Statistical Challenges of Modelling Covid-19.

Peter Dolton

(University of Sussex and NIESR)

National Institute Economic Review (2021), 1–37
doi:10.1017/nie.2021.22

NIER CAMBRIDGE UNIVERSITY PRESS

ARTICLE

THE STATISTICAL CHALLENGES OF MODELLING COVID-19

Peter Dolton[1,2]*

[1]University of Sussex, Brighton, United Kingdom
[2]National Institute of Economic and Social Research, London, United Kingdom
*Corresponding author. Email: p.dolton@niesr.ac.uk

(Received 14 April 2021; revised 25 June 2021; accepted 01 July 2021)

In 2020–2021, the world has been gripped by a pandemic that no living person has ever known. The coronavirus pandemic is undoubtedly the greatest challenge the world has faced in over a generation. The imperative of statistical modelling is not only to manage the short-run crisis for the health services, but also to explain the pandemic's course and establish the effectiveness of different policies, both non-pharmaceutical and with vaccines. This difficult task has been undertaken by the epidemiologists and others in the face of measurement data problems, behavioural complications and endogeneity issues. This paper proposes a simple taxonomy of the alternative different models and suggests how they may be used together to overcome limitations. This perspective may have important implications for how policy-makers cope with future waves or strains in the current pandemic, or future pandemics.

Keywords: COVID-19; econometric modelling.

JEL codes: C54; I18; C32; C33.

The statistical modelling of the Covid-19 pandemic is an extremely complex task.

”it’s like trying to build the ship whilst sailing it”

Prof Jonathan Van Tam

Daily new confirmed COVID-19 deaths per million people

Shown is the rolling 7-day average. Limited testing and challenges in the attribution of the cause of death means that the number of confirmed deaths may not be an accurate count of the true number of deaths from COVID-19.

