

Title: COVID-19: Detecting the Dark Matter of the Disease

Speakers: Bruce Bassett

Series: Cosmology & Gravitation

Date: June 16, 2020 - 11:00 AM

URL: <http://pirsa.org/20060022>

Abstract: COVID-19 is a mysterious disease associated with a large number of unanswered questions.&nbsp;

In this talk we review what is currently known, what is still a mystery and highlight some of our recent work on the role of climate, blood type and vaccinations on the transmission of the disease and on&nbsp;the extent of "dark infections", the asymptomatic and untested proportion of infections. We end with a list of open research questions that may be amenable to techniques from physics and data science.

# COVID-19: Detecting the Dark Matter of the Disease

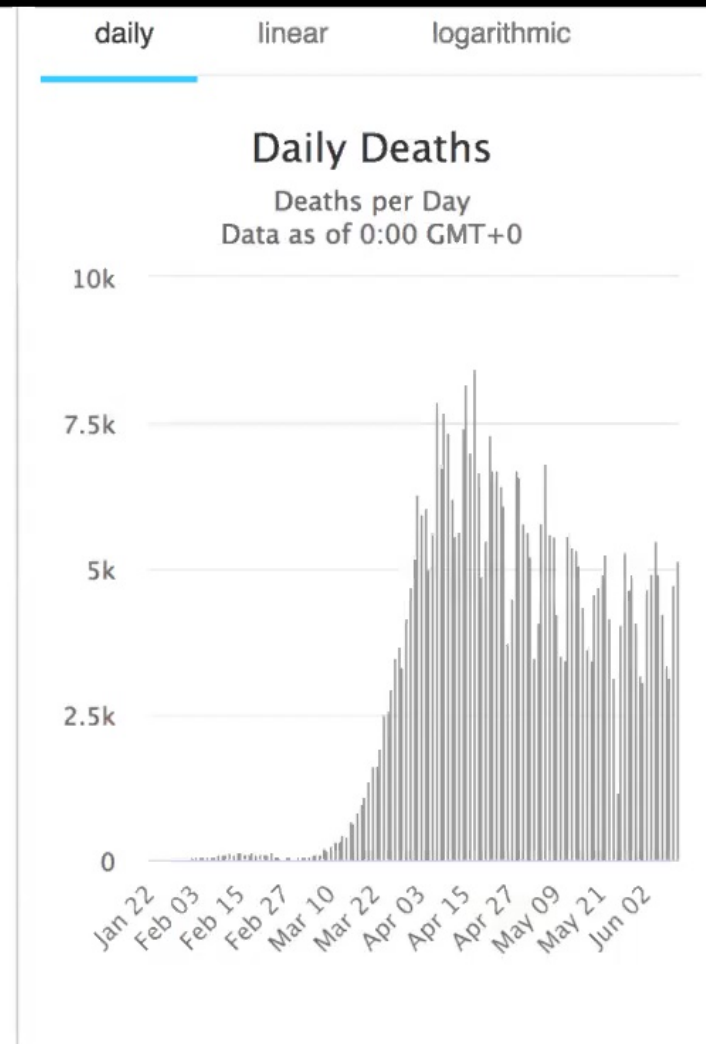
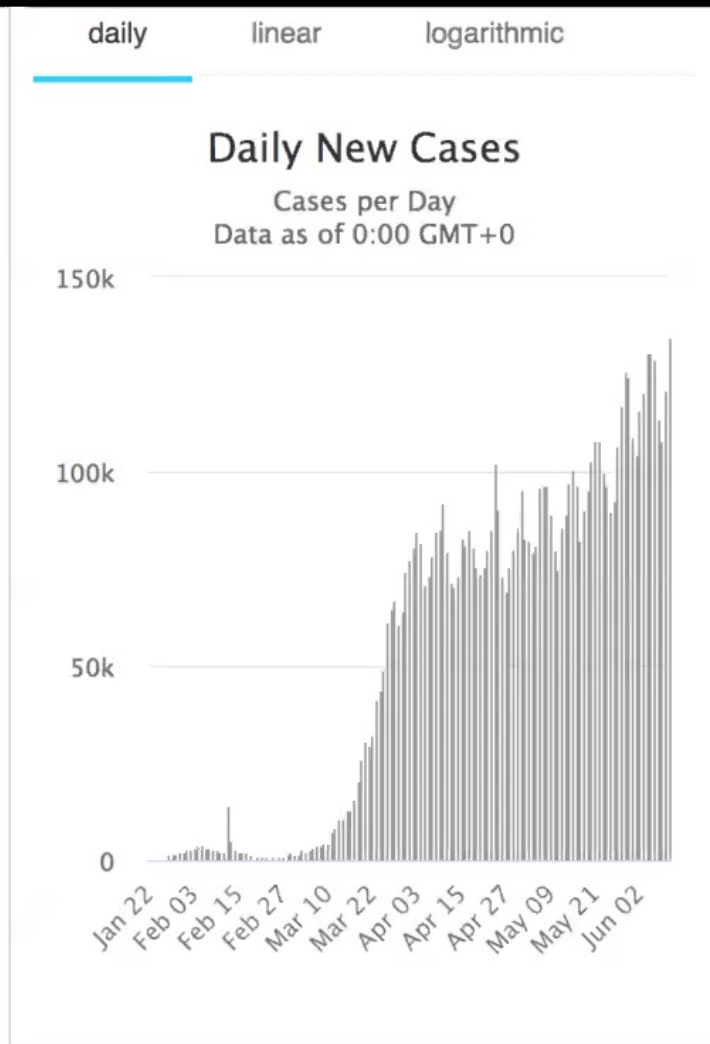
Bruce Bassett

AIMS-SAAO-SARAO-UCT

16 June 2020

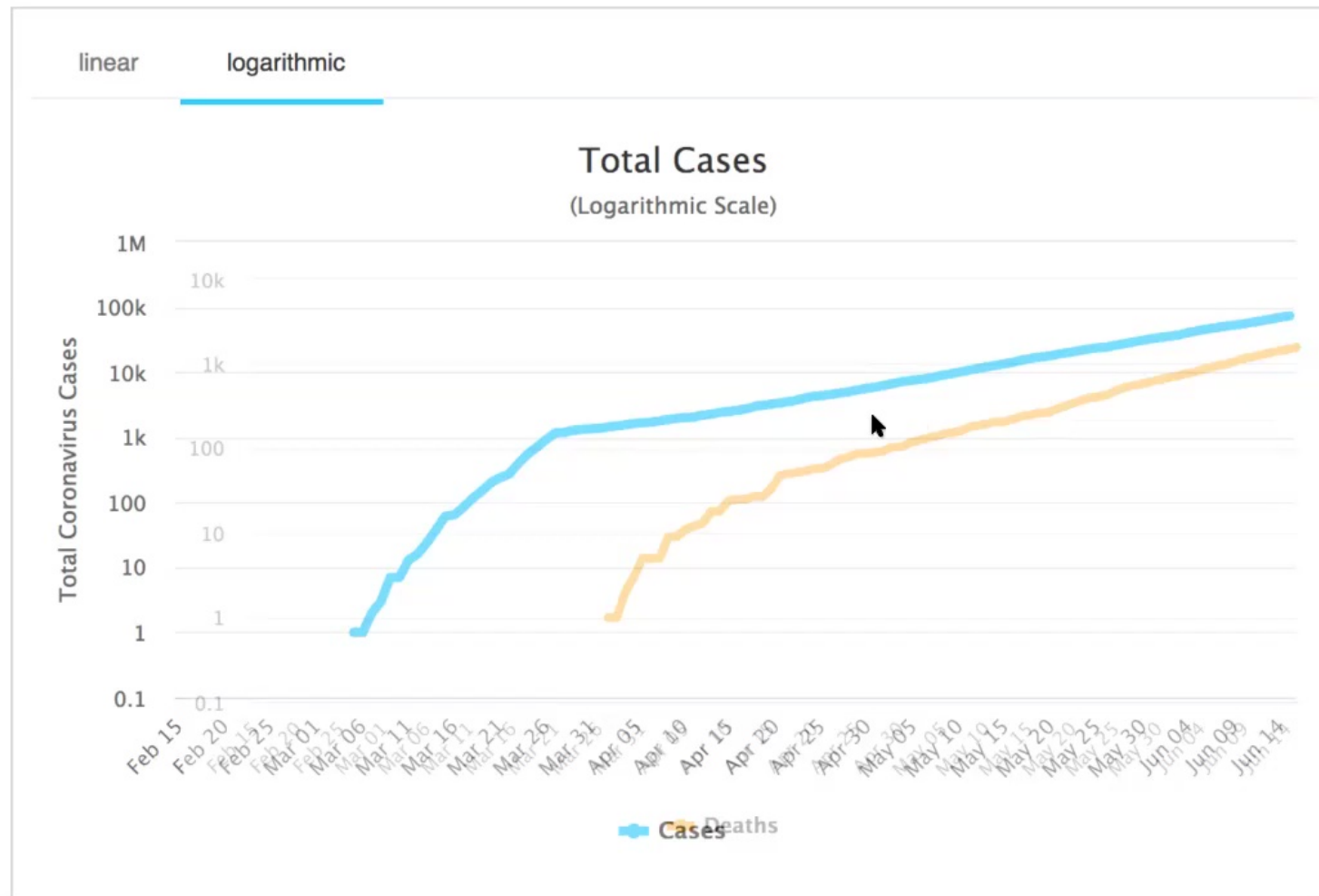
In collaboration with R. Armstrong, C. Finlay,  
N. Oozeer and F. Silwamba





<https://www.worldometers.info/coronavirus/>

## Total Coronavirus Cases in South Africa



## Some Mysteries

- Where did it originate? Is it natural or man-made?
- When did it start circulating in humans?
- How long are people infectious for?
- Does a positive PCR test indicate infectiousness?
- Are sicker people more infectious than mild cases?
- What factors are important in the spread? Climate, blood type, BCG?
- What is the main transmission route? How effective are masks?
- How infectious are asymptomatic people?
- Is there more than one main strain of the virus?
- Are superspreaders responsible for most infections?

## Some Mysteries - continued

- Why does it affect so many systems in the body?
- What are the long term effects and how long does it take to recover?
- What fraction of recovered people develop antibodies?
- How long will immunity last? Do some people have partial immunity?
- Can a vaccine be developed?
- How deadly is it? How bad is it to have HIV?

# Some Mysteries

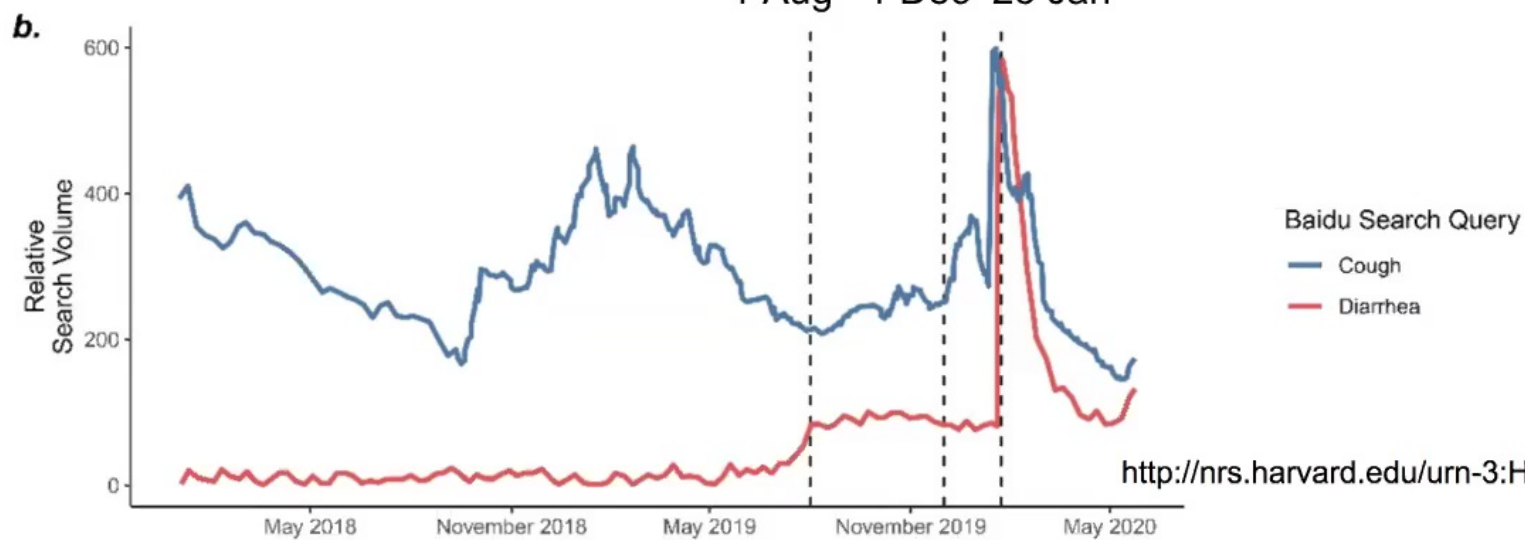
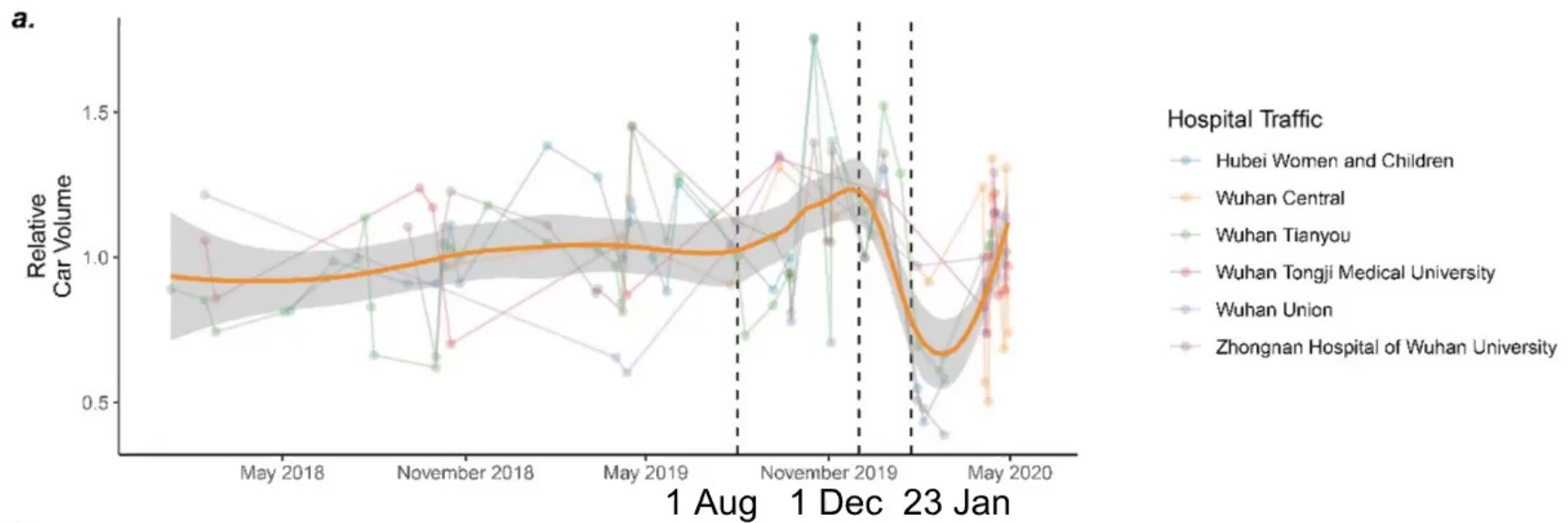
- When did it start circulating in humans?
- What are the long term effects and how long does it take to recover?
- What factors are important in the spread? Climate, blood type, BCG?
- How deadly is it? How bad is it to have HIV?

