when capable of infecting other susceptible individuals

Terminology



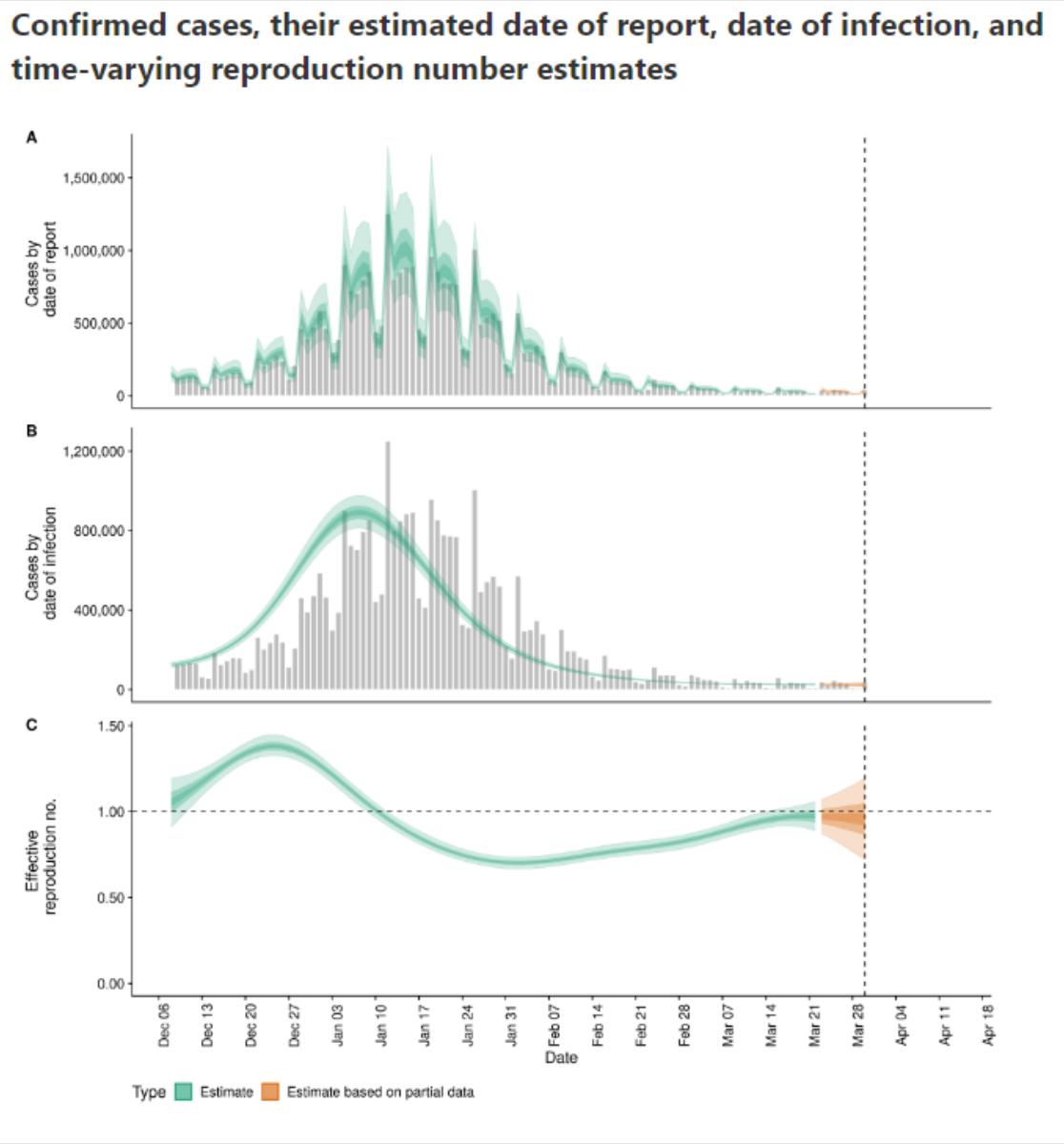
Reproductive number

- **Basic reproductive number (R_0):** average number infected by infected individual in an entirely susceptible population
- **Effective reproductive number (R_t):** average number infected by infected individual accounting for control measures, depletion of susceptible individuals, etc.

Application

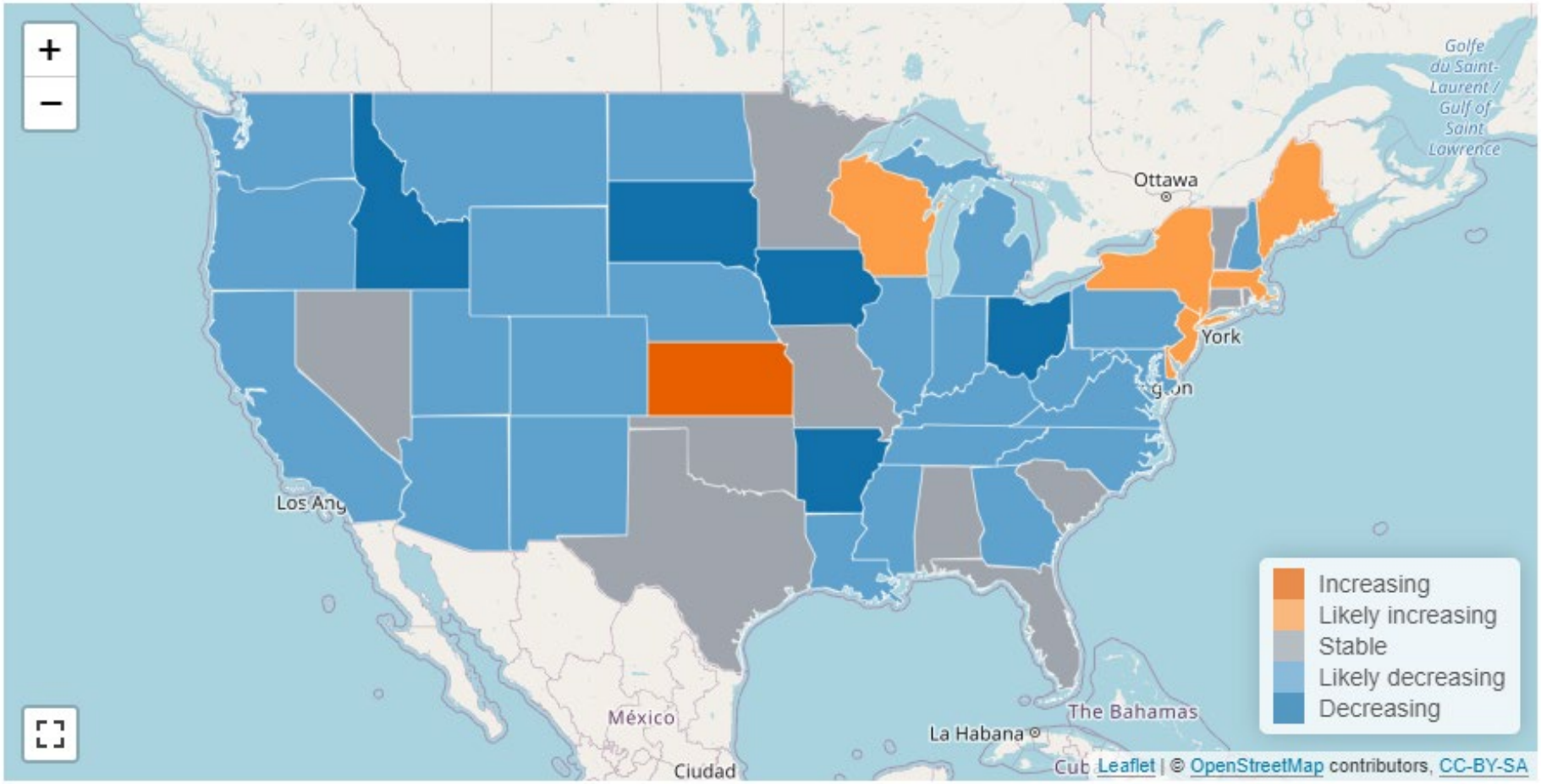
During COVID-19 several groups have published updated R_t estimates by country and state.