Our World in Data

LINEAR | LOG

Source: Johns Hopkins University CSSE COVID-19 Data

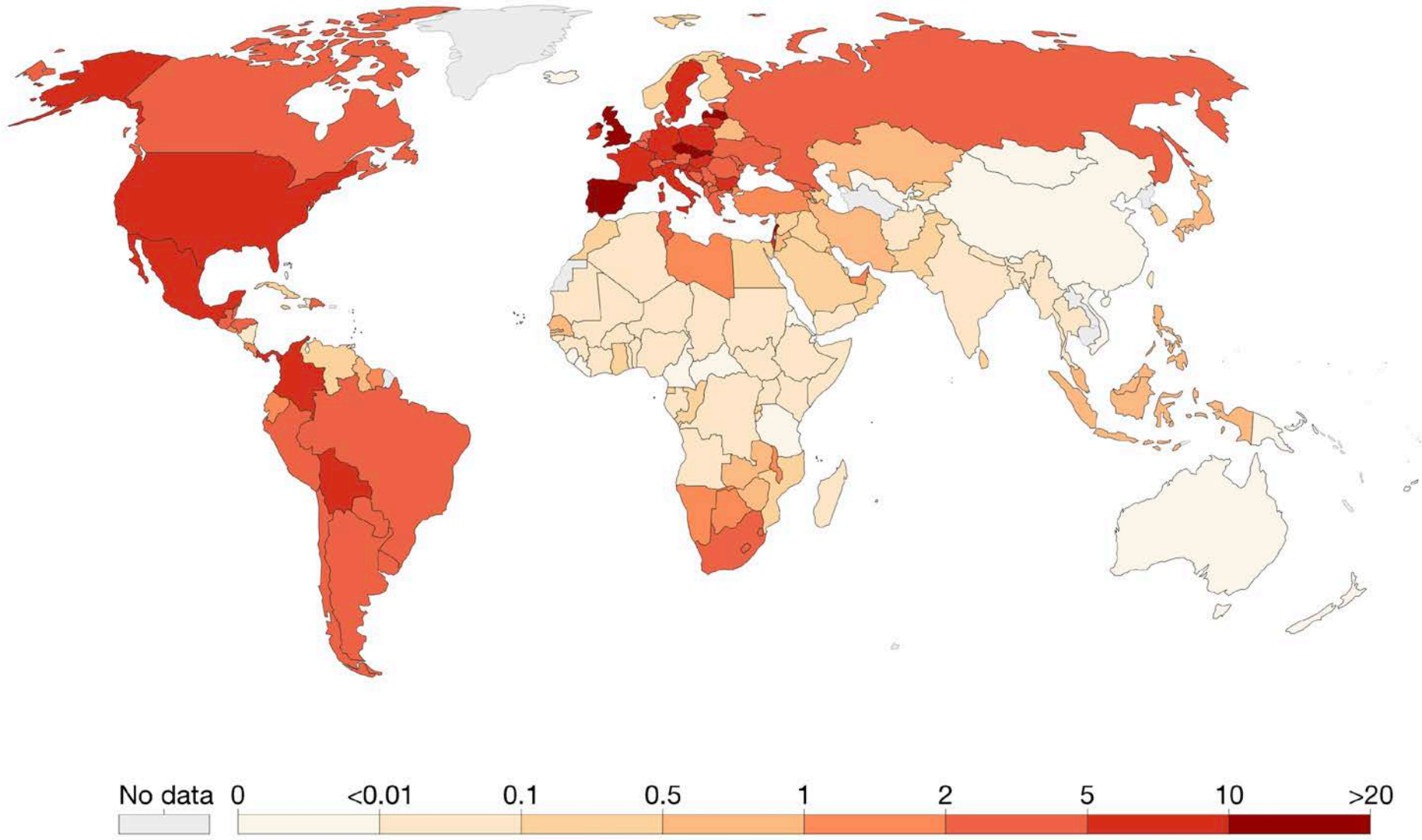
CC BY

Jan 28, 2020 — Oct 25, 2021

Daily new confirmed COVID-19 deaths per million people, Feb 10, 2021

Our World in Data

Shown is the rolling 7-day average. Limited testing and challenges in the attribution of the cause of death means that the number of confirmed deaths may not be an accurate count of the true number of deaths from COVID-19.



Source: Johns Hopkins University CSSE COVID-19 Data – Last updated 11 February, 10:03 (London time)