# How long has it been circulating?

- First case in Paris at least as early as December 27
- “We observe an upward trend in hospital traffic and search volume beginning in late Summer and early Fall 2019” - i.e. perhaps spreading as early as August 2019

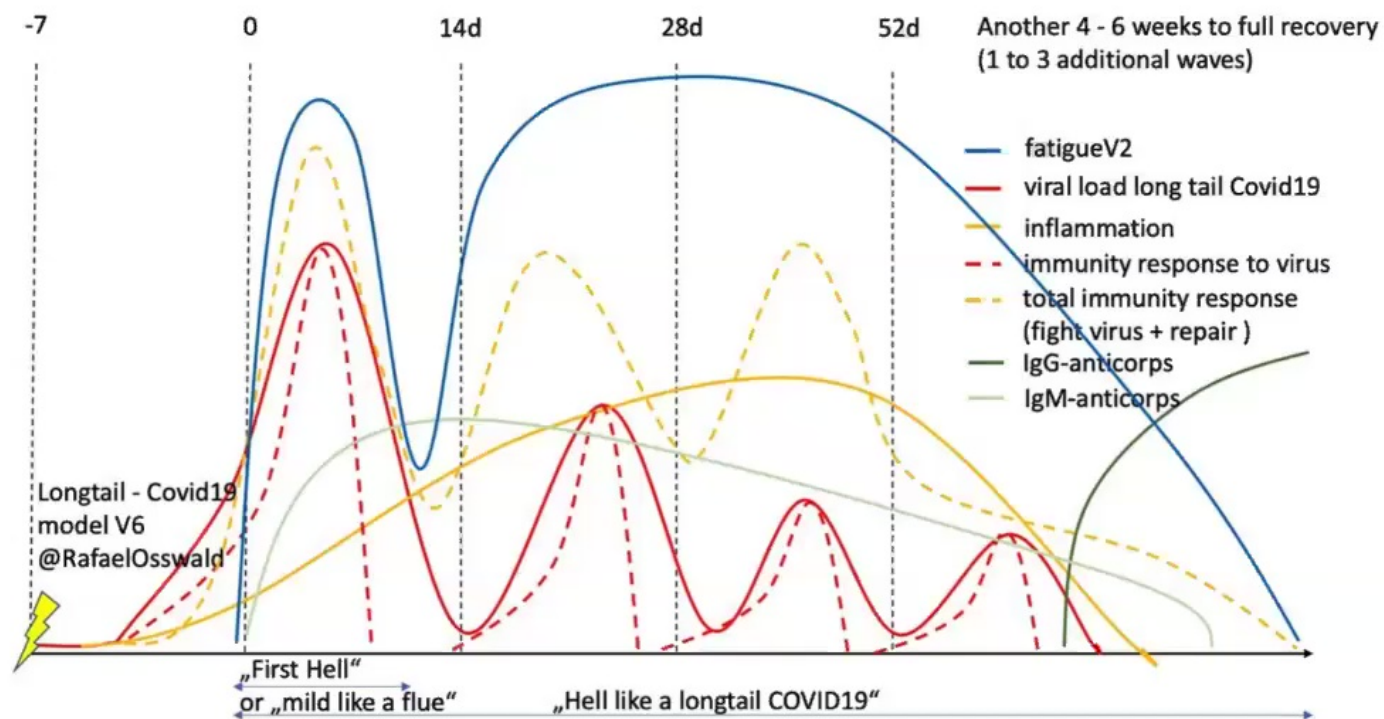
<http://nrs.harvard.edu/urn-3:HUL.InstRepos:42669767>





<http://nrs.harvard.edu/urn-3:HUL.InstRepos:42669767>

# How long does it take for mild cases to recover?



<https://twitter.com/RafaelOsswald/status/1264126160805474305>

# How deadly is the disease?

**What is the probability of you dying if you get infected? The Infection Fatality Rate (IFR)**

Sounds like an easy thing to calculate but is actually very difficult to assess for multiple reasons.

# Common Mistakes in Calculating the IFR

Estimate: Calculate the crude CFR: the number of deaths divide by the number of cases at the same time.

E.g. On June 16 Italy has had 34k deaths and 237k infections = 14.3%

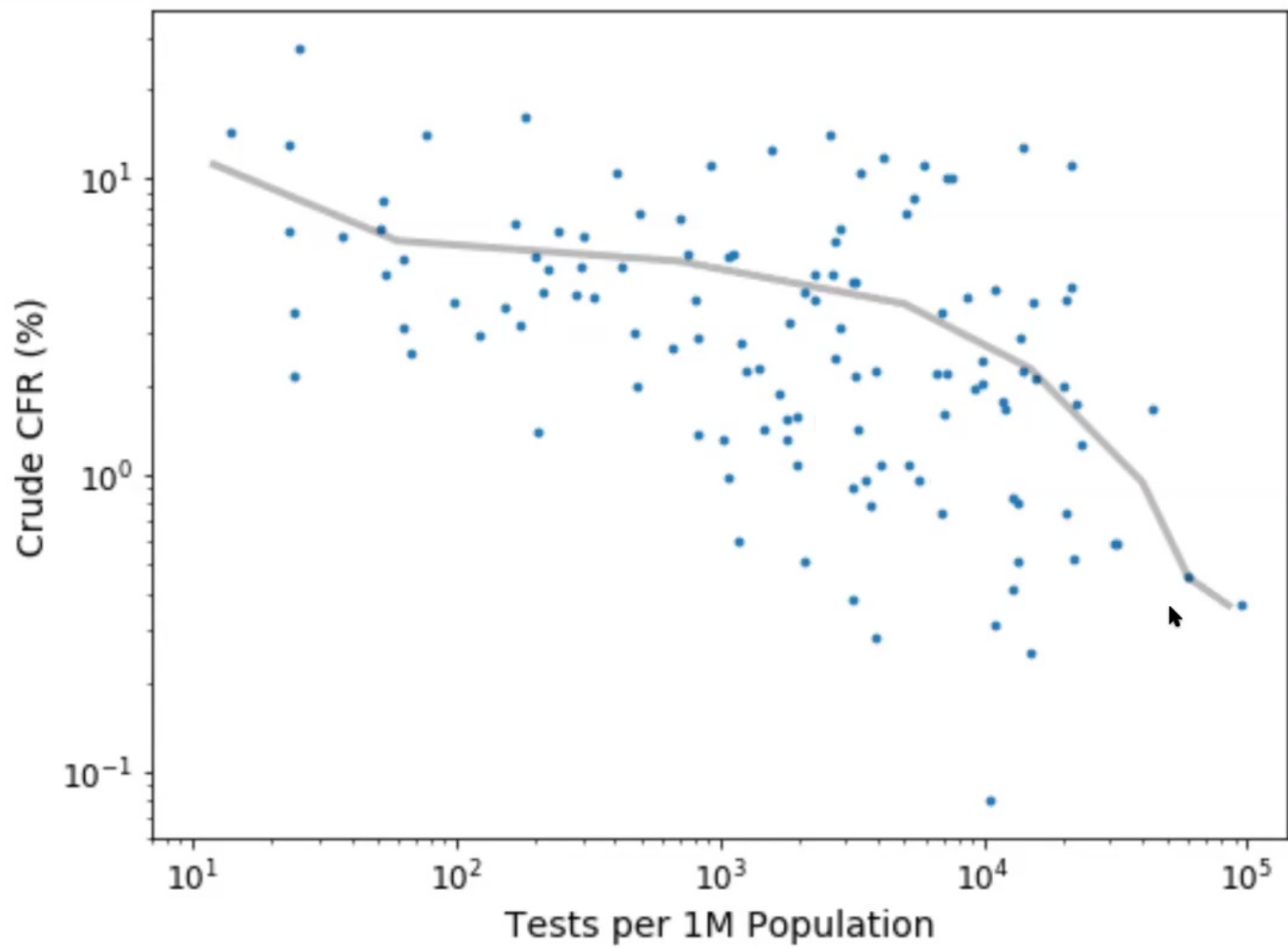
## Problems...

- Missing cases (like in astronomy; Malmquist bias)
- Delayed deaths (14 day mean time to death)
- Missing deaths (people who die at home/old age homes)
- Fundamentally gauge-dependent

# Missing Cases....

- The disease is very "clever" - a significant fraction of people have almost no symptoms (asymptomatic cases). Exact fraction is probably somewhere between 10 and 80%
- Not everyone can be tested - PCR testing is limited. Typically only the most severe get tested (Malmquist bias); and the False Negative Rate can be > 50% if delayed.
- How many infections are being missed?

Compare Belgium (**60,155 cases and 9,663 deaths on June 16**) and Qatar (**82,077 cases and 80 deaths**). Their Case Fatality Rate (CFR) differs by **100x** with similar testing protocols (90k and 107k tests per Million population)



# Is COVID-19 just like the flu? **No**

The IFR for influenza is typically quoted as being  $< 0.1\%$

Can we put a lower bound on the IFR for COVID-19?

Yes. Just assume that **everyone** in a region has been infected.

If we do this for NYC+Lombardy+Madrid we get 21k+16k+9k deaths among 25.1M people

→ **average IFR  $> 0.18\%$**  (caveat: overwhelmed healthcare systems)

D. Bressan <https://www.researchprotocols.org/2020/1/e194104/0000-04-00-00076026v1> and LinkedIn



# Antibody estimates of total infected...

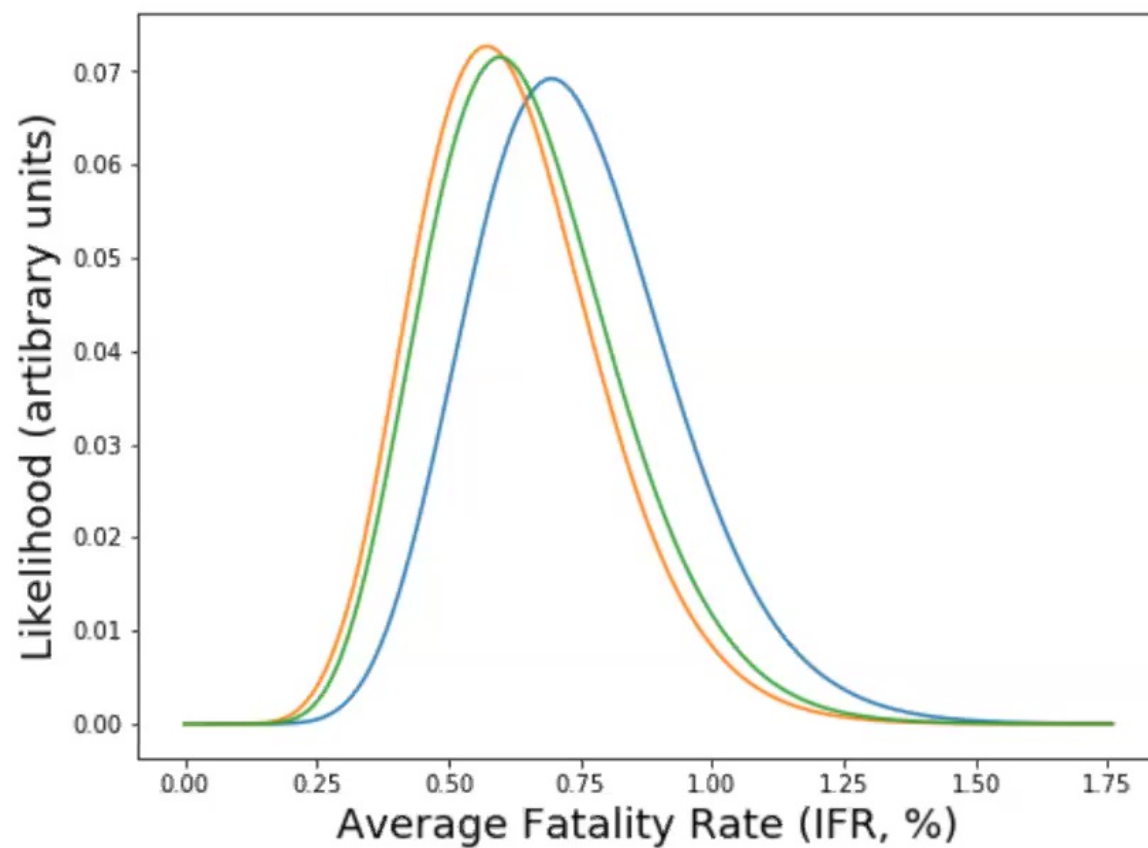
Serological studies suggest:

- Around 5% of Spanish population infected (but inhomogeneous) → 2.35M infections compared with 289k confirmed positive
- Around 4% infected in Wuhan

Suggests an IFR of around **0.6-1.3%** on average...though that doesn't account for missing deaths.