How is this done?



https://epiforecasts.io/covid/posts/national/united-states/

Subnational breakdown



State	New confirmed cases by infection date	Expected change in daily cases	Effective reproduction no.	Rate of growth	Doubling/halving time (days)
Alabama	366 (73 -- 2748)	Stable	1 (0.58 -- 1.9)	0.006 (-0.12 -- 0.24)	120 (2.9 -- -5.6)
Alaska	60 (4 -- 1024)	Likely decreasing	0.77 (0.28 -- 1.7)	-0.065 (-0.23 -- 0.18)	-11 (3.9 -- -3)
American Samoa	1339 (261 -- 7930)	Likely increasing	1.5 (0.99 -- 2.1)	0.14 (-0.0034 -- 0.27)	5.1 (2.5 -- -210)
Arizona	14 (2 -- 132)	Likely decreasing	0.57 (0.28 -- 1.1)	-0.13 (-0.23 -- 0.038)	-5.5 (18 -- -3)
Arkansas	56 (10 -- 269)	Decreasing	0.53 (0.23 -- 0.94)	-0.14 (-0.25 -- -0.016)	-5 (-44 -- -2.7)
California	1613 (364 -- 7574)	Likely decreasing	0.81 (0.45 -- 1.3)	-0.055 (-0.17 -- 0.087)	-13 (8 -- -4.1)
Colorado	783 (176 -- 3767)	Likely decreasing	0.94 (0.56 -- 1.4)	-0.018 (-0.13 -- 0.12)	-39 (5.9 -- -5.3)
Connecticut	340 (109 -- 1254)	Stable	0.98 (0.66 -- 1.5)	-0.0054 (-0.1 -- 0.12)	-130 (5.7 -- -6.9)
Delaware	98 (42 -- 254)	Likely increasing	1.1 (0.73 -- 1.6)	0.017 (-0.078 -- 0.14)	41 (4.8 -- -8.9)
District of				0.14 (-0.049 -	

How do we calculate the reproductive number?

If we have daily case counts:

1. **Real time:** EpiEstim method

2. **Retrospective**: Wallinga and Teunis

Alternative approach for getting at more nuance:
Mathematical models

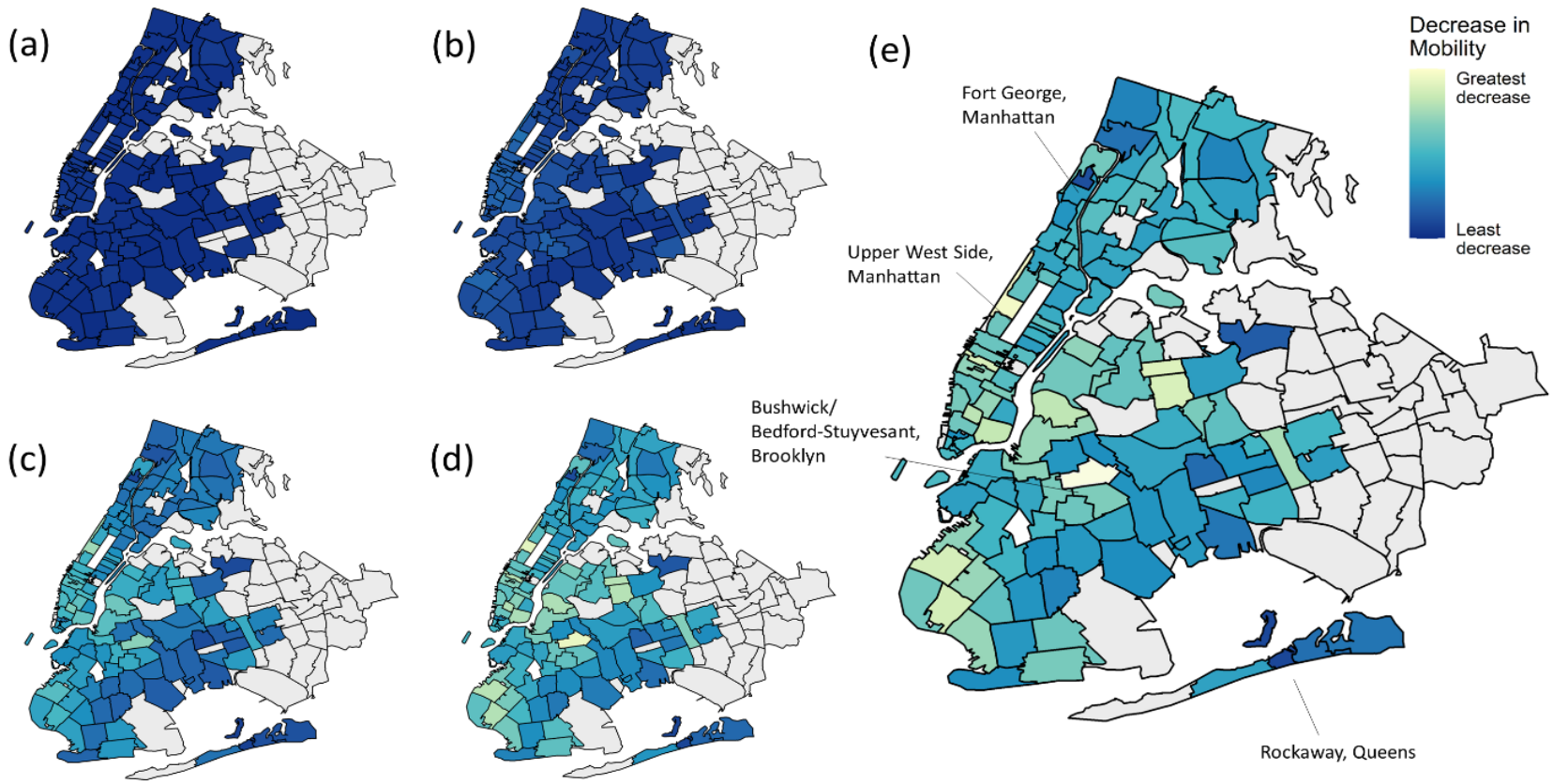
Modifications

This means that EpiEstim has been the preferred method to use in the current pandemic.