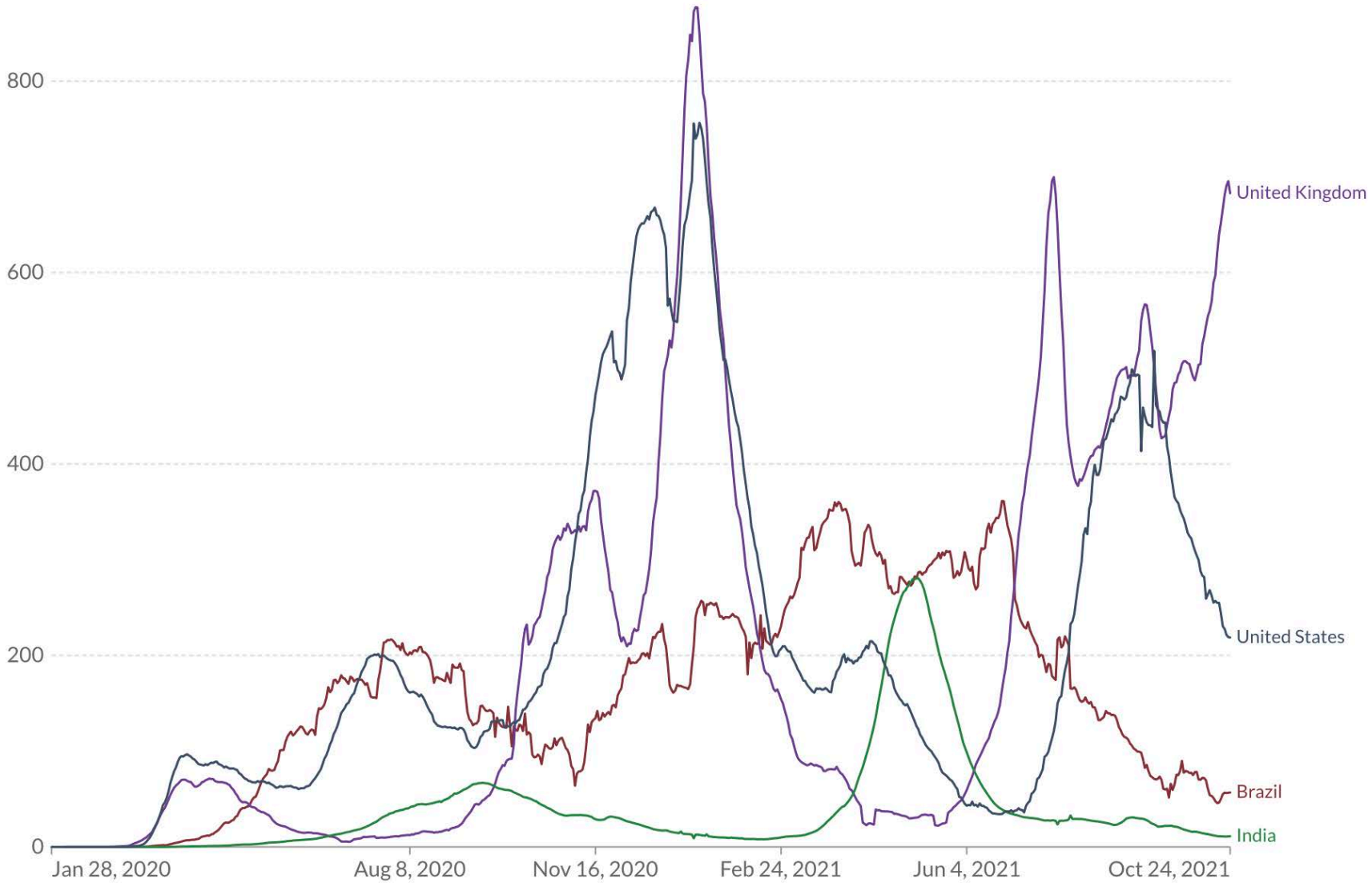
CC BY

Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World in Data

LINEAR LOG



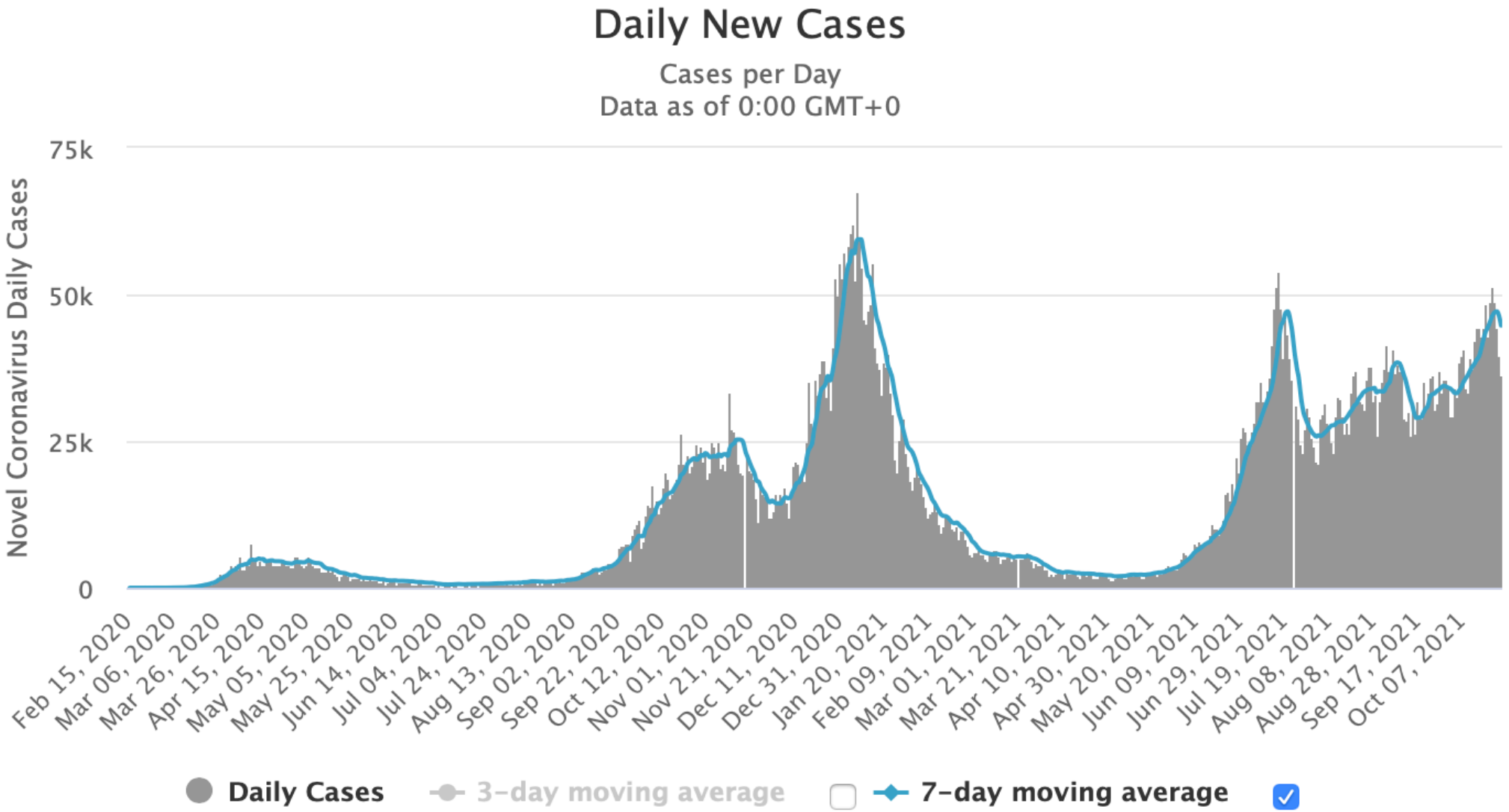
Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