## IFR from the Diamond Princess



# How many people will die in South Africa?

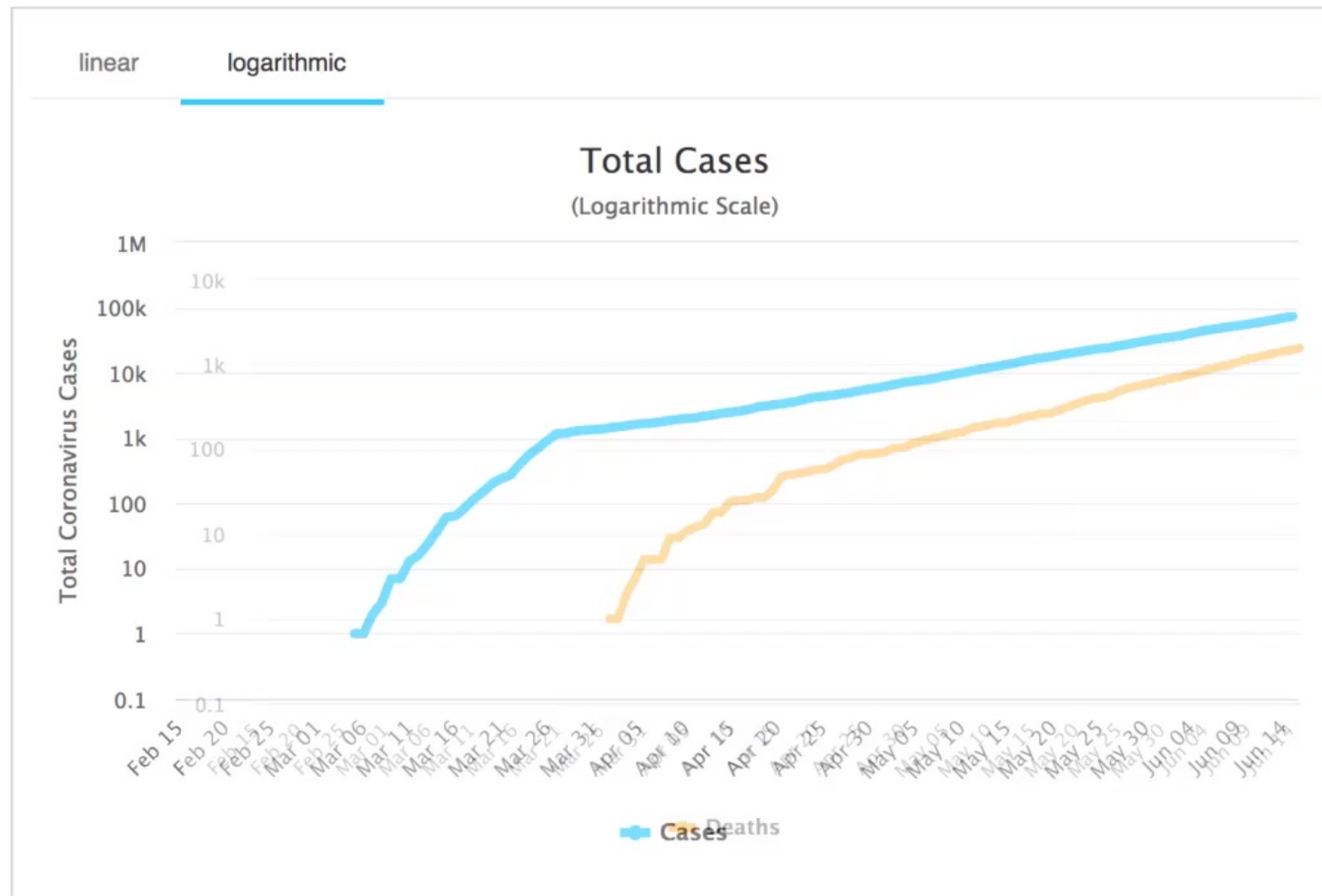
To answer this question we need to know two things:

1. What is the average IFR for South Africa
2. How many people will be infected?

# Some Mysteries

- Where did it originate? Is it natural or man-made?
- When did it start circulating in humans?

## Total Coronavirus Cases in South Africa



## Some Mysteries

- Where did it originate? Is it natural or man-made?
- When did it start circulating in humans?
- How long are people infectious for?
- Does a positive PCR test indicate infectiousness?
- Are sicker people more infectious than mild cases?
- What factors are important in the spread? Climate, blood type, BCG?
- What is the main transmission route? How effective are masks?
- How infectious are asymptomatic people?
- Is there more than one main strain of the virus?

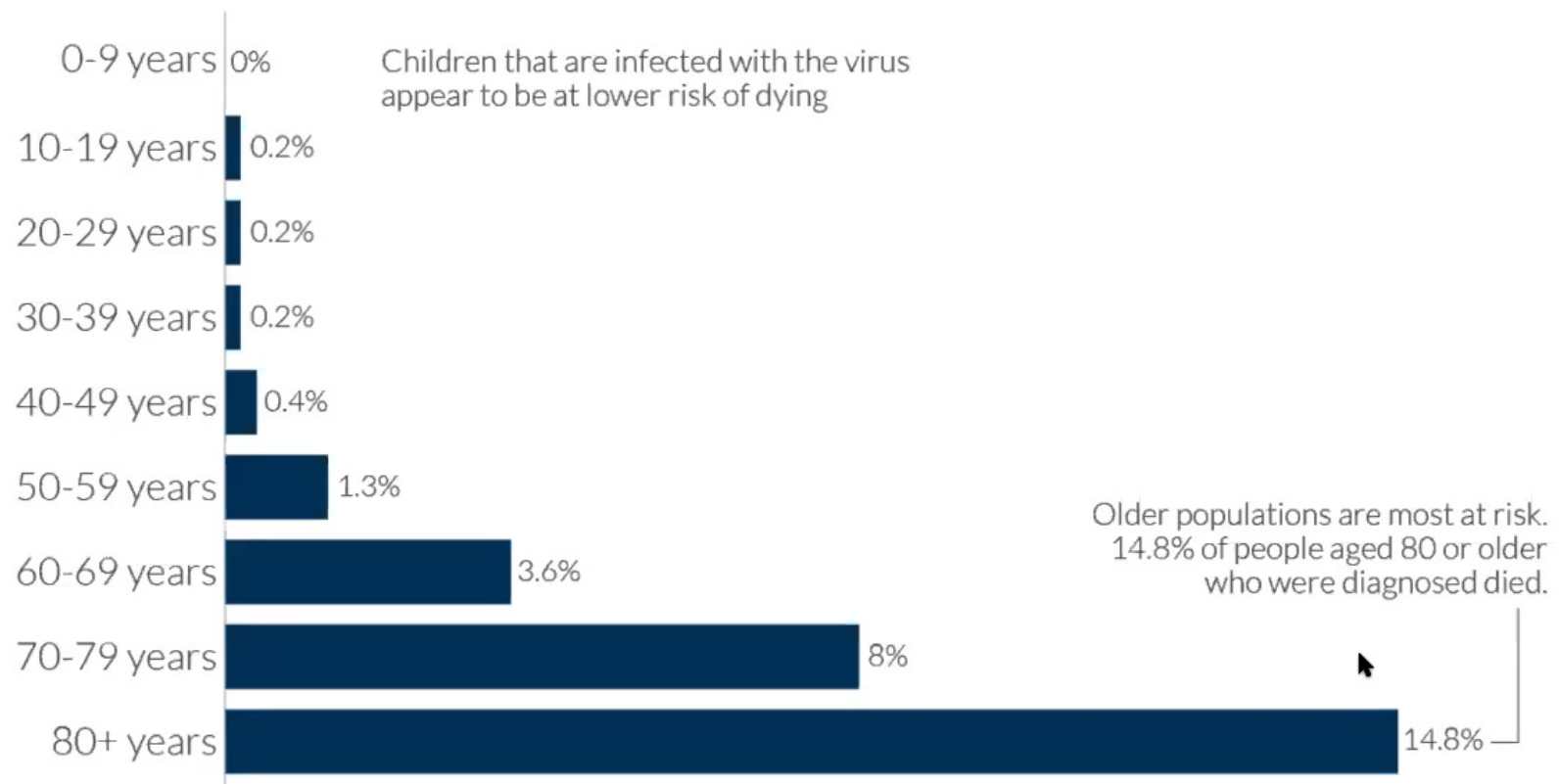
# Common Mistakes in Calculating the IFR

Estimate: Calculate the crude CFR: the number of deaths divide by the number of cases at the same time.

E.g. On June 16 Italy has had 34k deaths and 237k infections = 14.3%

# Coronavirus: early-stage case fatality rates by age-group in China

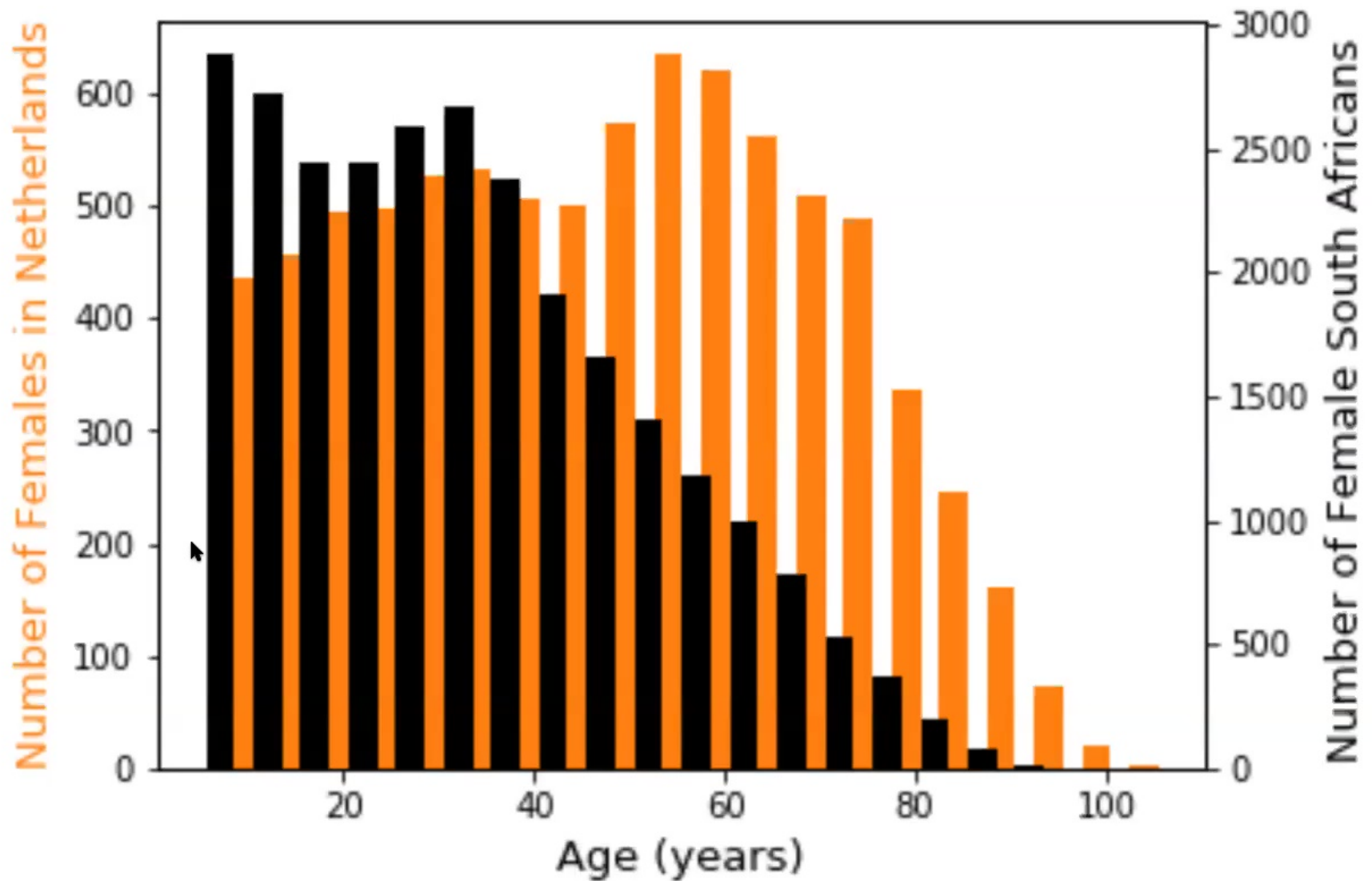
Case fatality rate (CFR) is calculated by dividing the total number of deaths from a disease by the number of confirmed cases. Data is based on early-stage analysis of the COVID-19 outbreak in China in the period up to February 11, 2020.