Modifications needed:

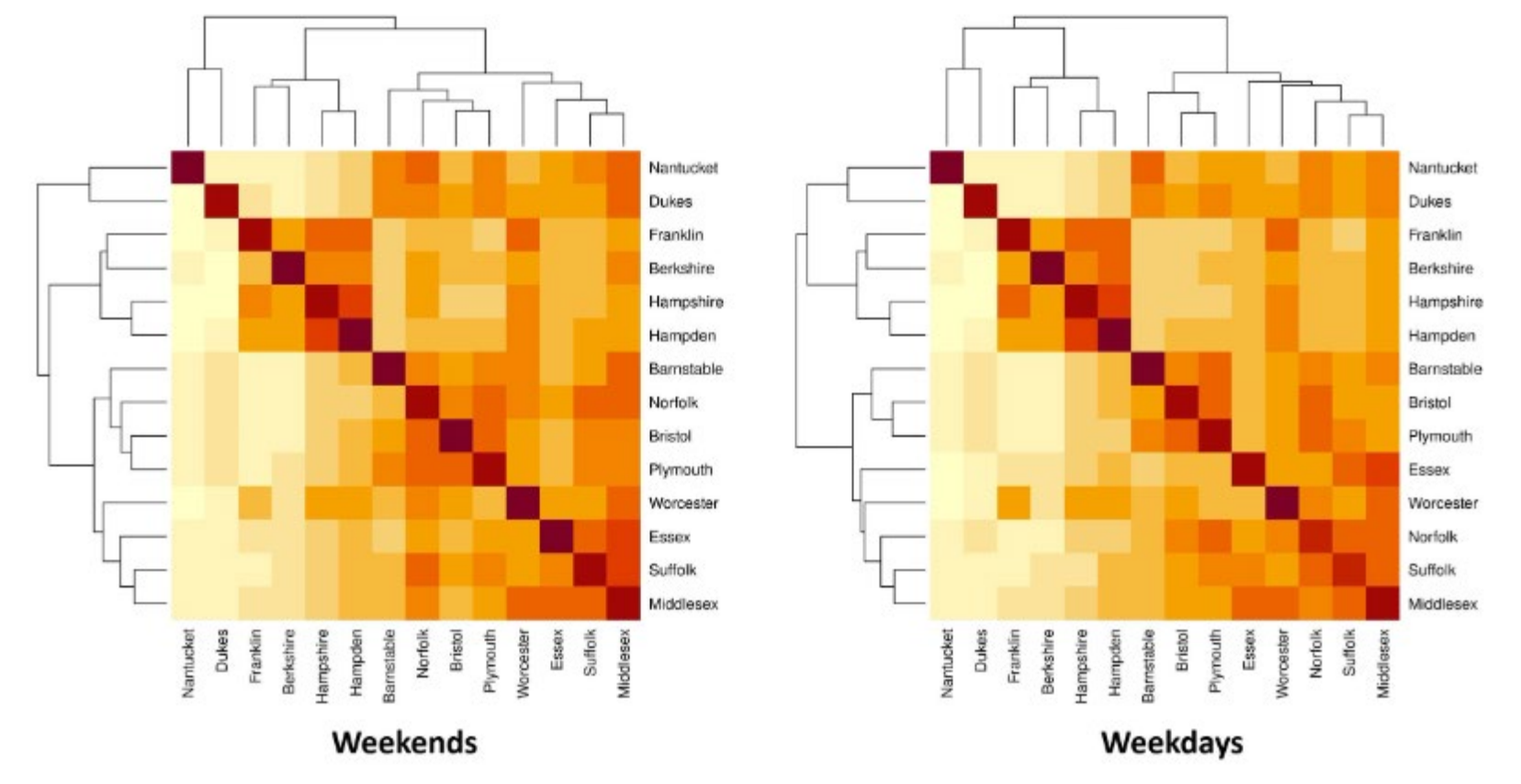
1. Backcalculation of cases to infection time
2. Heterogeneity
3. Immigration
4. Smoothing

Mobility matters

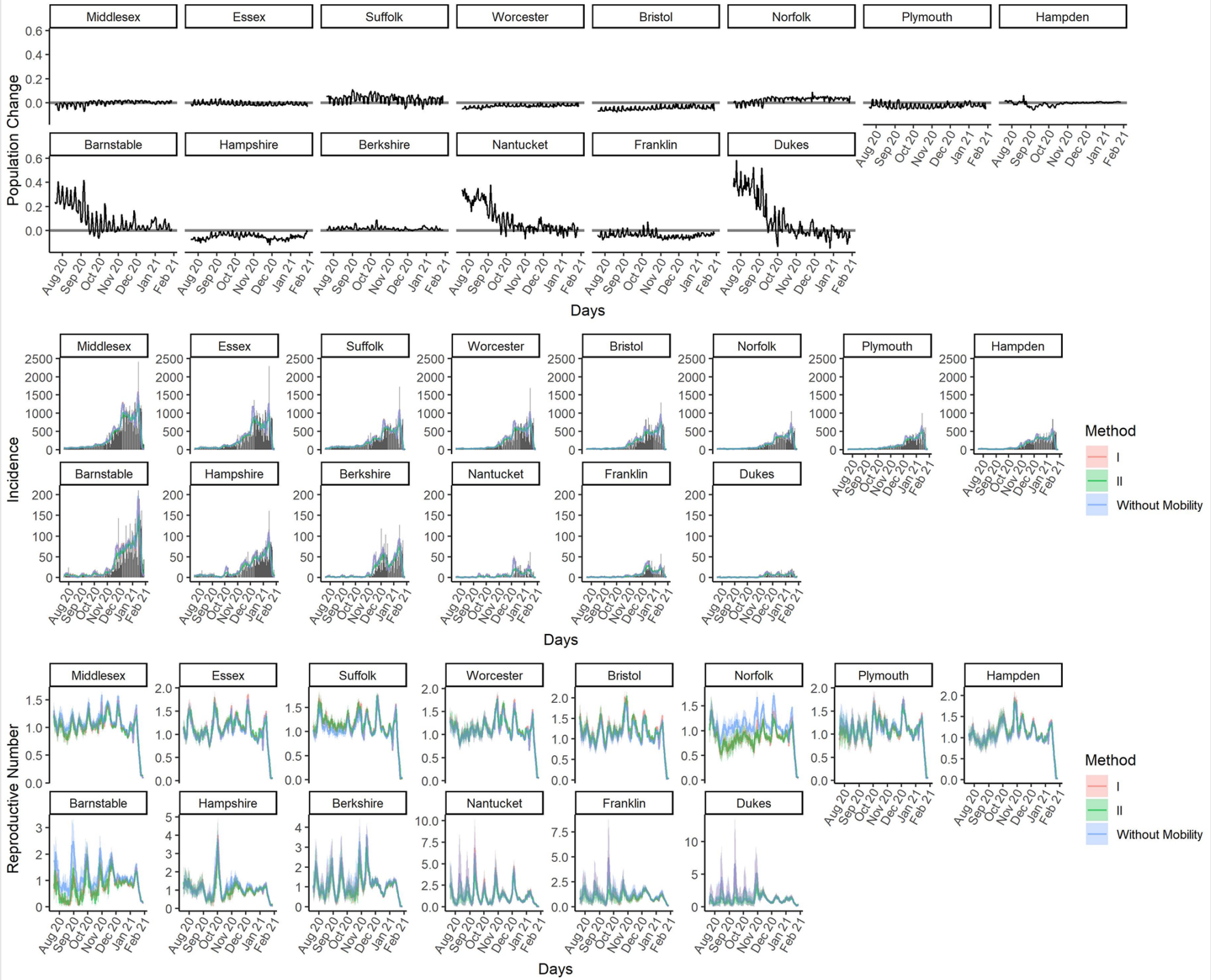


New York City reduction in subway use in ZCTAs during the COVID-19 outbreak on the week of (a) February 29, 2020 (b) March 7, 2020 (c) March 14, 2020 (d) March 21, 2020 (e) April 11, 2020. Reductions are calculated as the change in subway use relative to the pre-shutdown period and standardized by the pre-shutdown standard deviation. Panels b-d correspond to key New York City executive orders (b) Local state of emergency; restricted gatherings exceeding 500 persons; (c) City schools closure, (d) Stay-at-home order, non-essential businesses closure.