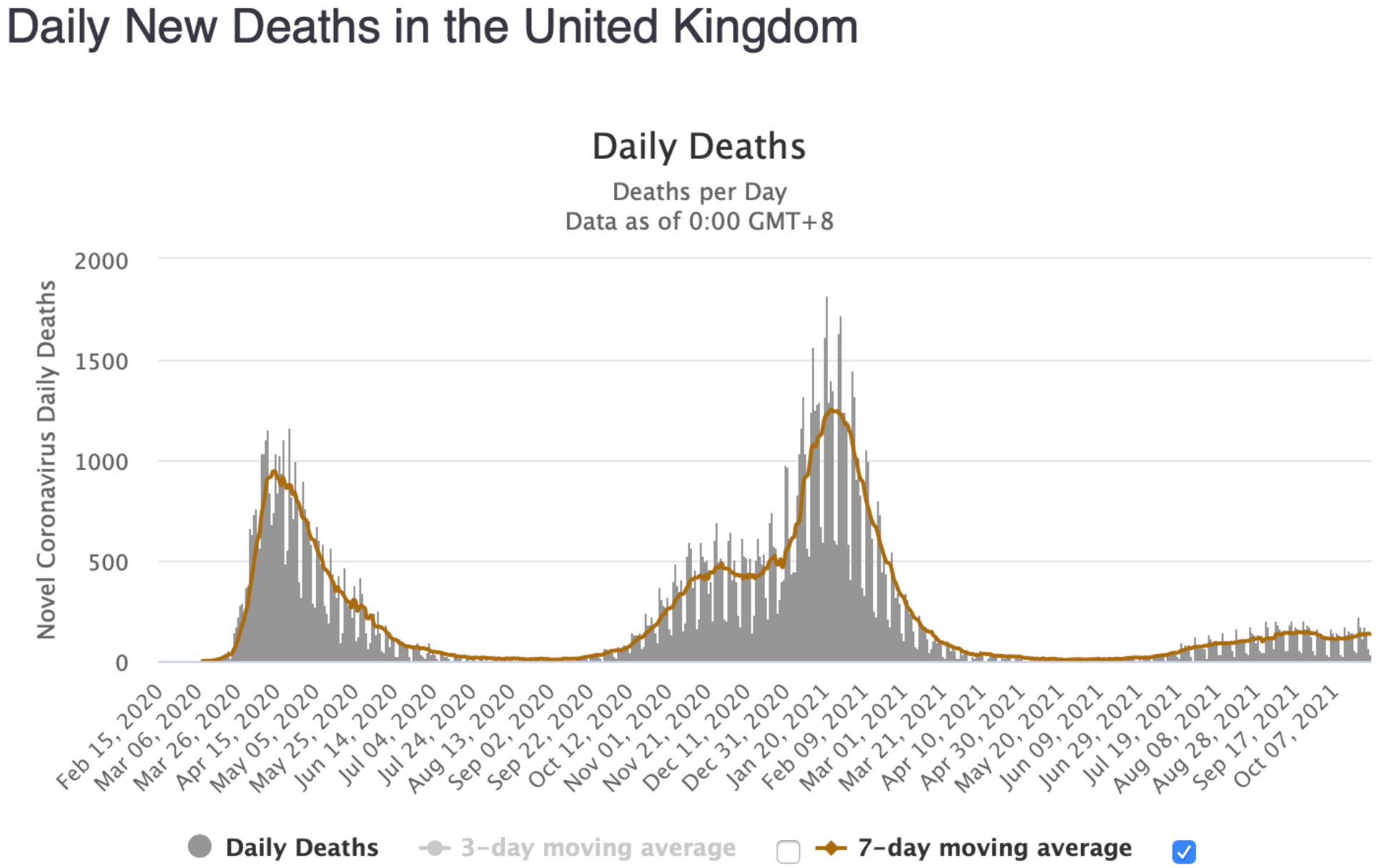
Jan 28, 2020 Oct 25, 2021

Motivation

Daily New Cases in the United Kingdom



Motivation


Daily New Deaths in the United Kingdom
Daily Deaths
Deaths per Day
Data as of 0:00 GMT+8
Novel Coronavirus Daily Deaths
2000
1500
1000
500
0
Feb 15, 2020
Mar 06, 2020
Mar 26, 2020
Apr 15, 2020
May 05, 2020
May 25, 2020
Jun 14, 2020
Jul 04, 2020
Jul 24, 2020
Aug 13, 2020
Sep 02, 2020
Sep 22, 2020
Oct 12, 2020
Nov 01, 2020
Nov 21, 2020
Dec 11, 2020
Dec 31, 2020
Jan 20, 2021
Feb 09, 2021
Mar 01, 2021
Mar 21, 2021
Apr 10, 2021
Apr 30, 2021
May 20, 2021
Jun 09, 2021
Jun 29, 2021
Jul 19, 2021
Aug 08, 2021
Aug 28, 2021
Sep 17, 2021
Oct 07, 2021
Daily Deaths
3-day moving average
7-day moving average

Roadmap of Paper

- Why Model? Rationale for Modelling
- Data Sources
- Modelling Approaches : A Taxonomy
- Statistical Data &
  - Measurement Error
  - Endogeneity
- Endogeneity & How do we think about Policies
- SIR Epidemiological Models
- Other Models
- Comparing Models: Traffic Lights
- Conclusion