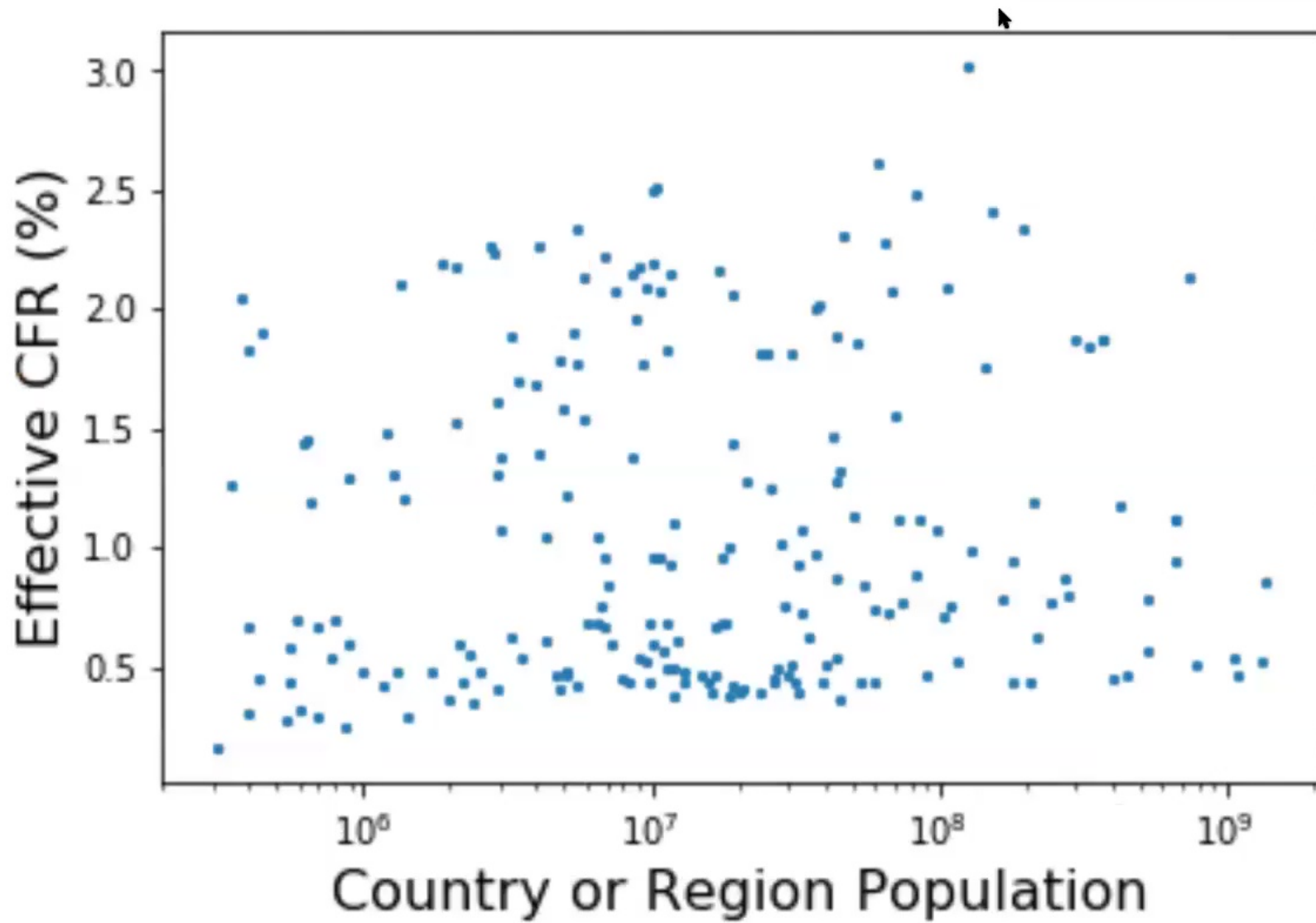
Data source: Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. *Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020.* China CDC Weekly.

OurWorldinData.org - Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors.







## Effective CFR's by region...

- Italy - 15,812 - 2.6%
- USA - 61,276 - 1.9%
- UK - 14,128 - 2.1%
- Europe - 159,681 - 2.1%
- Africa - 70,540 - 0.5%

## Effective CFR's by region...

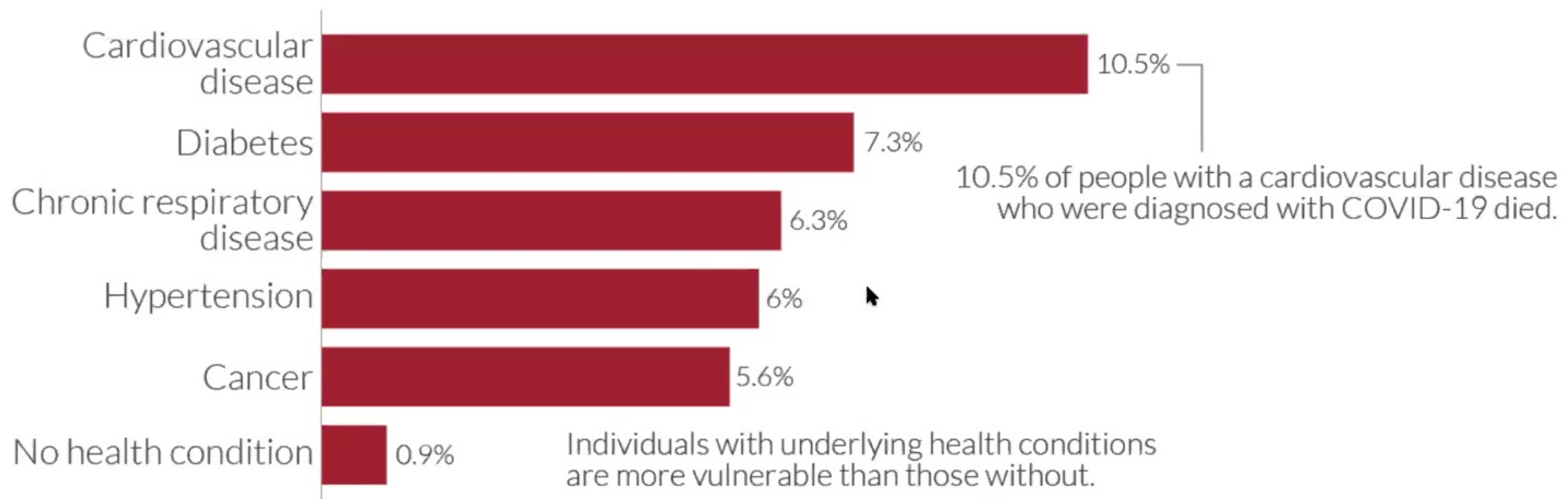
- Italy - 15,812 - 2.6%
- USA - 61,276 - 1.9%
- UK - 14,128 - 2.1%
- Europe - 159,681 - 2.1%
- Africa - 70,540 - 0.5%

For Africa the average IFR might therefore be 0.1%

# Coronavirus: early-stage case fatality rates by underlying health condition in China

Our World  
in Data

Case fatality rate (CFR) is calculated by dividing the total number of deaths from a disease by the number of confirmed cases. Data is based on early-stage analysis of the COVID-19 outbreak in China in the period up to February 11, 2020.



Data source: Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. *Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020*. China CDC Weekly.

[OurWorldinData.org](https://ourworldindata.org) – Research and data to make progress against the world's largest problems.

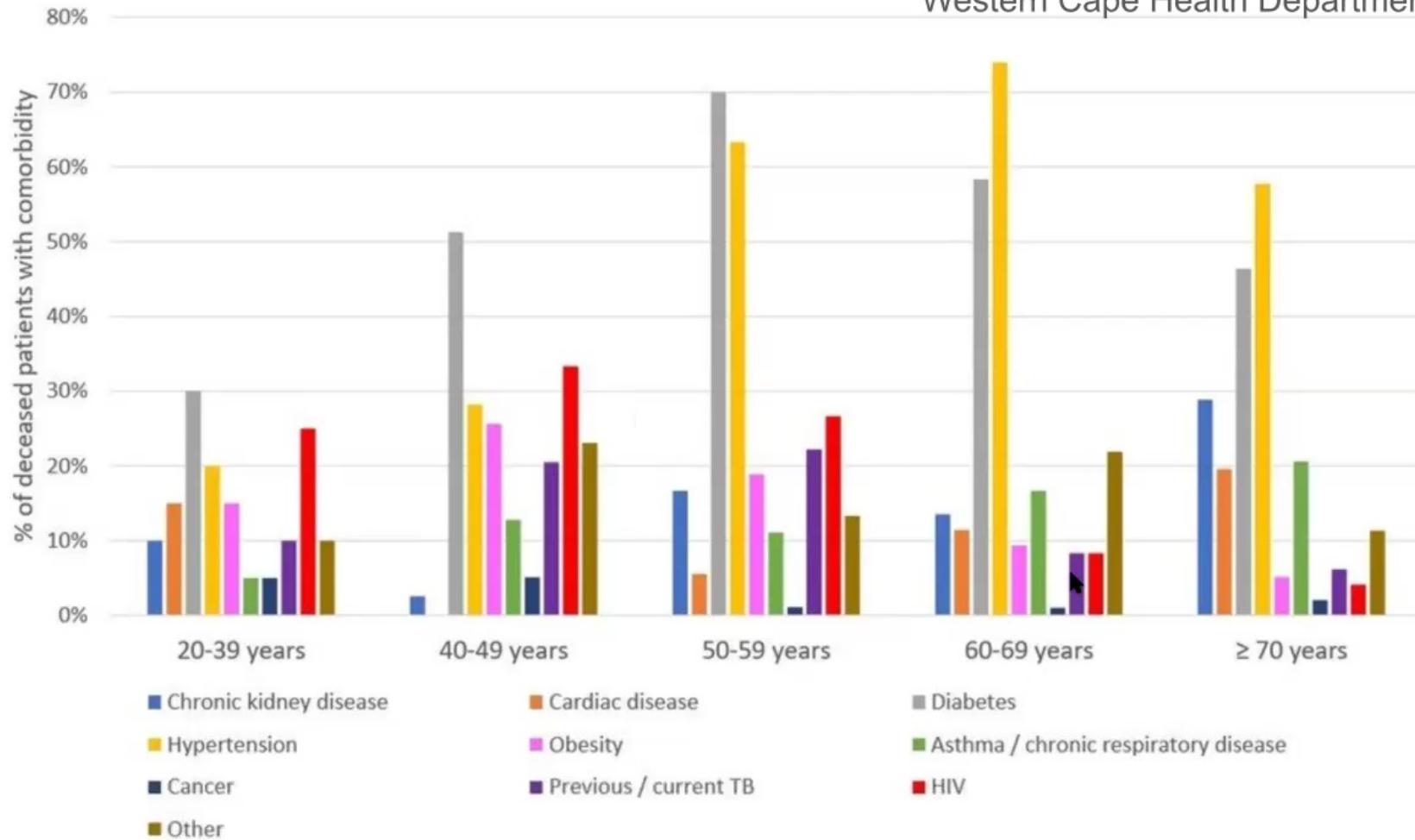
Licensed under [CC-BY](https://creativecommons.org/licenses/by/4.0/) by the authors.

# How will comorbidity affect South Africa?

- 8M HIV patients
- High hypertension
- High TB
- Malnutrition
- ...

Percent of deceased patients aged >20 years with different comorbidities by age group at death

Western Cape Health Department



So how many people will be infected?

## Key Variables: $R_0$ and $R_t$

- $R_0$  is the reproductive number initially (when there are no interventions)

For COVID-19 we have something like:  $2 < R_0 < 6$

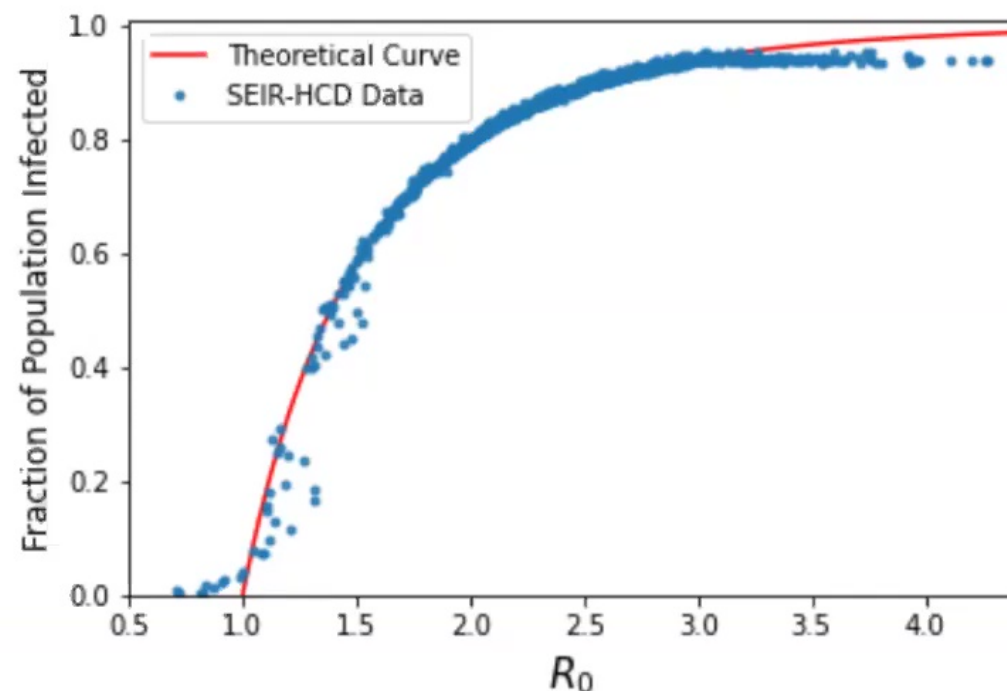
- $R_t$  is the reproductive number at time  $t$  (usually after interventions or after herd-immunity kicks in). For countries around the world we see  $0.5 < R_t < 2$



## Key Variables: $R_0$ and $R_t$

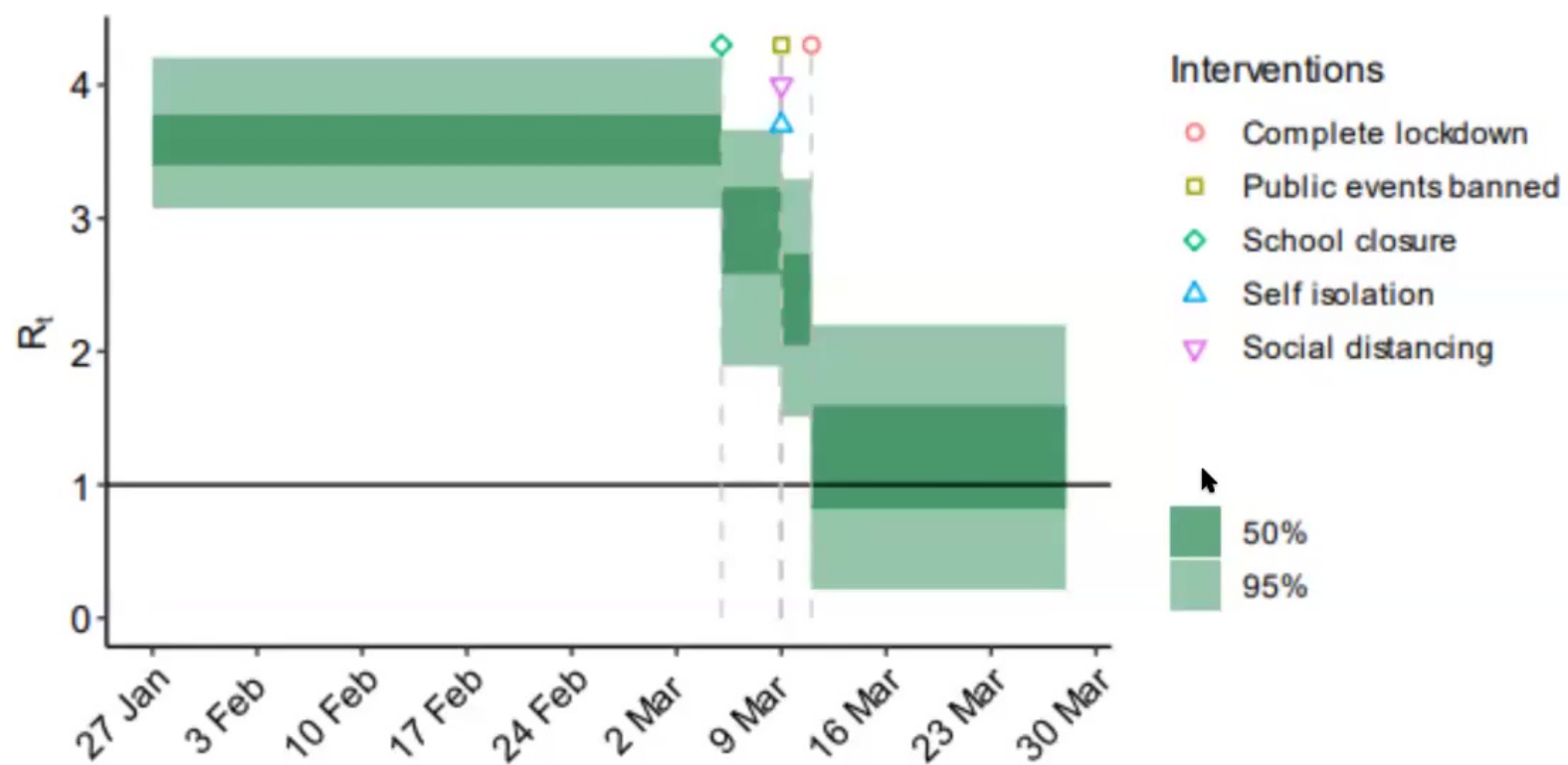
- If  $R_0 > 1$  then growth will be exponential until herd immunity fraction is infected:  $1 - 1/R_0$
- Actual number of infected is significantly larger though...