Example: mobility in MA

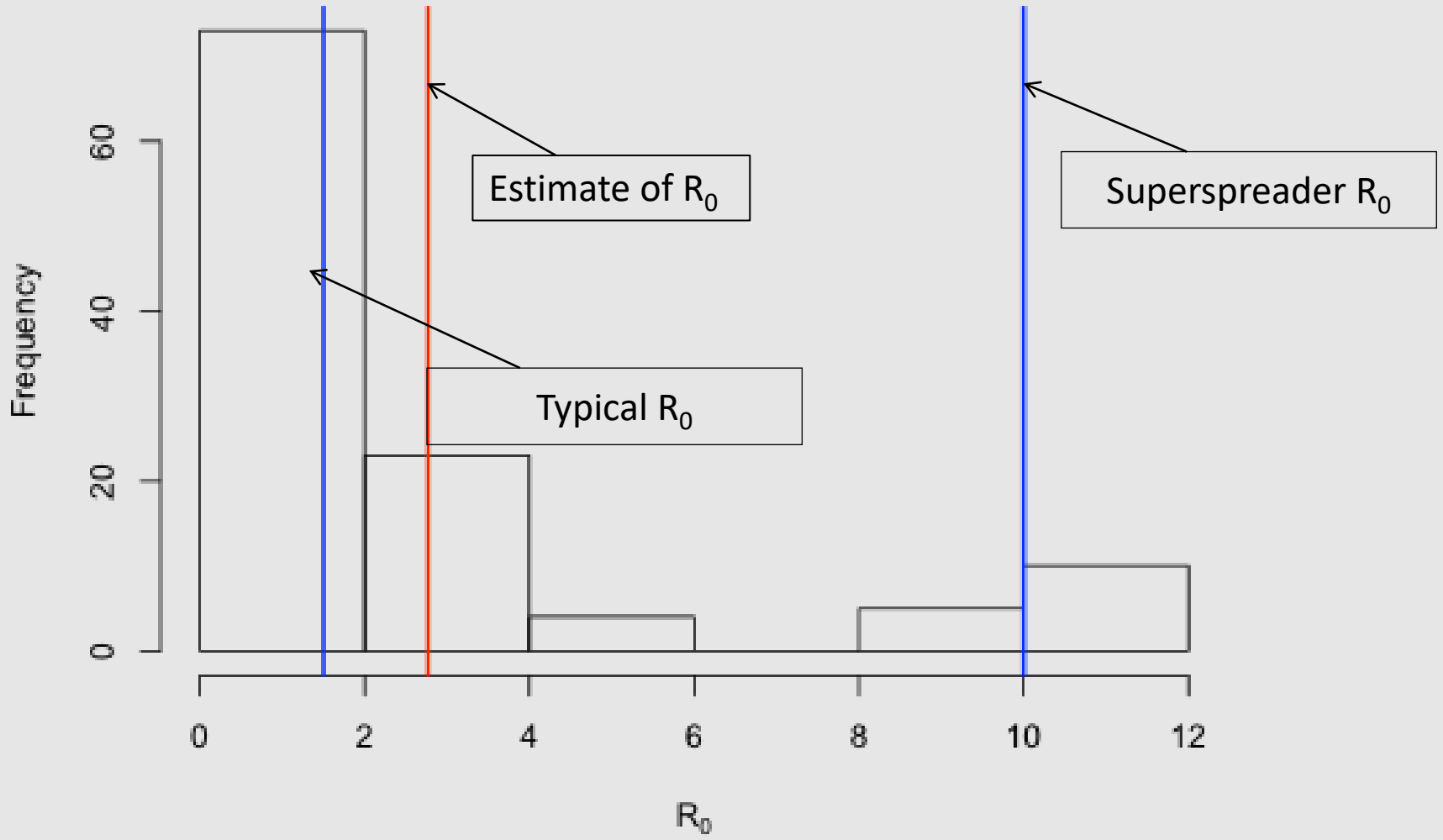


Zhou et al, 2022

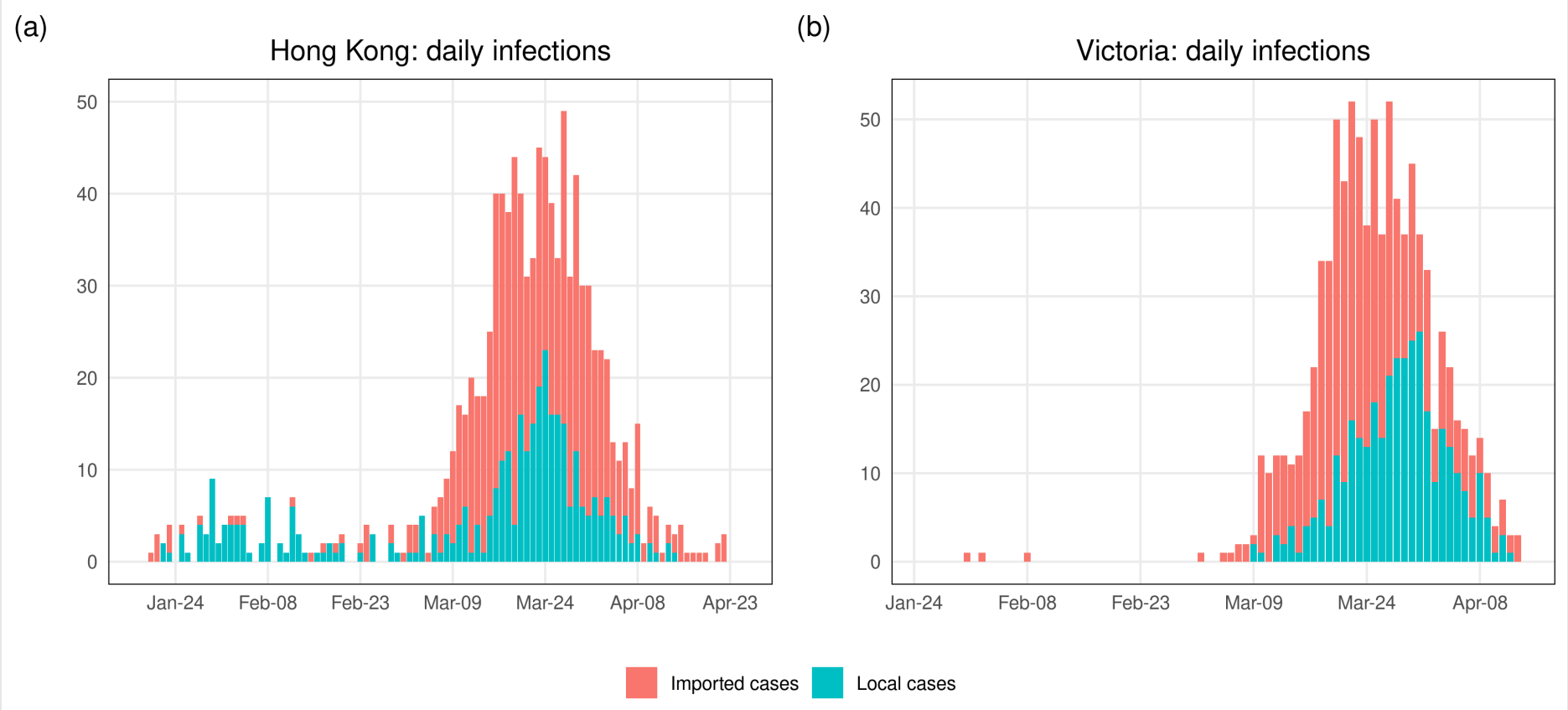

Middlesex
Essex
Suffolk
Worcester
Bristol
Norfolk
Plymouth
Hampden
Barnstable
Hampshire
Berkshire
Nantucket
Franklin
Dukes
Population Change
Incidence
Reproductive Number
Days
Aug 20
Sep 20
Oct 20
Nov 20
Dec 20
Jan 21
Feb 21
Method
I
II
Without Mobility