Measurement Error

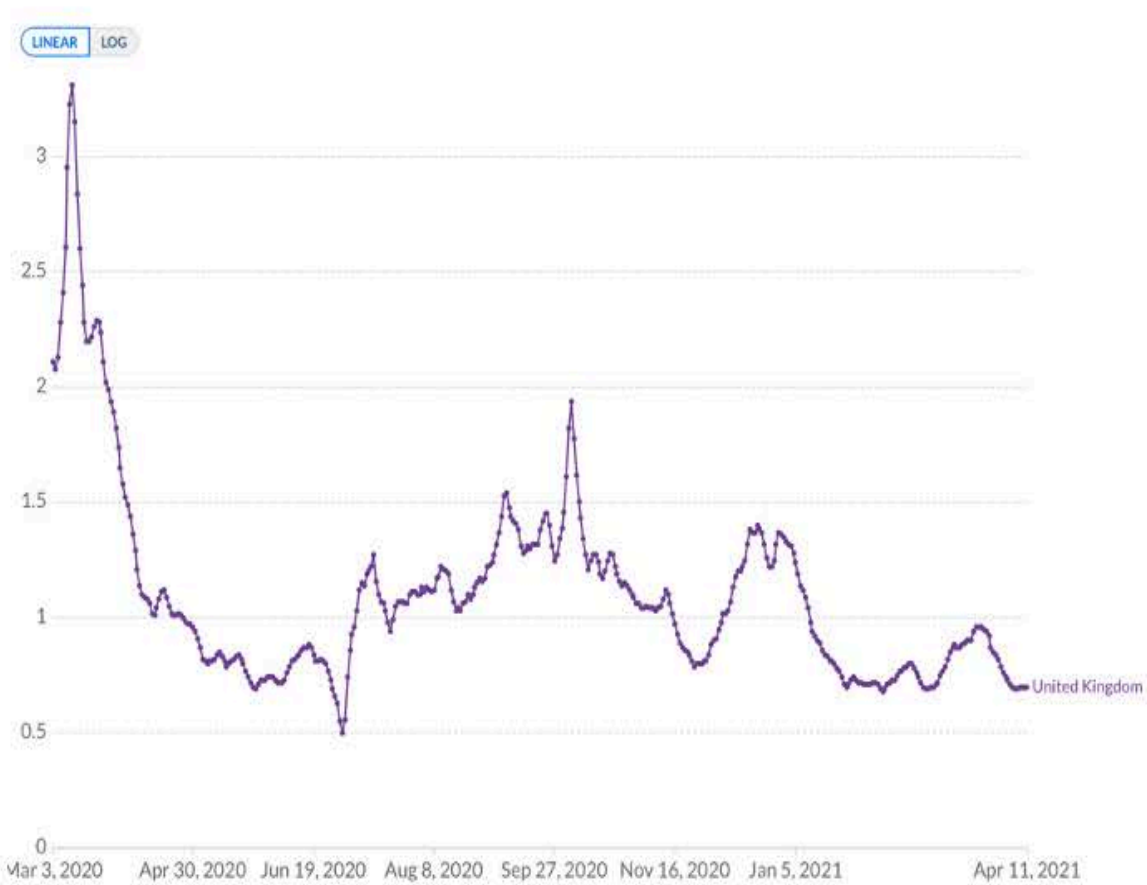
- What is a death from Covid – death certificate issues.
- What is a case? – based on a test or not?
- How many cases - if not enough tests.
  - How many tests are done
  - Who are tested
  - How are you counting them - people can have more than one test
- In comparing data across countries there is no uniformity.
- Endogeneity of reporting reason to believe that Iran was mis-reporting.