e.g. 84% for  $R_0 = 2.2$



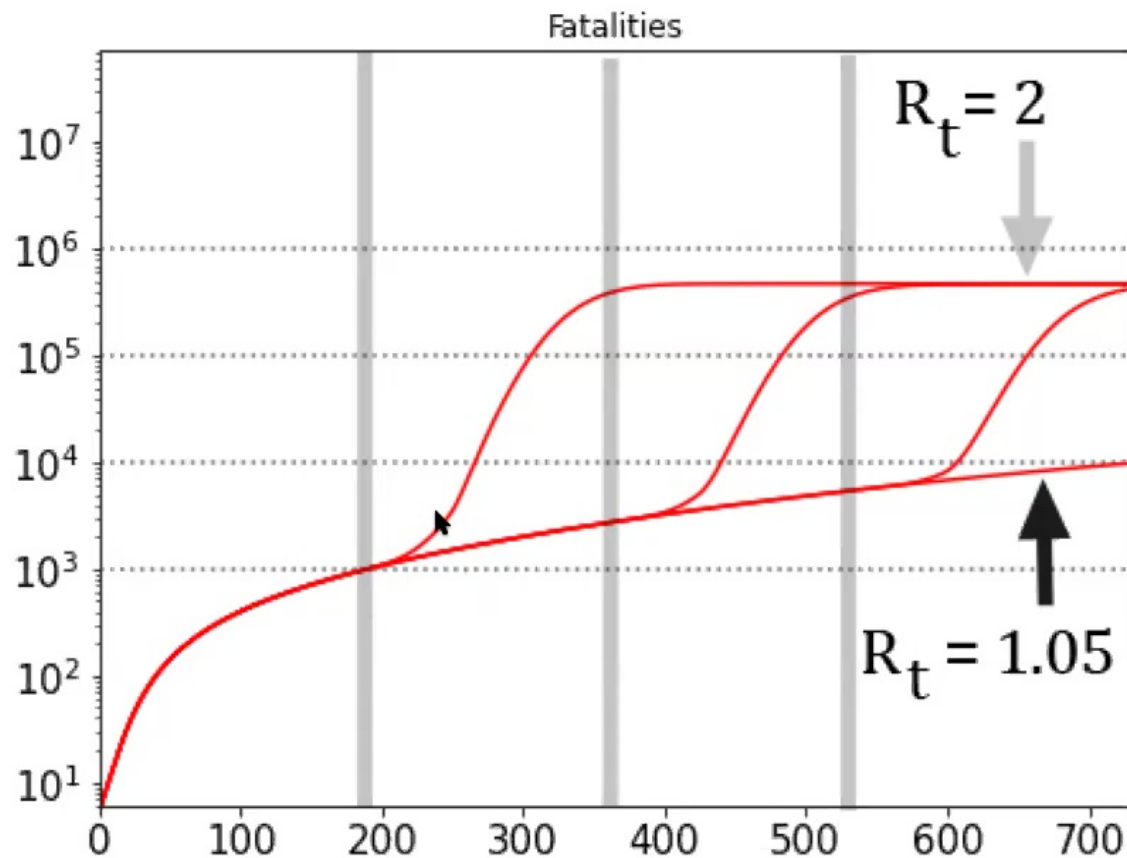
**Question:** is there a simple predictor of final attack rate for a time-varying  $R_t$ ?  
(c.f. Dark energy and  $w(z)$ )

## Key Variables: $R_t$ - ITALY



Imperial Covid Report 13

## A bitter pill to swallow...



<https://www.linkedin.com/pulse/long-term-covid-19-forecasts-south-africa-bruce-bassett/>



Essentially, all models are wrong, but some are useful.

(George E. P. Box)

izquotes.com

# Going beyond mean-field

- We know there are highly connected people who will have  $R \gg R_0$ . Won't this completely change our mean-field predictions? (Especially if we consider fitting to early data around the world)

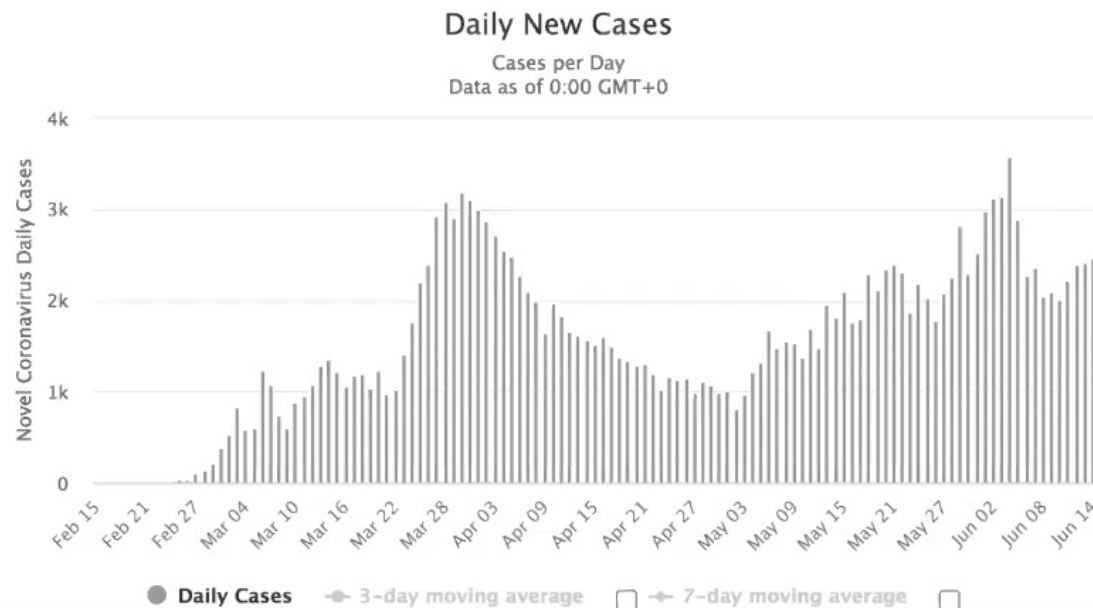
# Going beyond mean-field

- We know there are highly connected people who will have  $R \gg R_0$ . Won't this completely change our mean-field predictions? (Especially if we consider fitting to early data around the world)
- We have a problem very similar to the averaging/backreaction problem in cosmology. Does averaging a full inhomogeneous model give the same results as a homogeneous model with averaged parameters? ( $R_0$ )
- Is it possible that herd immunity will kick in at 20%? Could it already have kicked in in Spain, Italy, Sweden, Iran, NYC?

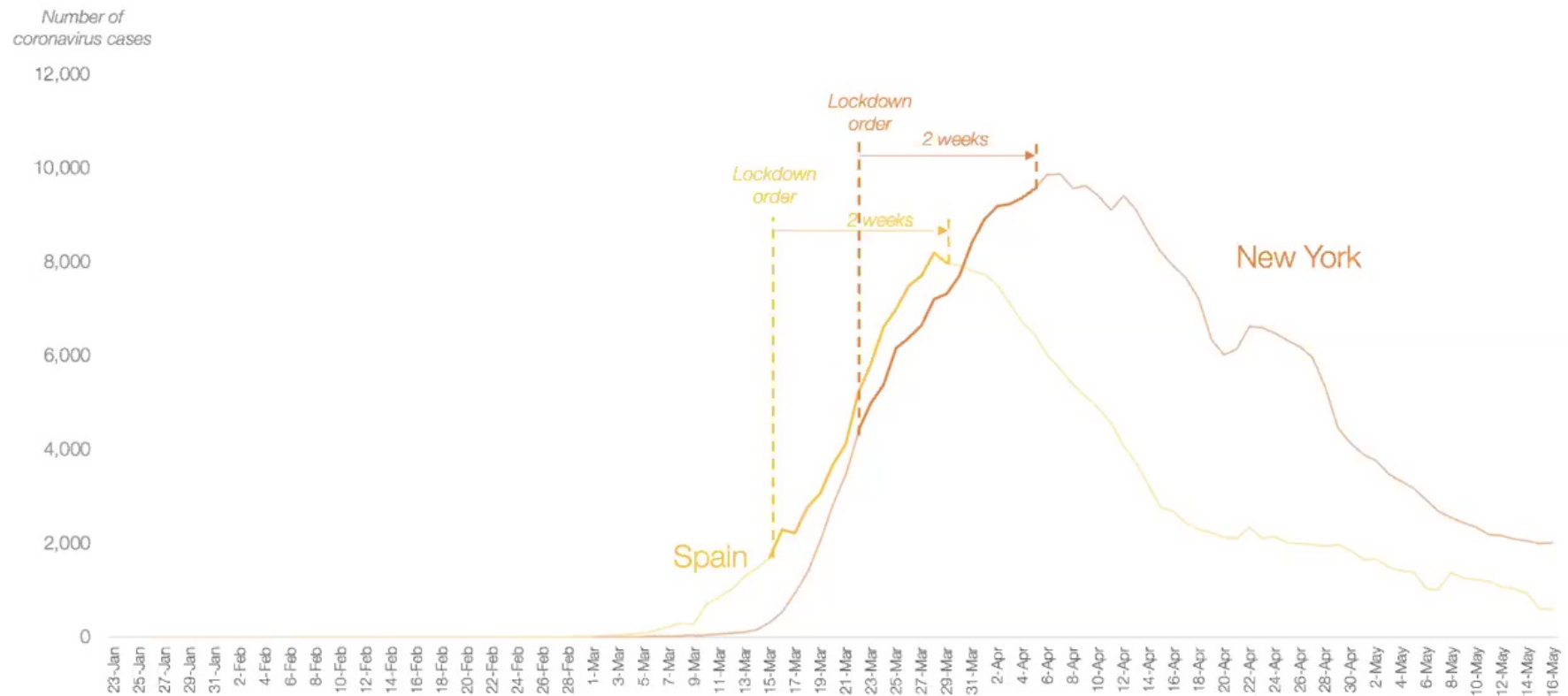
# Places that are opening up...

- Spain and Italy have already started opening up and have not seen second waves, but Iran has....