Superspreading

Superspreading

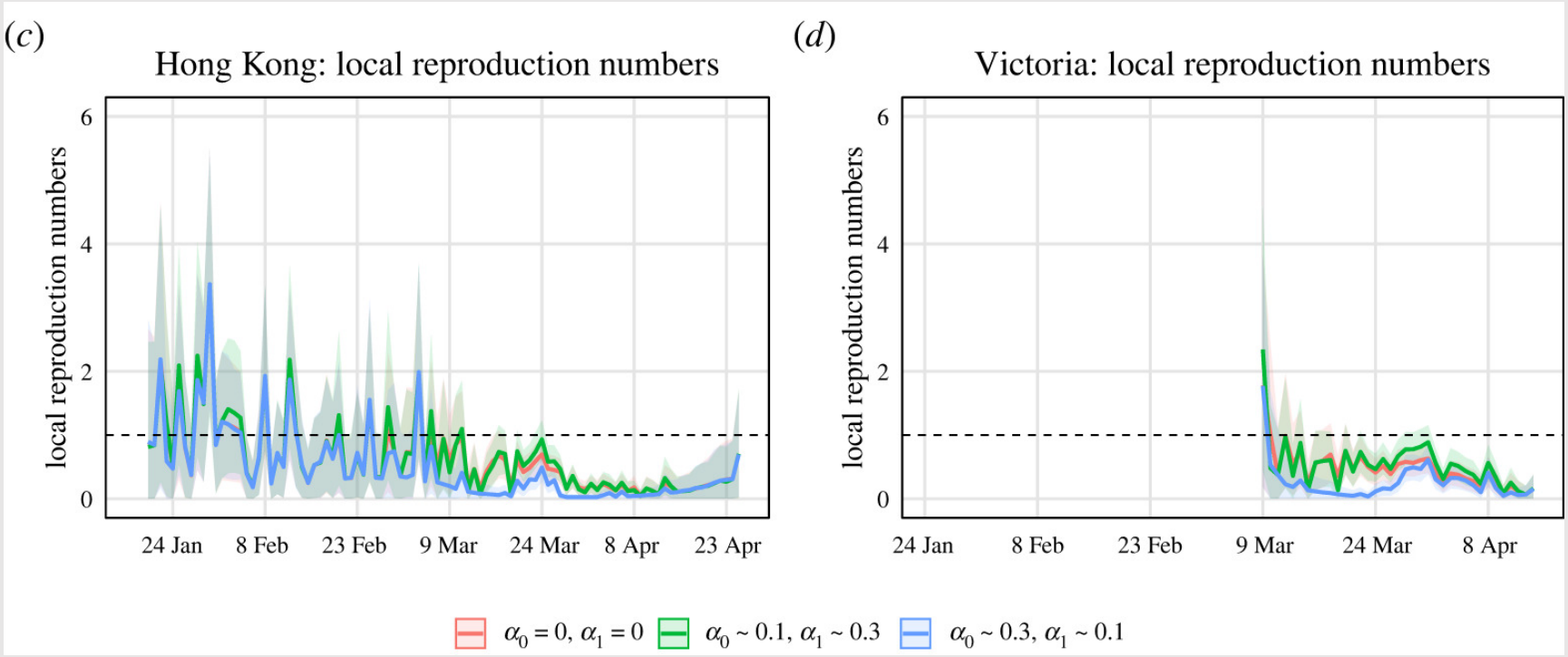


Example: imported cases

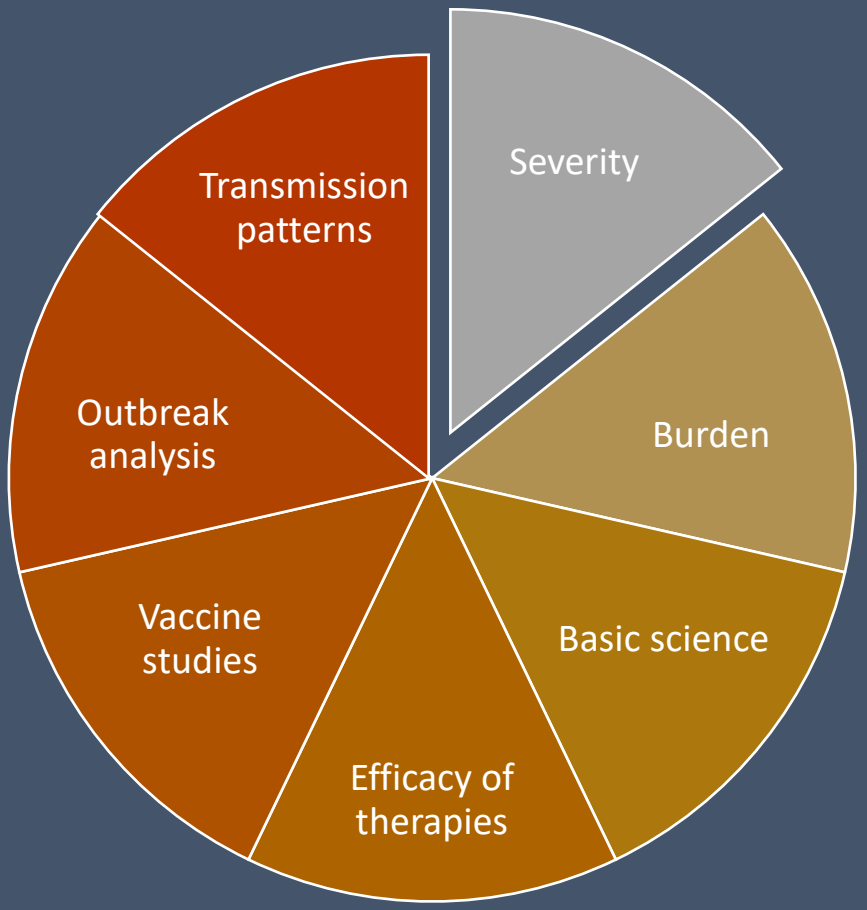


Li et al, 2022

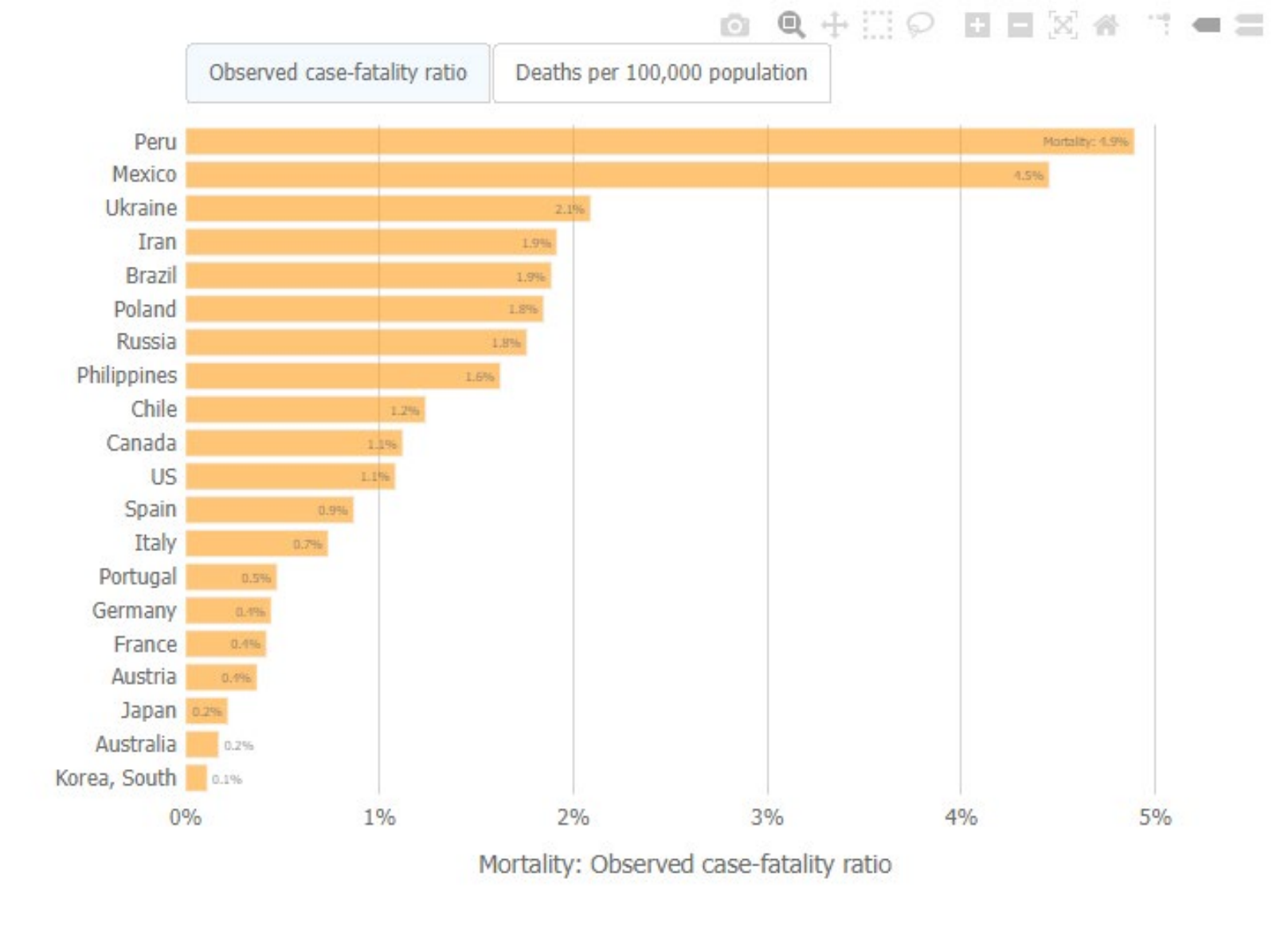
Example: imported cases



Li et al, 2022



Estimates of severity



https://coronavirus.jhu.edu/data/mortality

Severity

Case fatality ratio: ratio of **confirmed deaths from disease** to **confirmed cases of disease**.

Infection fatality ratio: ratio of ***total*** **deaths from disease** to ***total*** **cases of disease**.

Severity: Case (or infection) fatality ratio

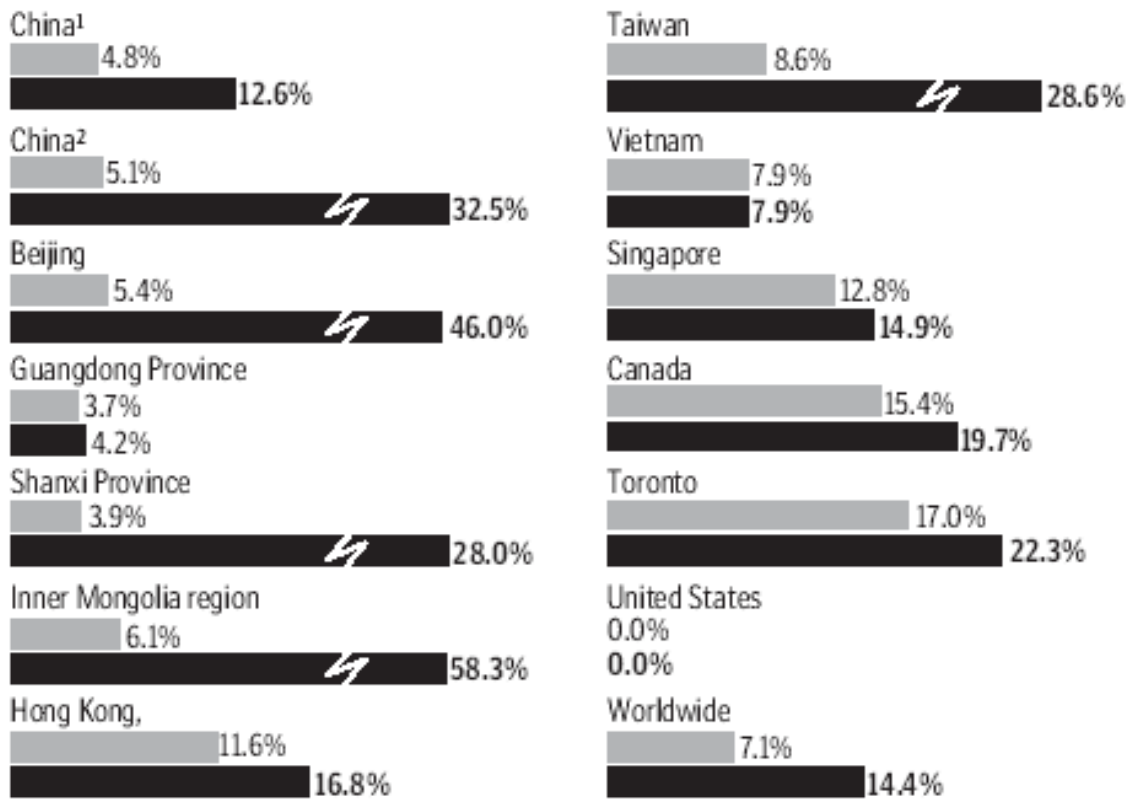
- Case-fatality ratio = deaths/cases
- Hard to estimate both numerator and denominator
- Cases (denominator)
  - Mild ones missed
  - Asymptomatic ones impossible to detect without serosurveys
  - Even severe cases may not be tested; test may be insensitive; may not be reported; etc.
- Deaths (numerator)
  - Not all deaths will have virus present at time of test
  - Not all will be tested
  - Censoring bias!

Estimates of severity (CFR)

- Influenza
  - 1918 Influenza: CFR $\geq$ 2.5%
  - Other pandemics: <0.1% (Taubenberger et al, 2009)
  - H5N1: 60% (WHO< 2006); Li et al (2008): 14-33%
- Seasonal Influenza: 0.1-0.2% in US
- Legionnaires (Dominquez et al, 2009)
  - Of 1938 cases reported, 164 died (CFR: 8.5%).
  - Has decreased though time due to improved diagnostics and detection of outbreaks
- Ebola Zaire virus: CFR $\geq$ 90%
- MERS-CoV: 34% (Munster et al, 2020)
- SARS-CoV-1: 10% (Munster el al, 2020)

Changes in estimates of CFR in SARS

- What percentage of SARS patients die? How do you know? What about the patients who haven't died yet?