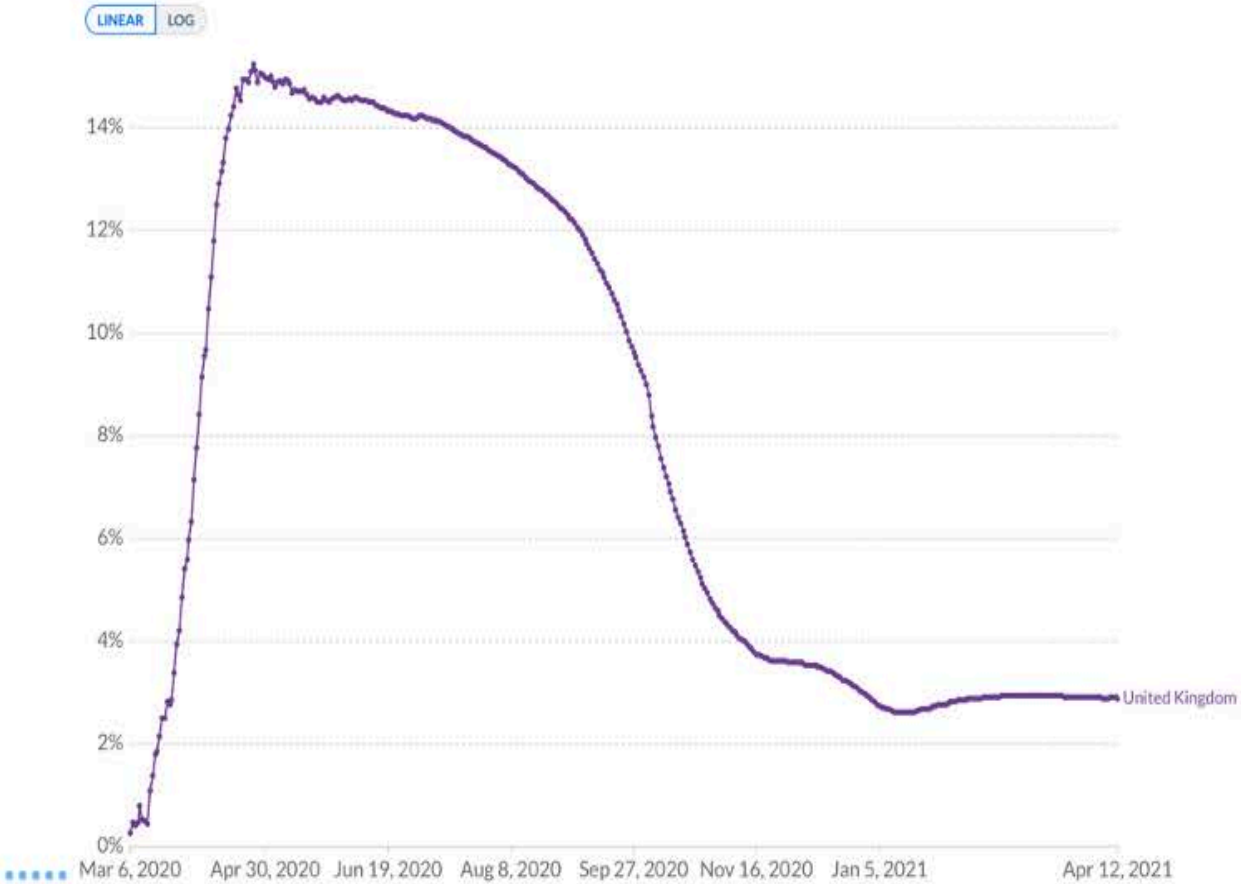
What is the right dependent variable?

- Deaths per day
- Moving Average of deaths per day (3 or 7 day?)
- New Deaths per Million of the Population
- New Deaths per day over Cases
- Excess Deaths
- Hospitalisations