Daily New Cases in Iran



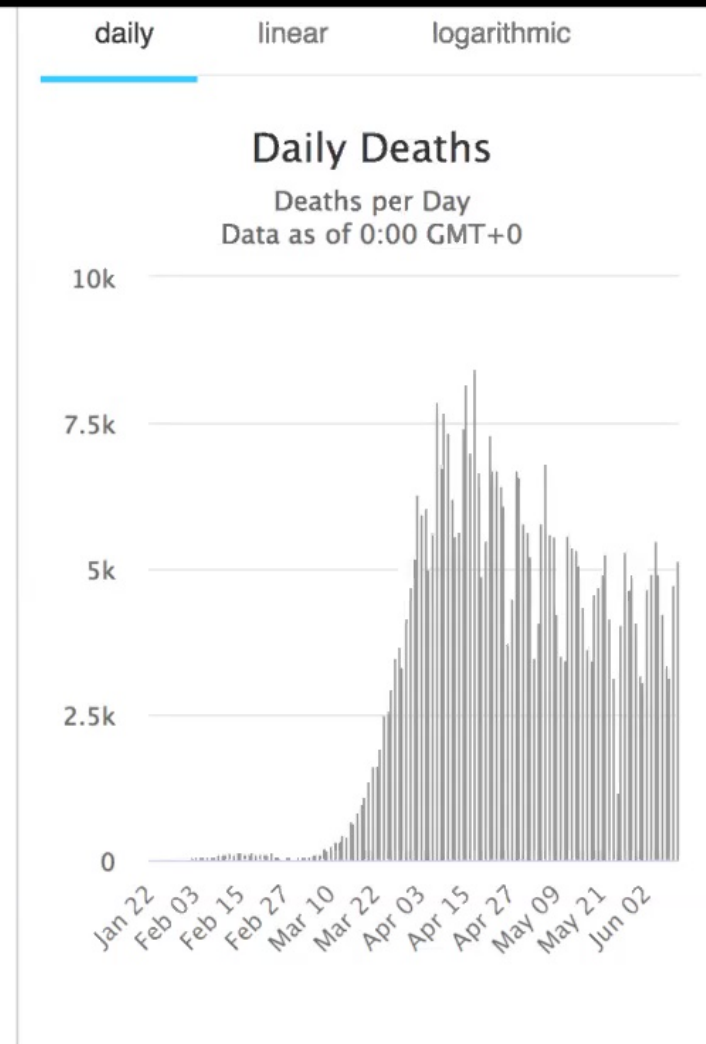
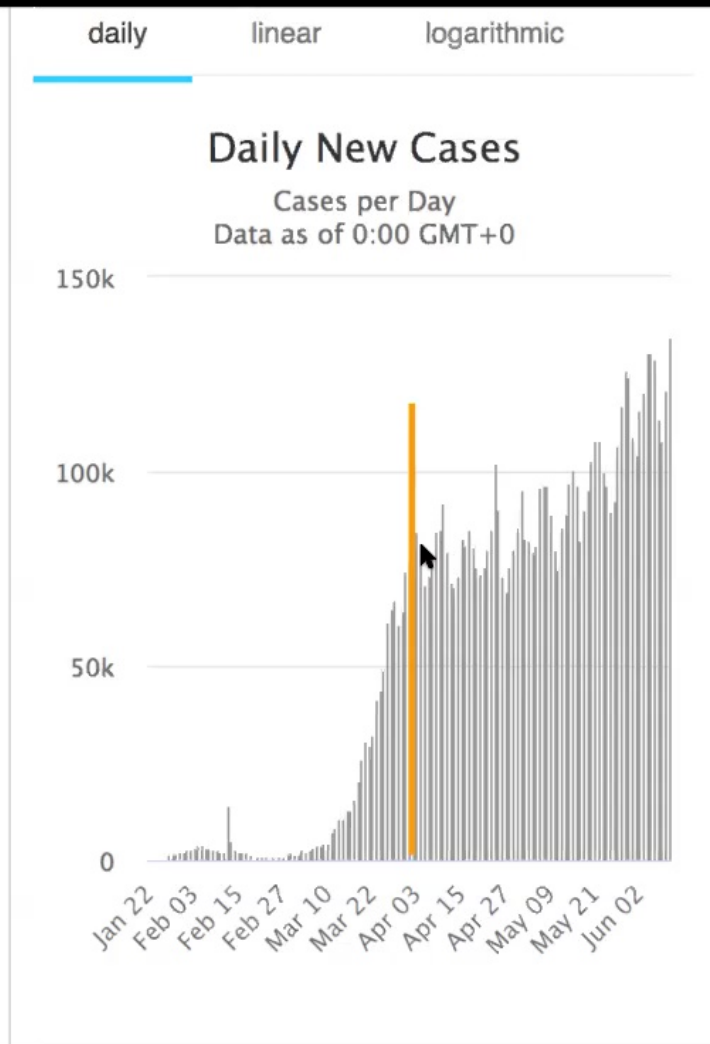
## Chart 7.b: Days Between Heavy Lockdown and Peak of Coronavirus Cases

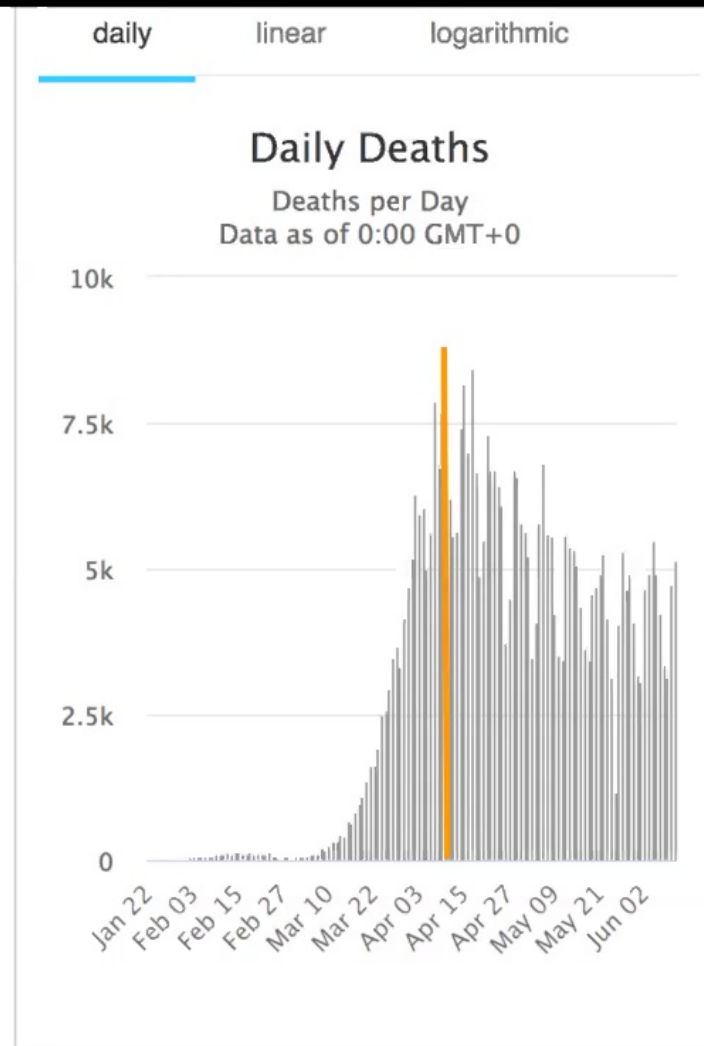
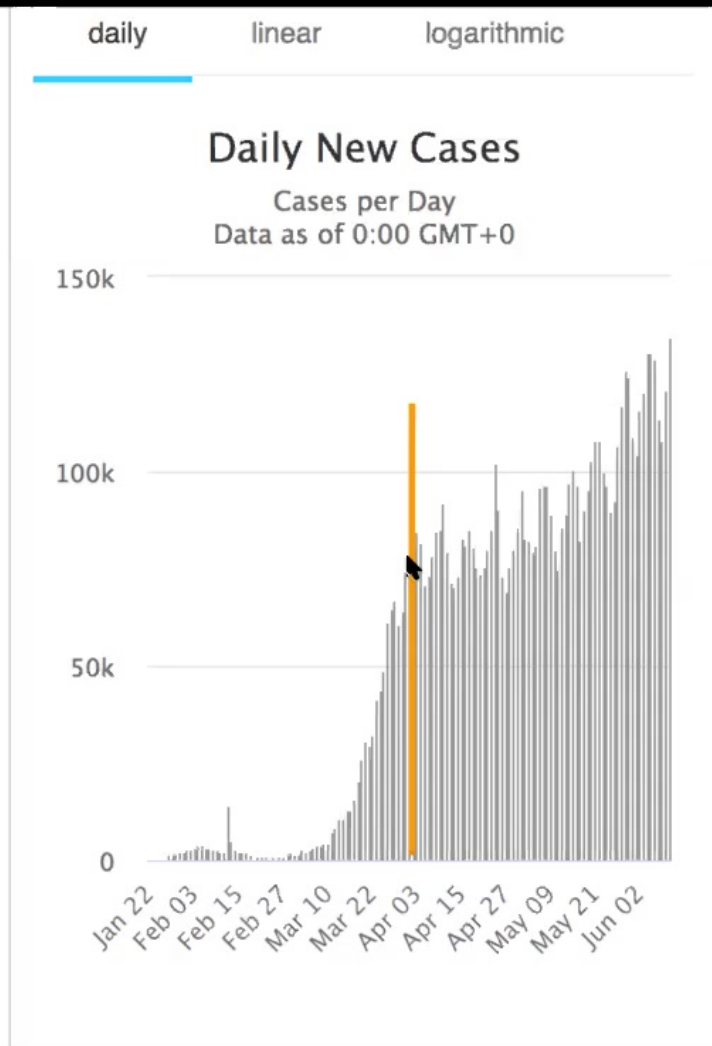


Note: 7-day averages for all countries, centered around the actual date (so for example March 10<sup>th</sup> is the average of March 7<sup>th</sup> through March 13<sup>th</sup>). For China, France and Spain, which have had statistical artifacts on their case curves (eg, adding nursing home cases for France, adding seroprevalence cases and withdrawing them for Spain), I eliminated the effects by smoothing the curve around the statistical artifact and applying the gap in cases to the entire curve. The curve overall shape doesn't change, except for eliminating the artifacts.

Source: Tomas Pueyo Analysis from Johns Hopkins data for countries via GitHub ([https://github.com/CSSEGISandData/COVID-19/blob/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series/time\\_series\\_covid19\\_confirmed\\_global.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv)), and New York Times data for New York (<https://github.com/nytimes/covid-19-data/blob/master/us-states.csv>)