Deaths / Cases

Deaths /
(Cases with known outcome)

Statistically correct method:
CA Donnelly et al.
Lancet June 2003

Boston Globe, May 6, 2003

CFR for SARS

- SARS
  - WHO, 2003: 12-17%
  - Donnelly et al, more statistically correct approach
    - 13.2% (9.8-16.8) for under 65 years old; 43.3% (35.2-52.4) for at least 65 years old
    - Using a nonparametric approach: 6.8% (4.0-9.6) and 55.0% (45.3-64.7)

2009 Influenza: example of the difficulties

- Mexico, May 4
  - 509 confirmed cases
  - 19 deaths (CFR=4%)
- US, May 4
  - 268+786 confirmed + probable cases
  - 1 death (CFR=0.1%)
- What was going on?

Issues in estimating severity

- Censoring bias (missing deaths; underestimate severity)
- Mild cases not detected (overestimate severity)

- Improvements
  - Garske et al. *BMJ* 14 Jul
    - CFR 0.2-1.2%
    - Focus on censoring bias

EUROSURVEILLANCE Vol. 14 - Issue 26 - 2 July 2009

THE EMERGING INFLUENZA PANDEMIC: ESTIMATING THE CASE FATALITY RATIO

N Wilson (nick.wilson@otago.ac.nz)[1], M G Baker[1]
1. Department of Public Health, University of Otago, Wellington, New Zealand

the plausible range of the CFR for symptomatic infection by this pandemic strain in developed countries. All of the methods produce substantially lower values (range 0.06% to 0.0004%) than a previously published estimate for Mexico (0.4%). As these

People take time to die (SARS)

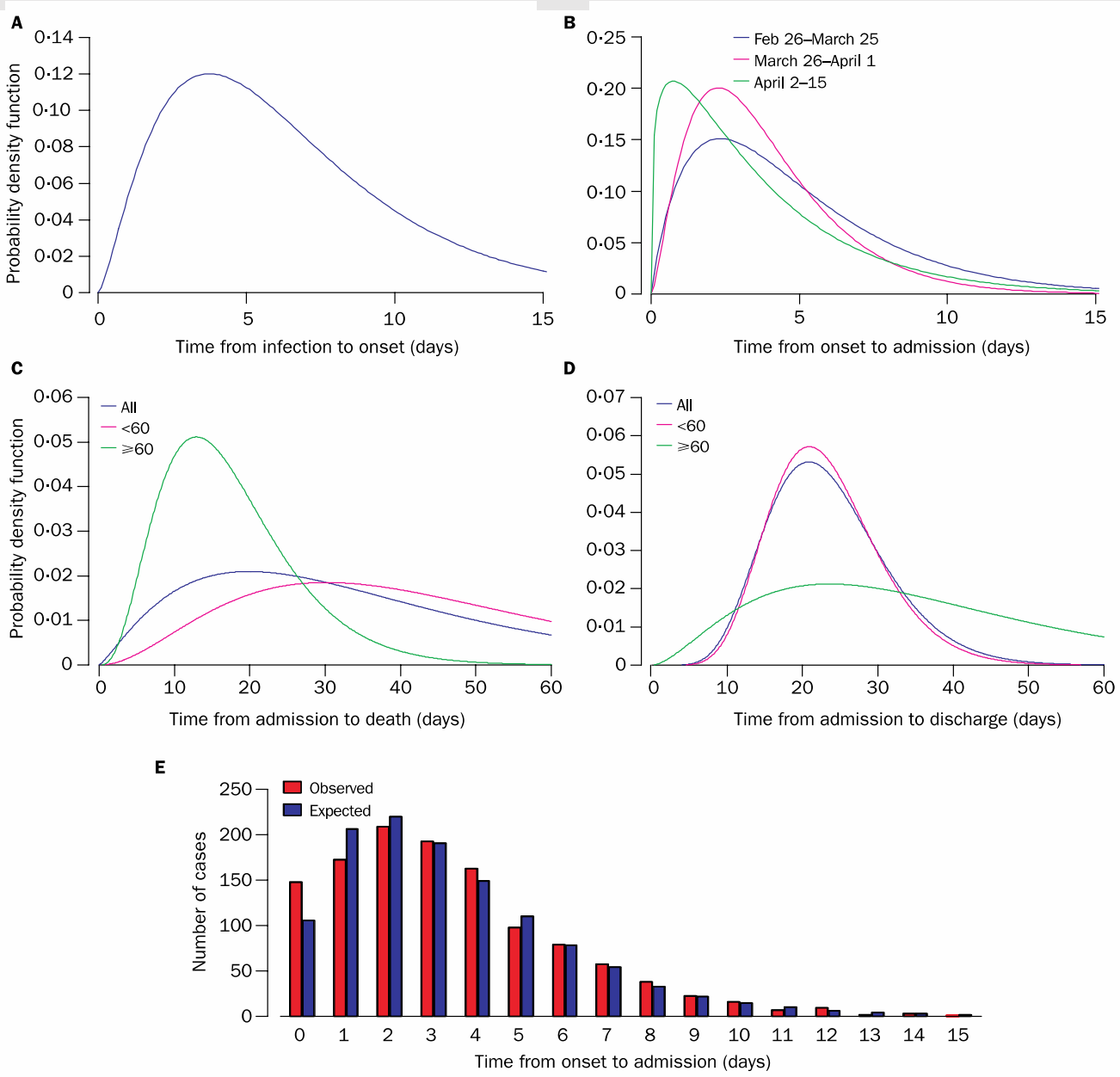
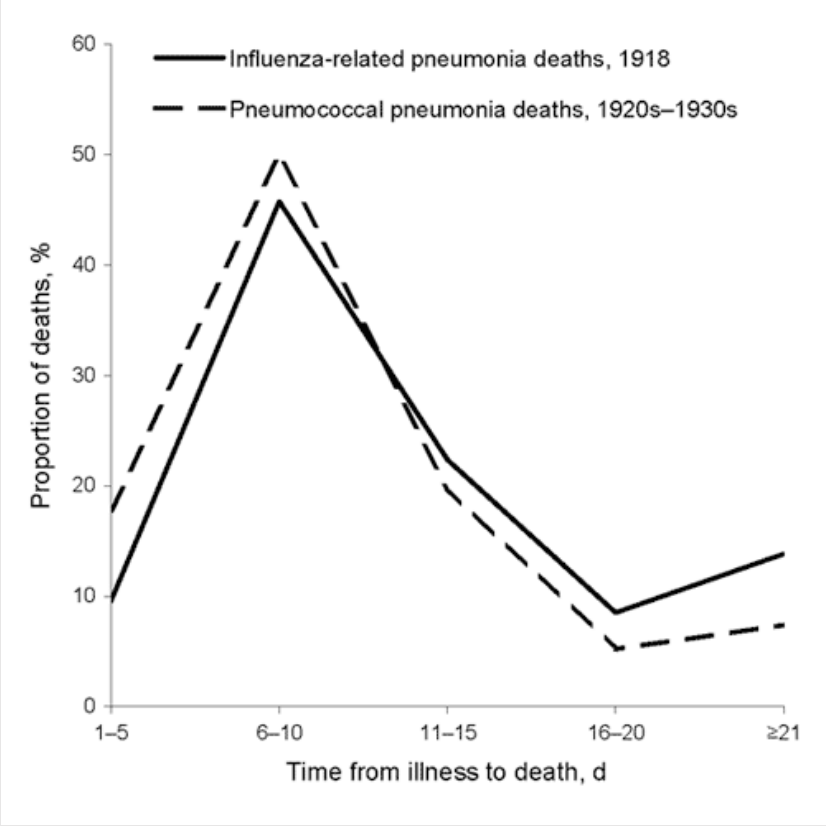


Figure 2: **Maximum likelihood estimates**
A: Infection-to-onset distribution. B: Time-dependent onset-to-admission distribution as a function of time of onset of clinical symptoms. C: Admission-to-death distribution by patients' age. D: Admission-to-discharge distribution by patients' age. E: Observed and maximum likelihood estimated onset-to-admission intervals in presence of censoring.