- New Infections/Cases
- R number, Infections per 100,000 of population

Figure 3: Alternative Dependent Variable for the UK

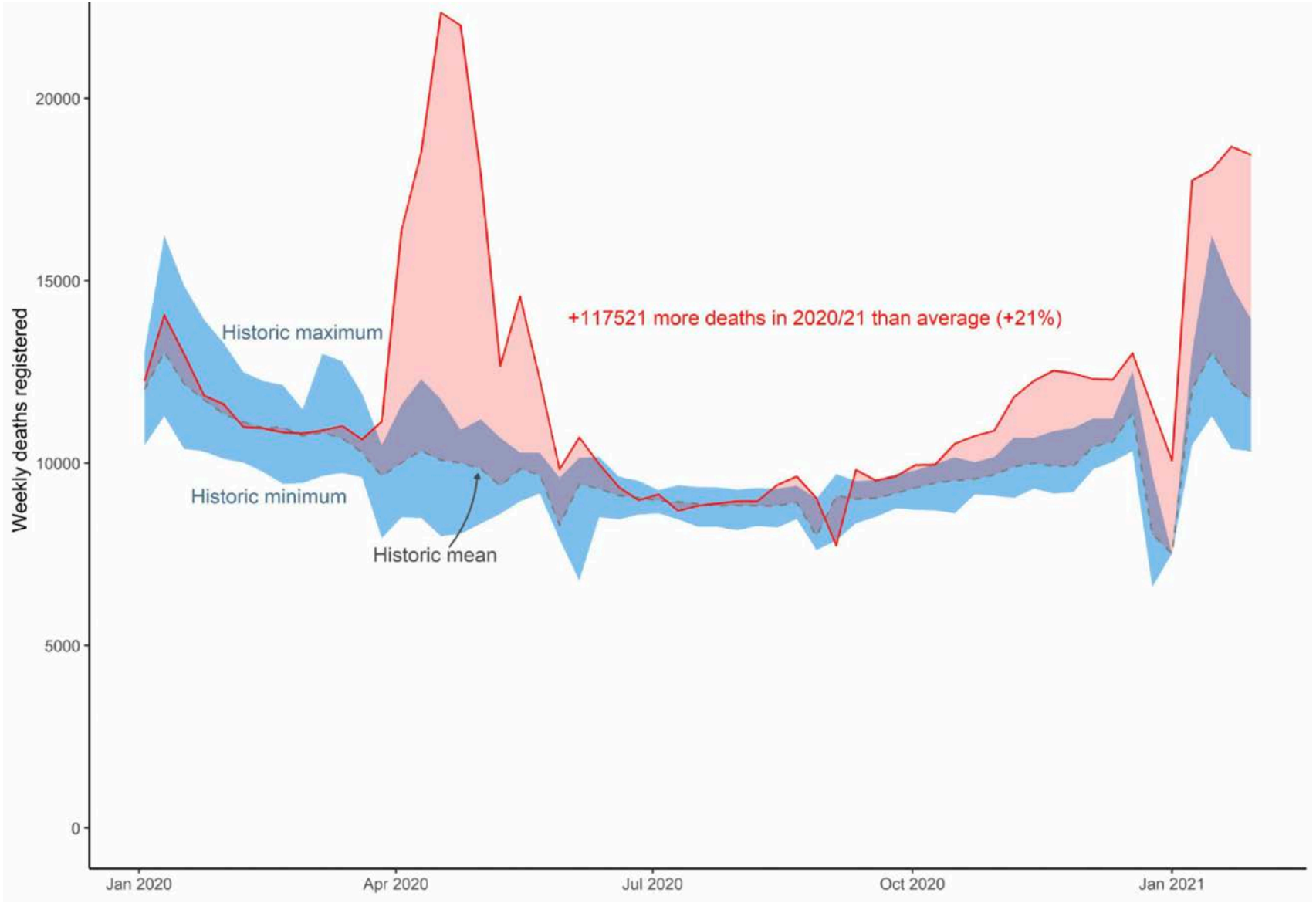
Panel A: The Reproduction Rate, $\mathcal{R}_t$ → → Panel B: Case Fatality Rate