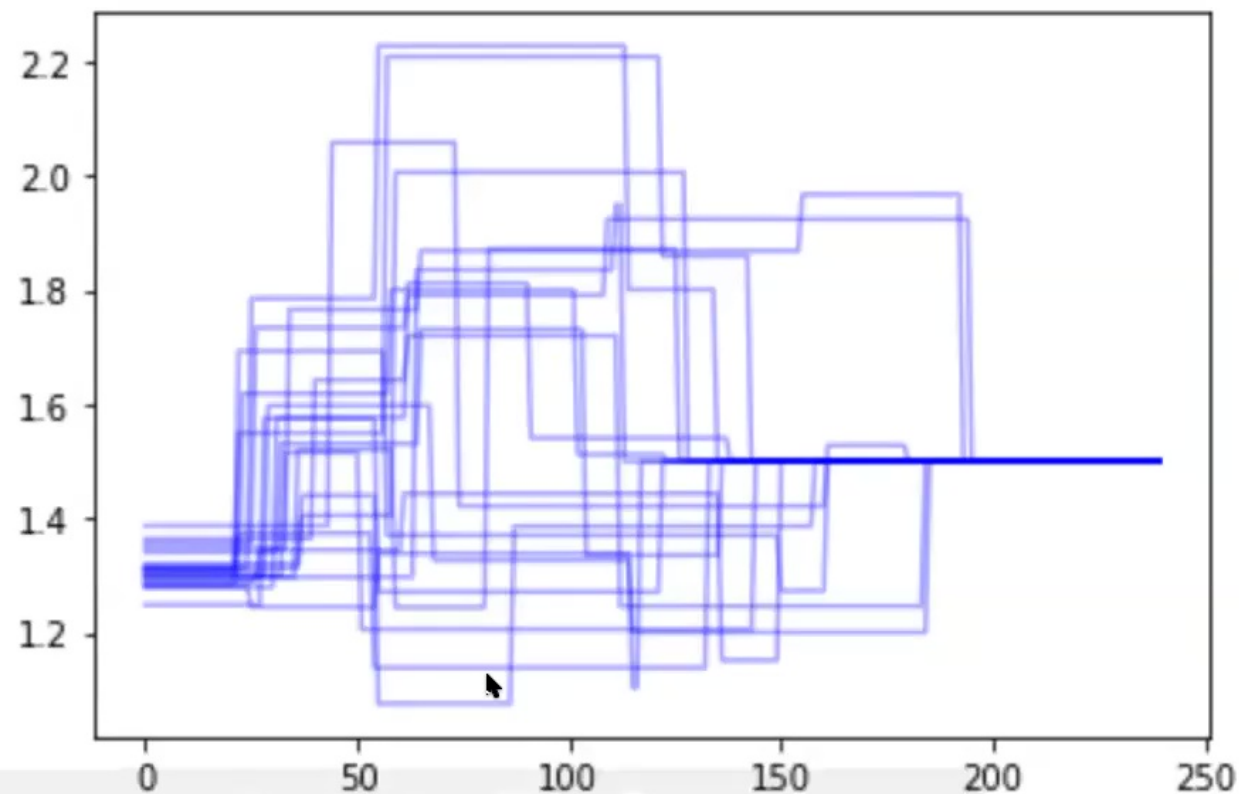






<https://www.worldometers.info/coronavirus/>

## Some example $R_t$ curves over time...

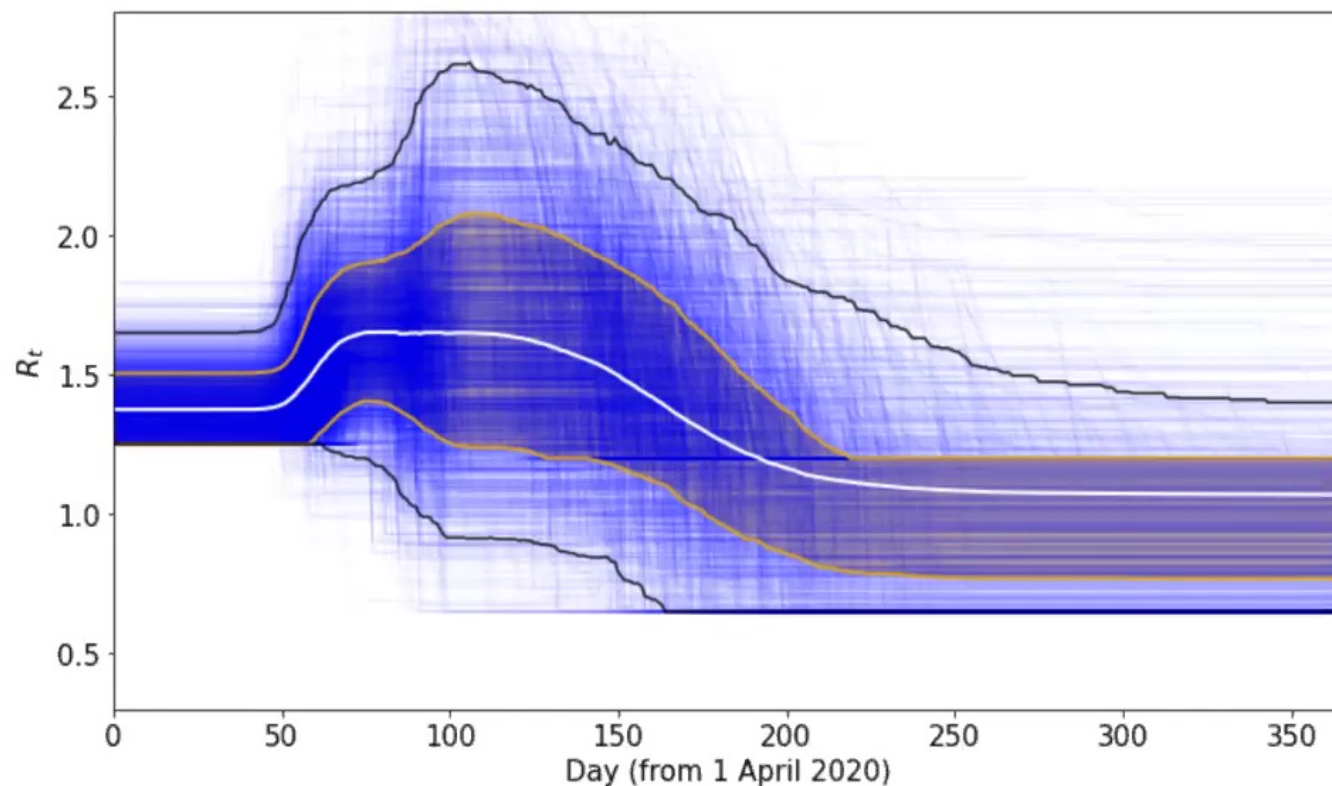


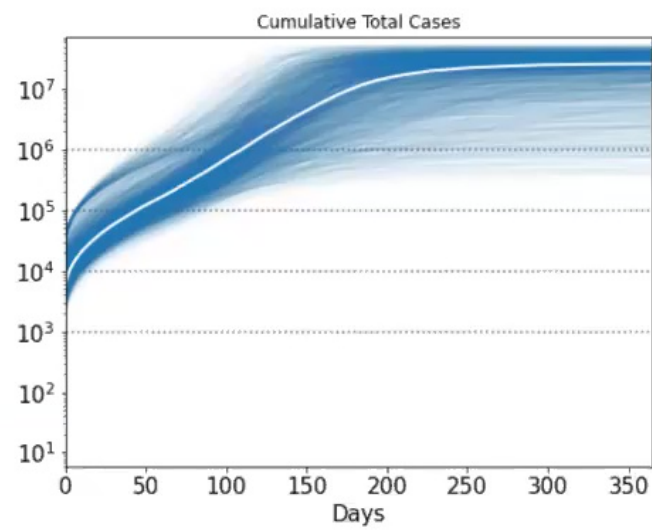
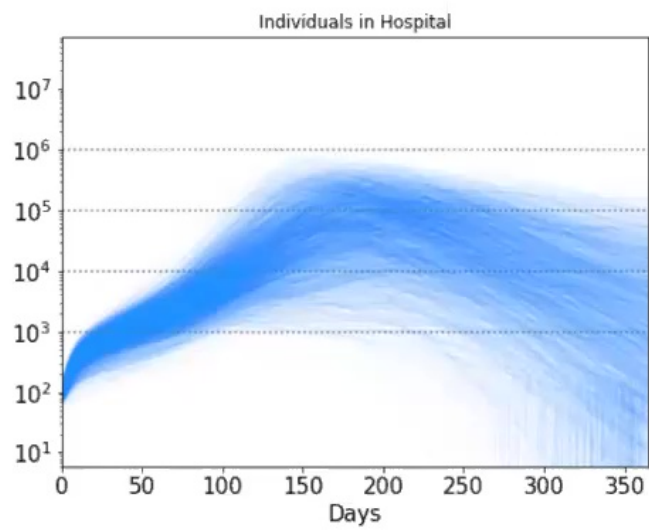
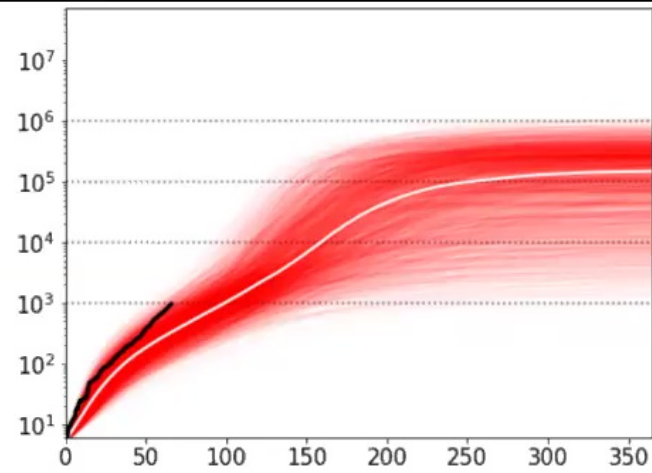
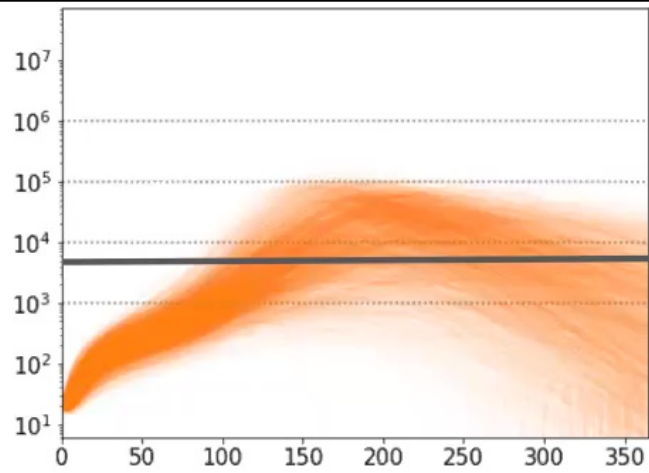
# Rt over time...

How much effect will:

- Climate
- Masks
- Compliance
- Contact tracing
- Quarantine
- Gov policy
- Multiple strains

Have?



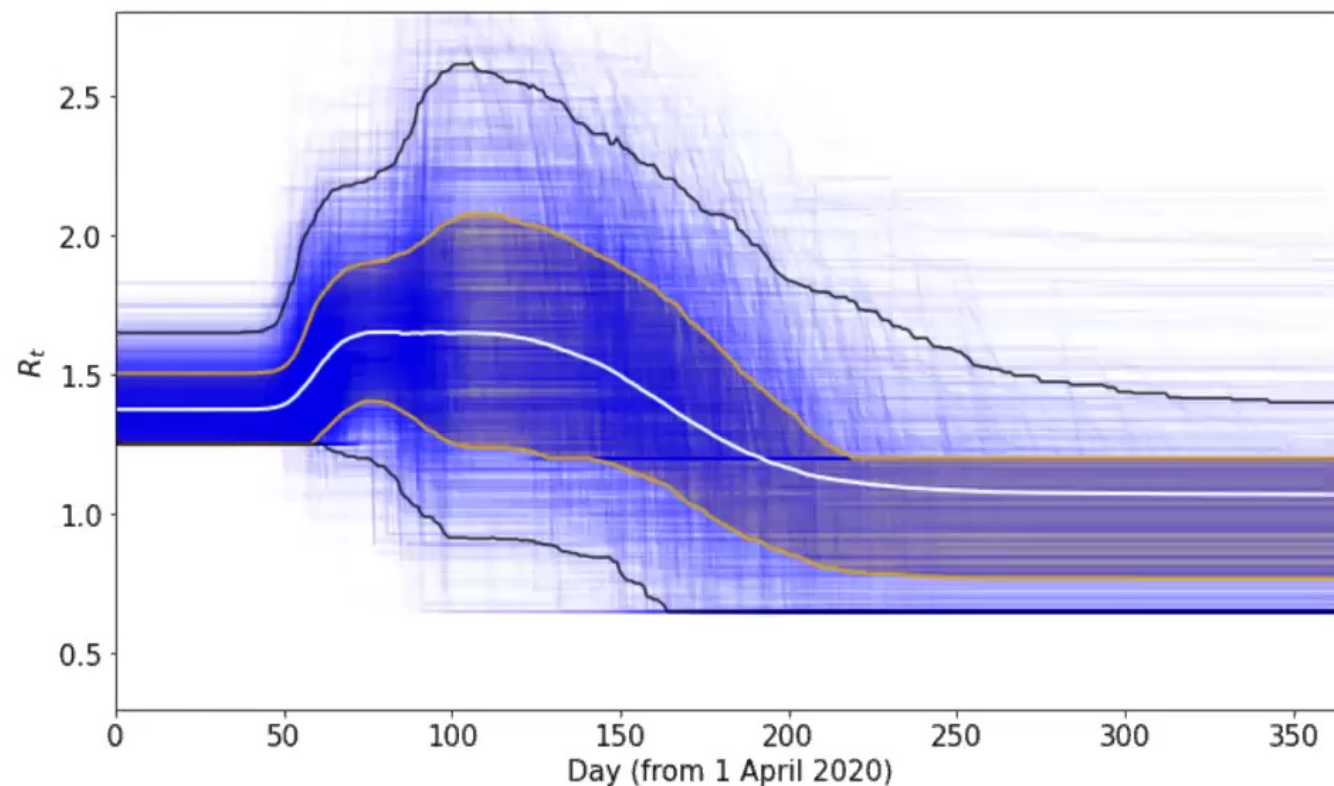


# Rt over time...

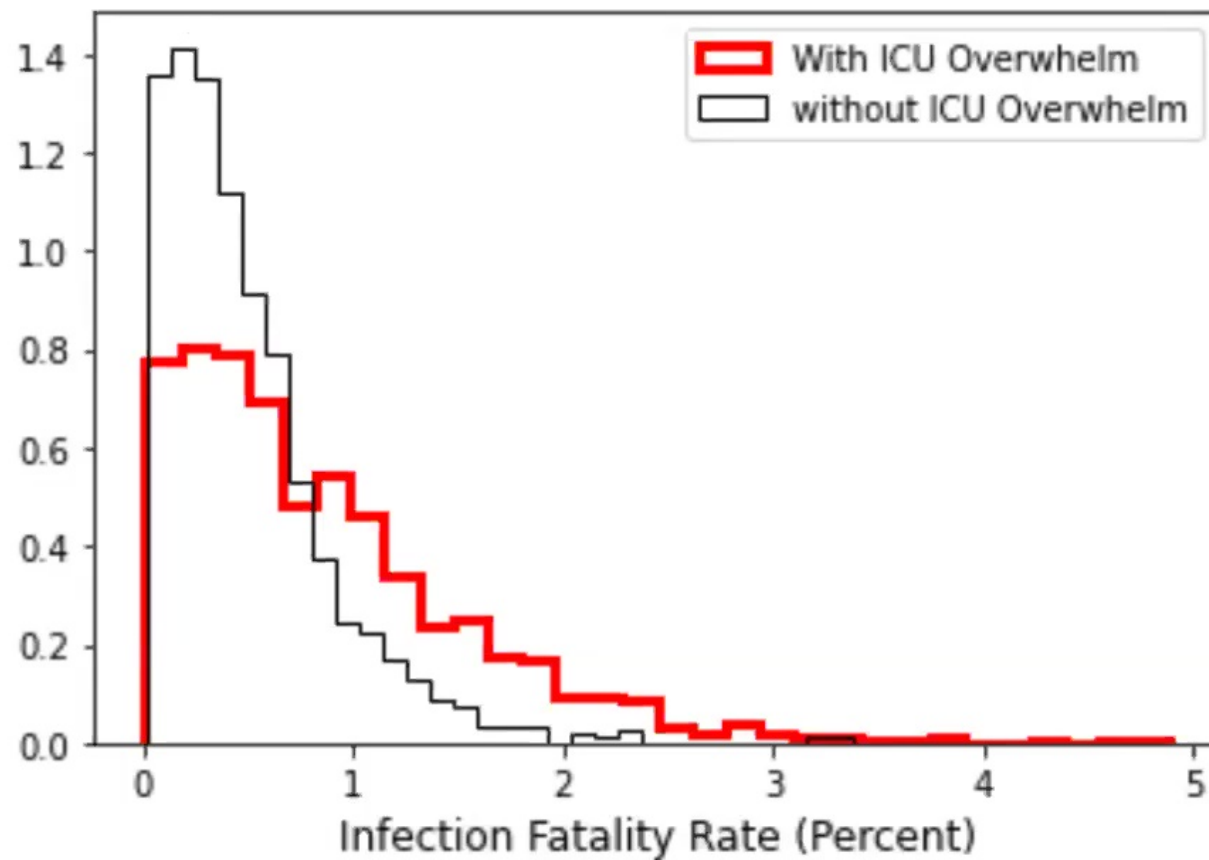
How much effect will:

- Climate
- Masks
- Compliance
- Contact tracing
- Quarantine
- Gov policy
- Multiple strains

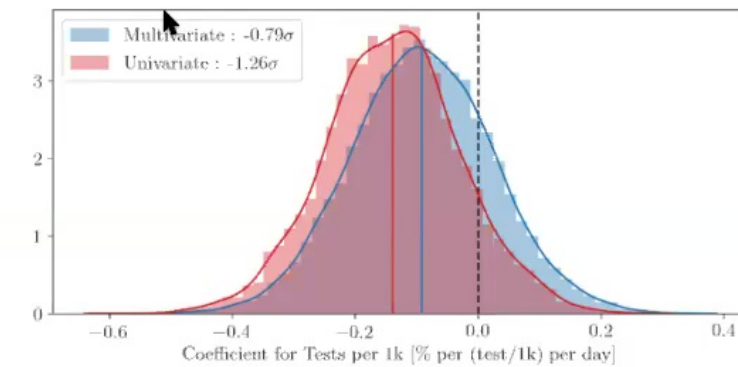
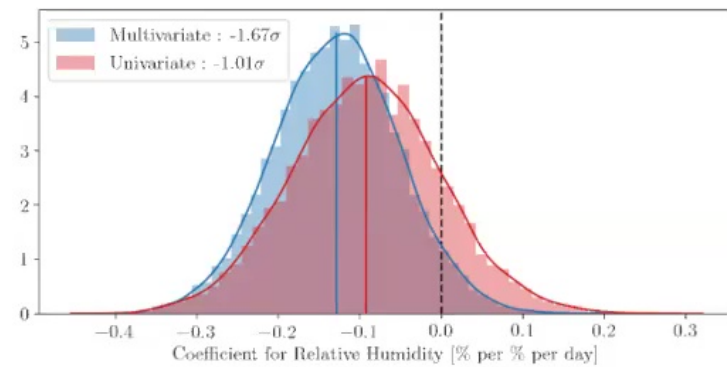
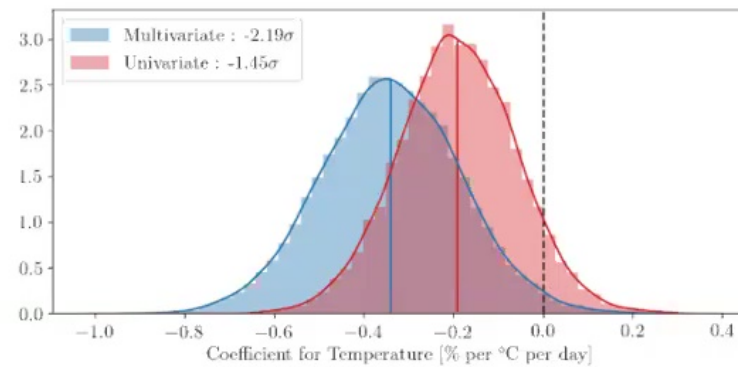
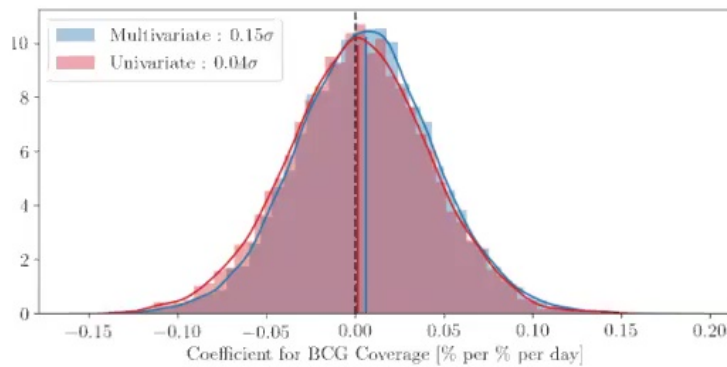
Have?