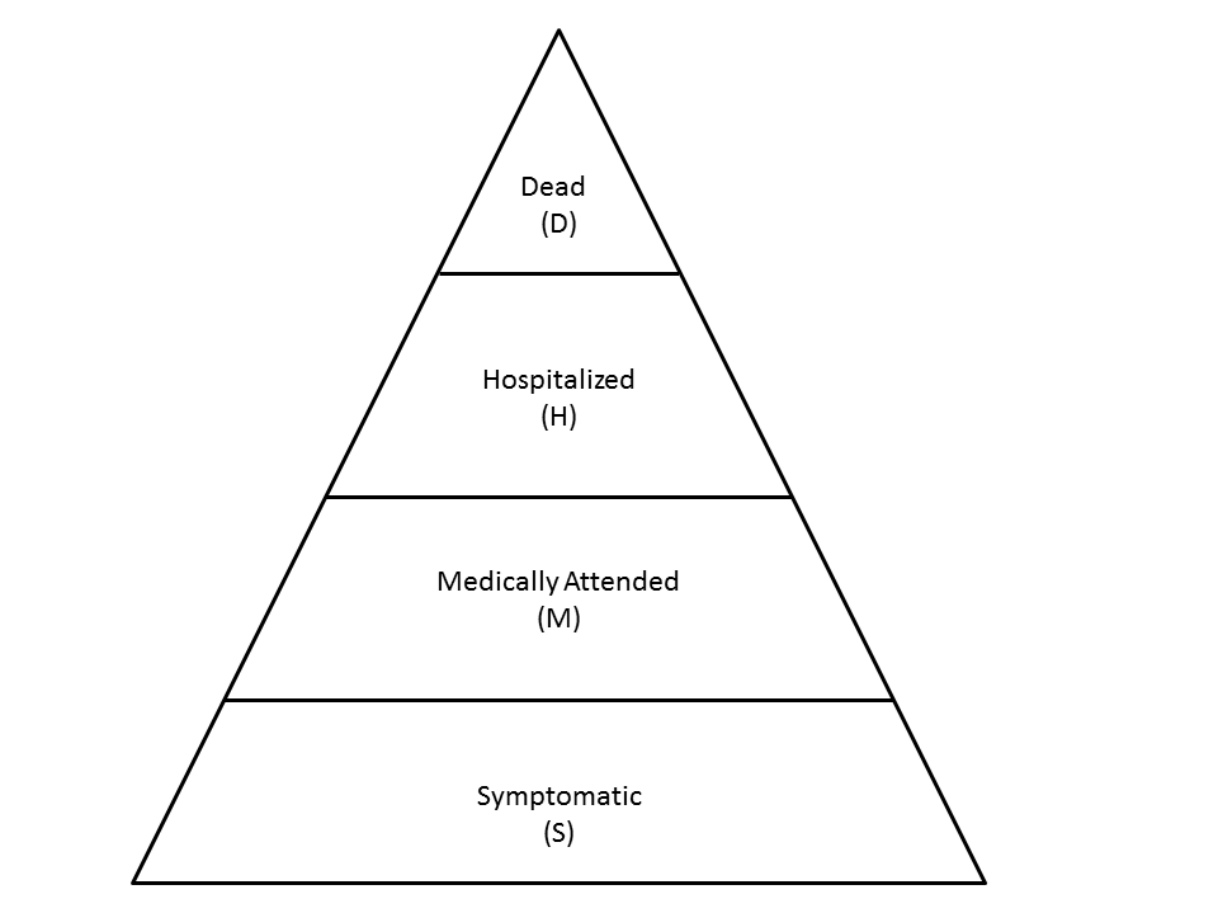
Source: Donnelly et al (2003)

People take time to die (P&I mortality)



Klugman Astley Lipsitch *Emerg Inf Dis* 2009

Illness Pyramid



Reported cases are the tip of a pyramid (the technical term for an iceberg)

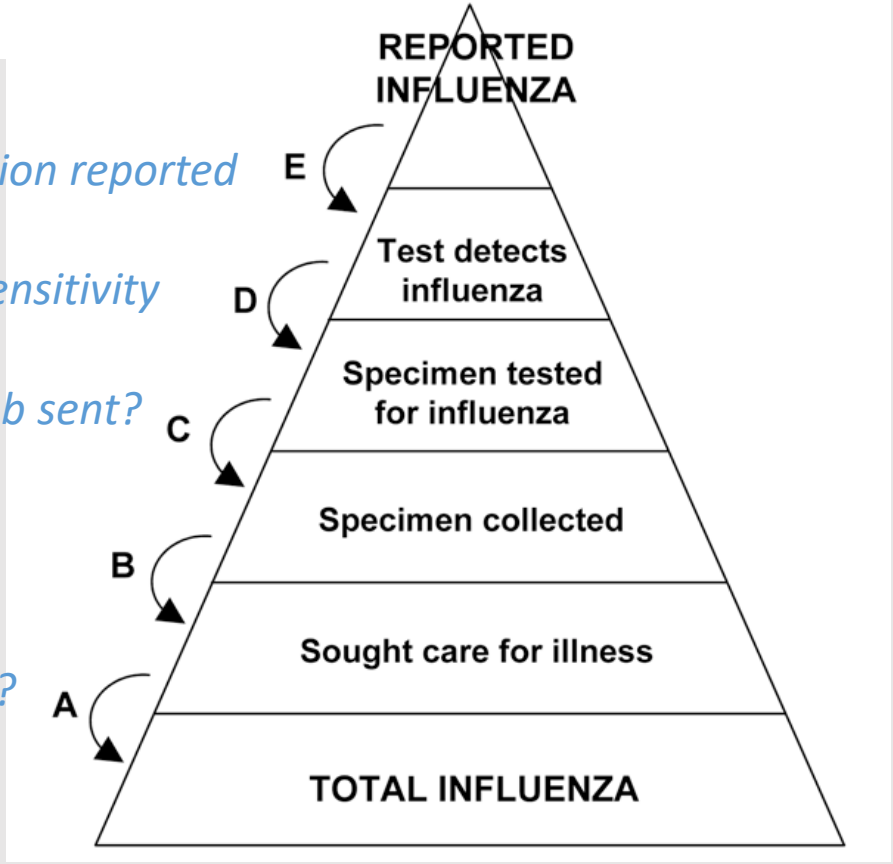
U(.95,1) *Educated guess: fraction reported*

U(.9,1) *Literature review: test sensitivity*

U(.2,.3) *Outbreak investigation: was swab sent?*

U(.19,.34) *(to yes) Did the provider take a swab?*

U(.42,.58) *Survey: did you have fever+cough or sore throat? (to yes) did you seek medical care?*



C Reed et al. Emerg Inf Dis 2009