## Effect of ICU Overwhelm for (30%, 75% )

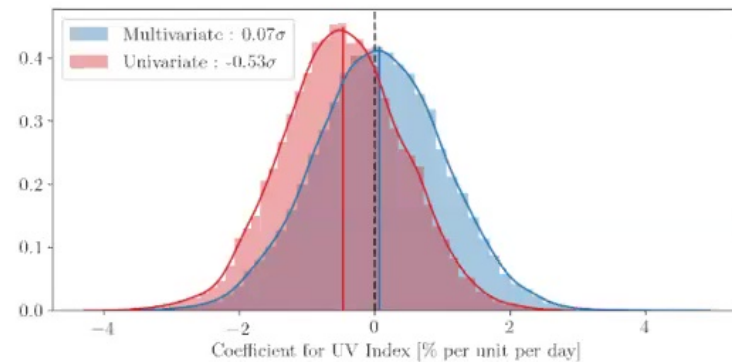
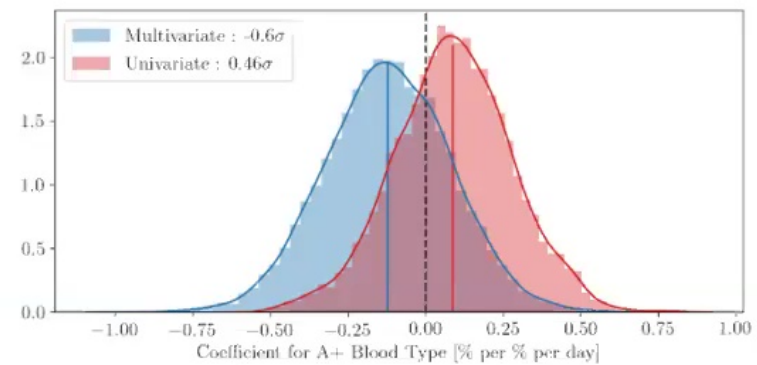
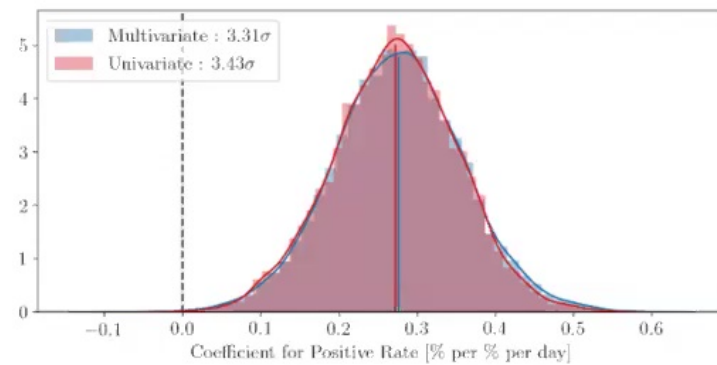


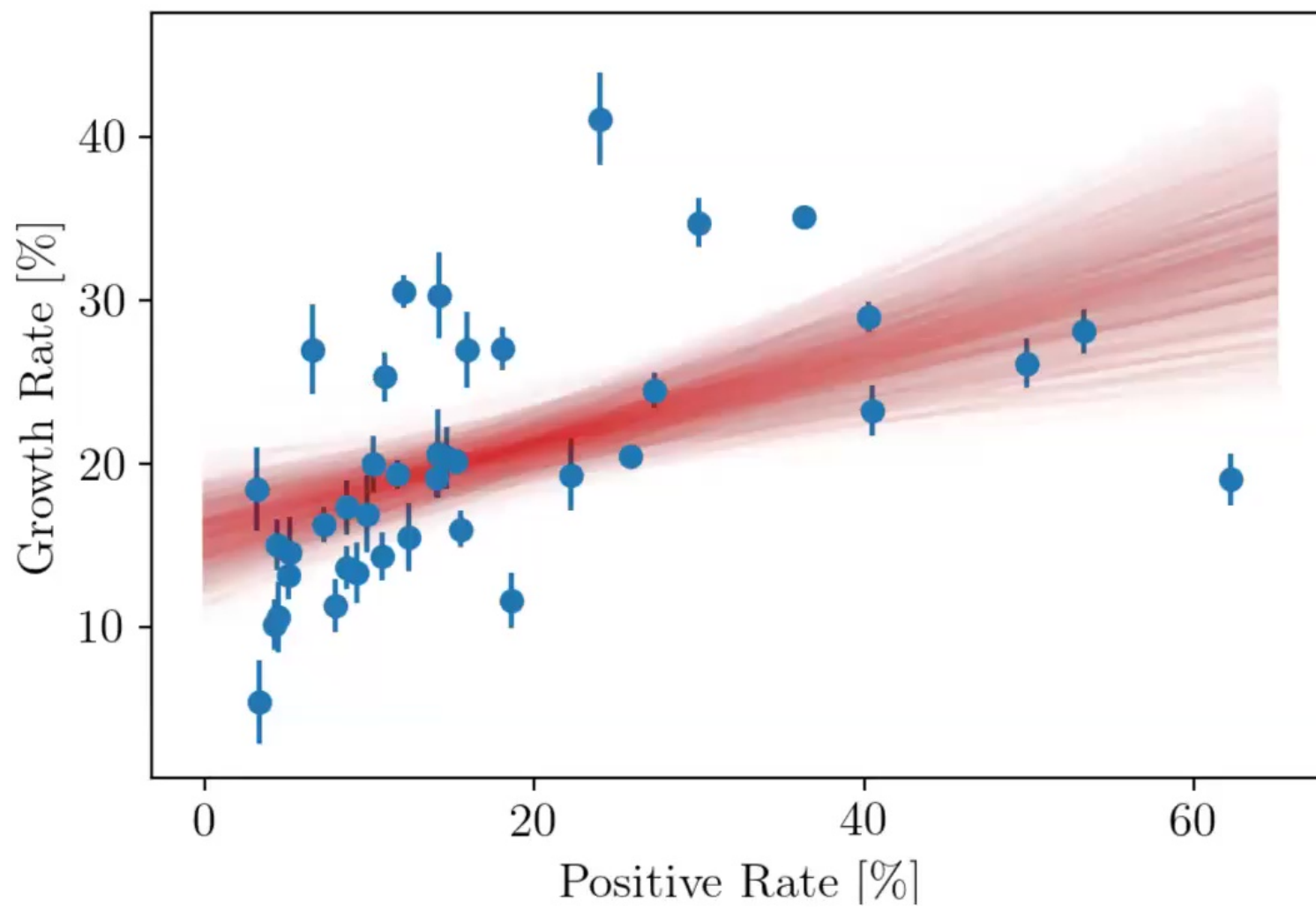
# BCG, Climate and Testing...





# Testing, blood and UVindex





# The Takeaways...

Precise answers are not possible currently. Model predictions are **exponentially** sensitive to **assumptions** about:

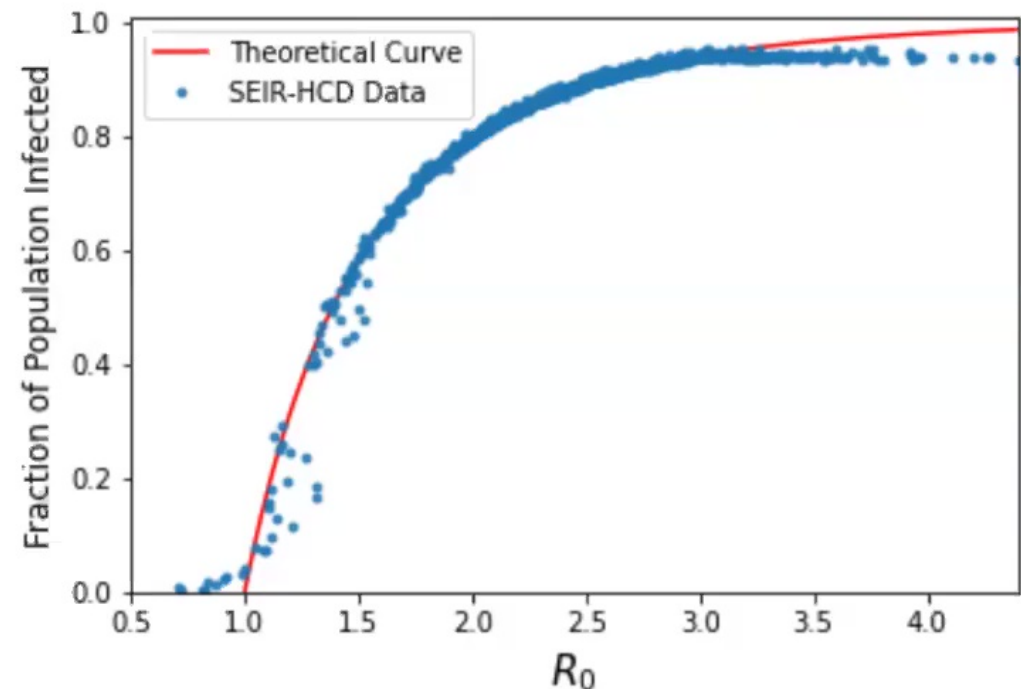
- Virus parameters (infection time, climate dependence, IFR...)
- Future government policy and community compliance
- Effectiveness of interventions in the face of reopening the economy (contact tracing, quarantining, testing) etc....

Don't rely on predictions of single models or when people claim that "63490 people will die..."

## Conclusions II

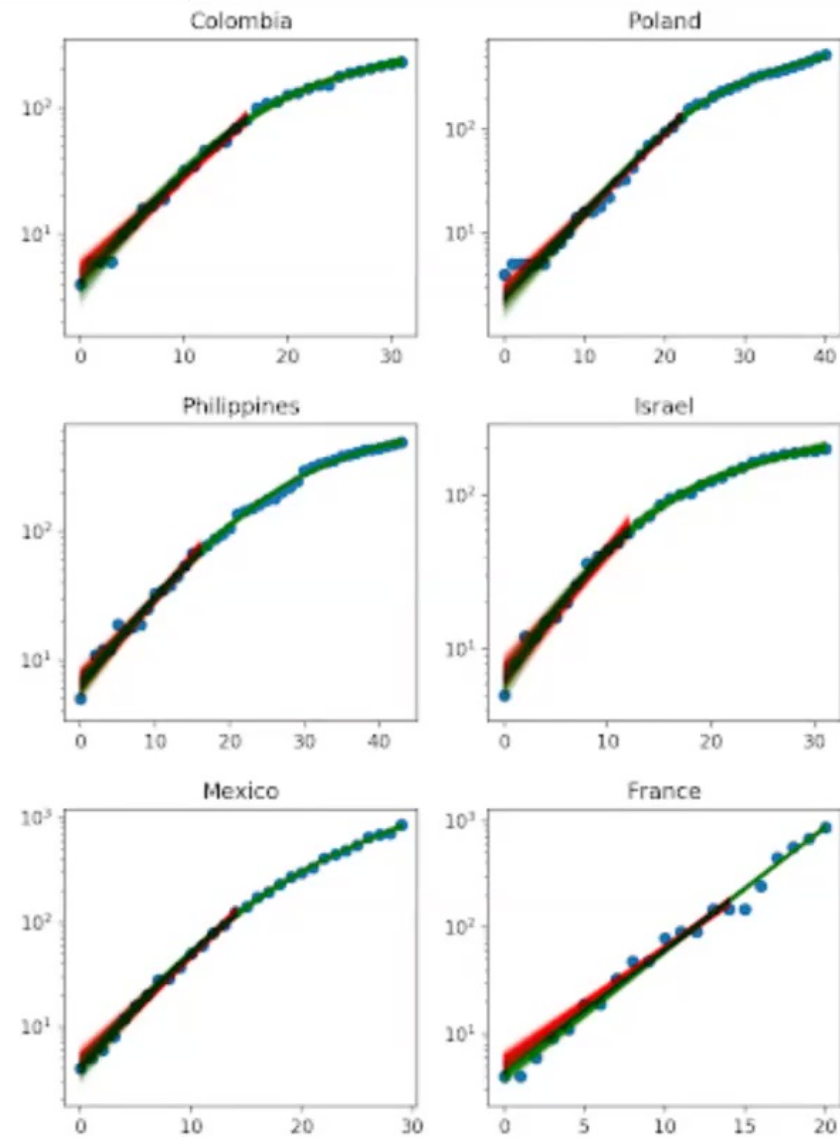
- Covid-19 is definitely worse than the flu (IFR > 0.18%).
- No evidence for mitigating effects of weather/BCG or blood (yet)
- Lots of mysteries still to solve...

**Question:** what is the optimal framework for deciding about ending lockdown given all the uncertainties?



<https://www.linkedin.com/pulse/long-term-covid-19-forecasts-south-africa-bruce-bassett/>

# Fitting the death data





# Fitting the death data

We don't fit infection counts since they are so affected by testing. We hope that deaths will be more reliable...