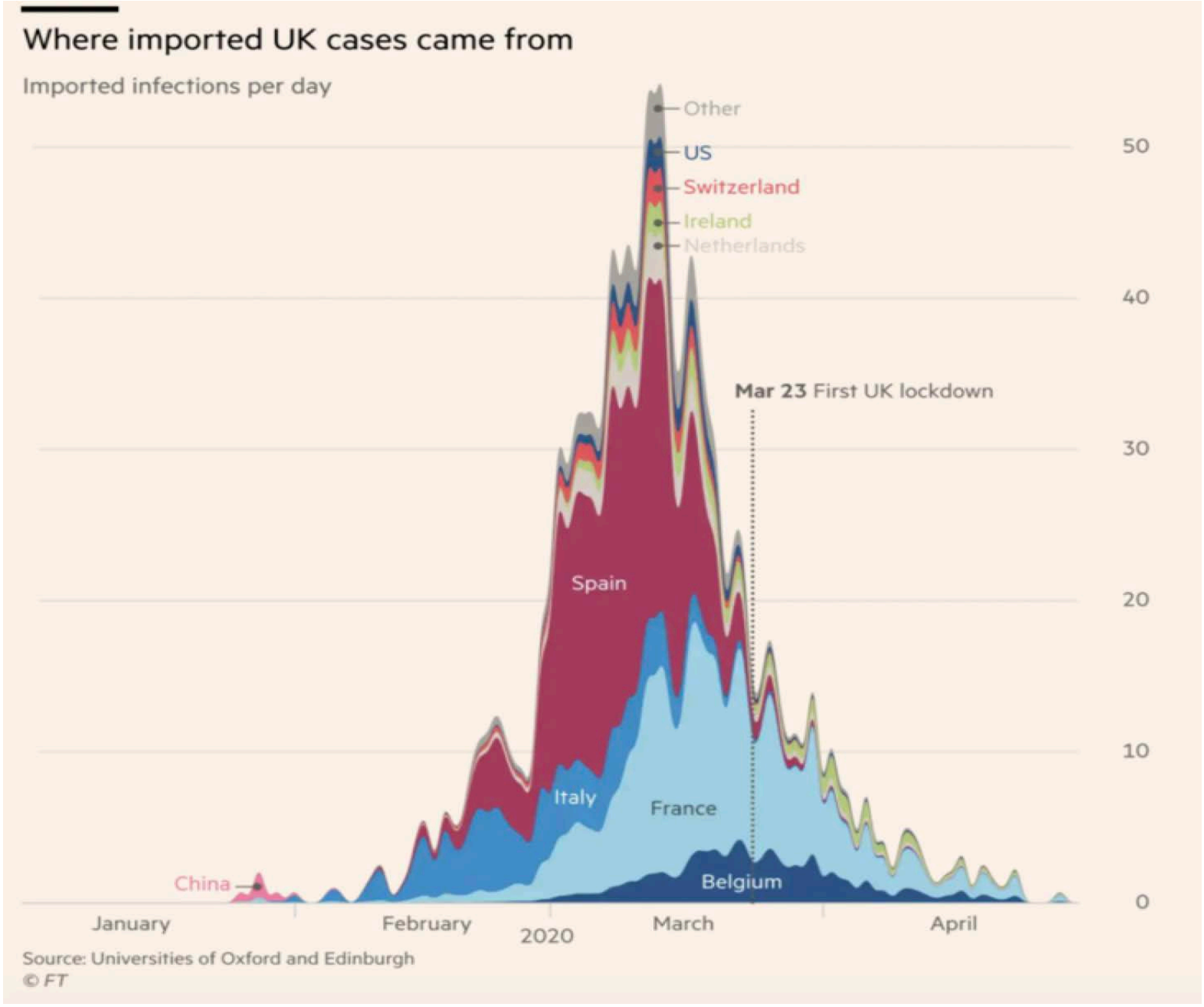


Excess Mortality England & Wales

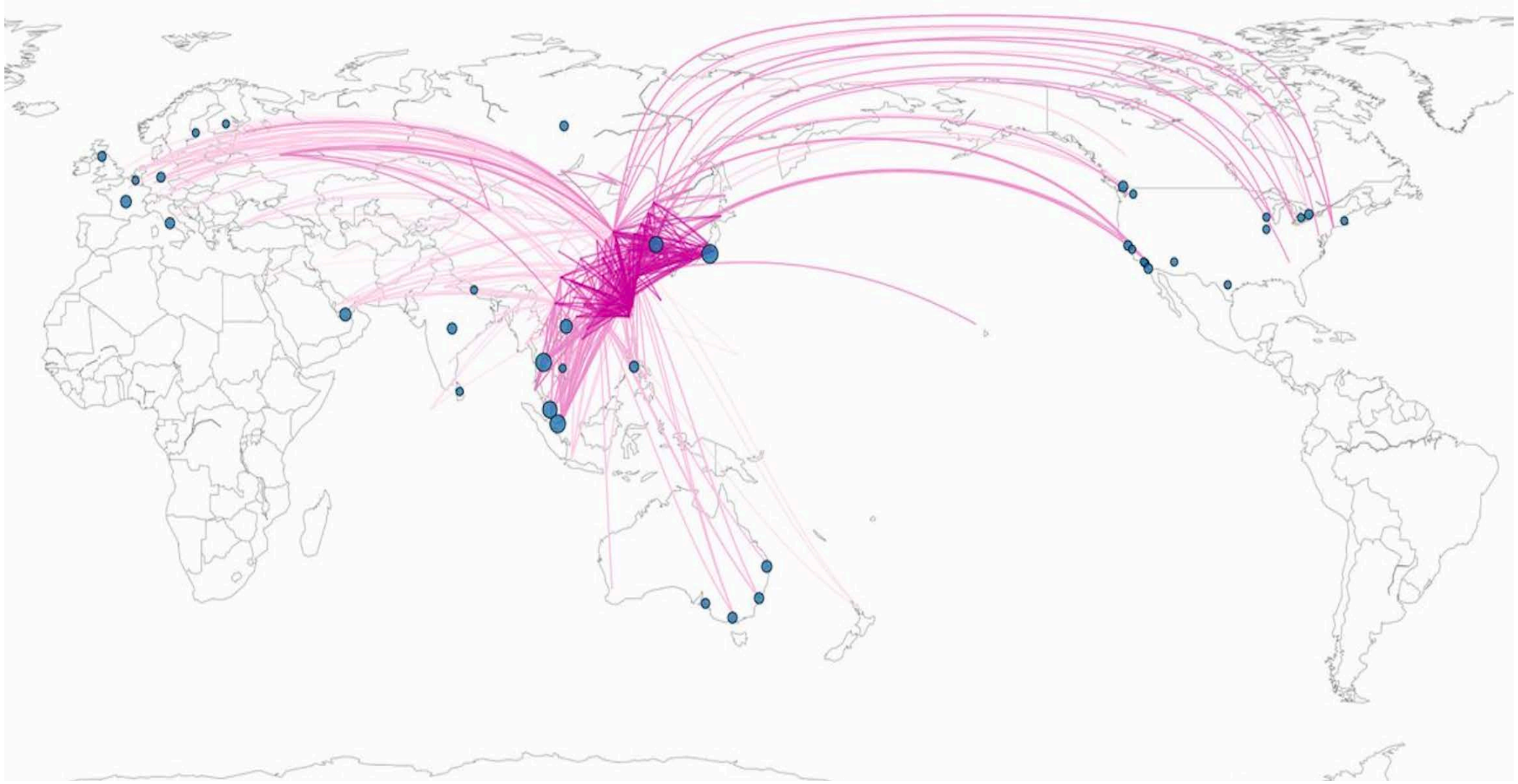
2020/21 compared to 2010-19.



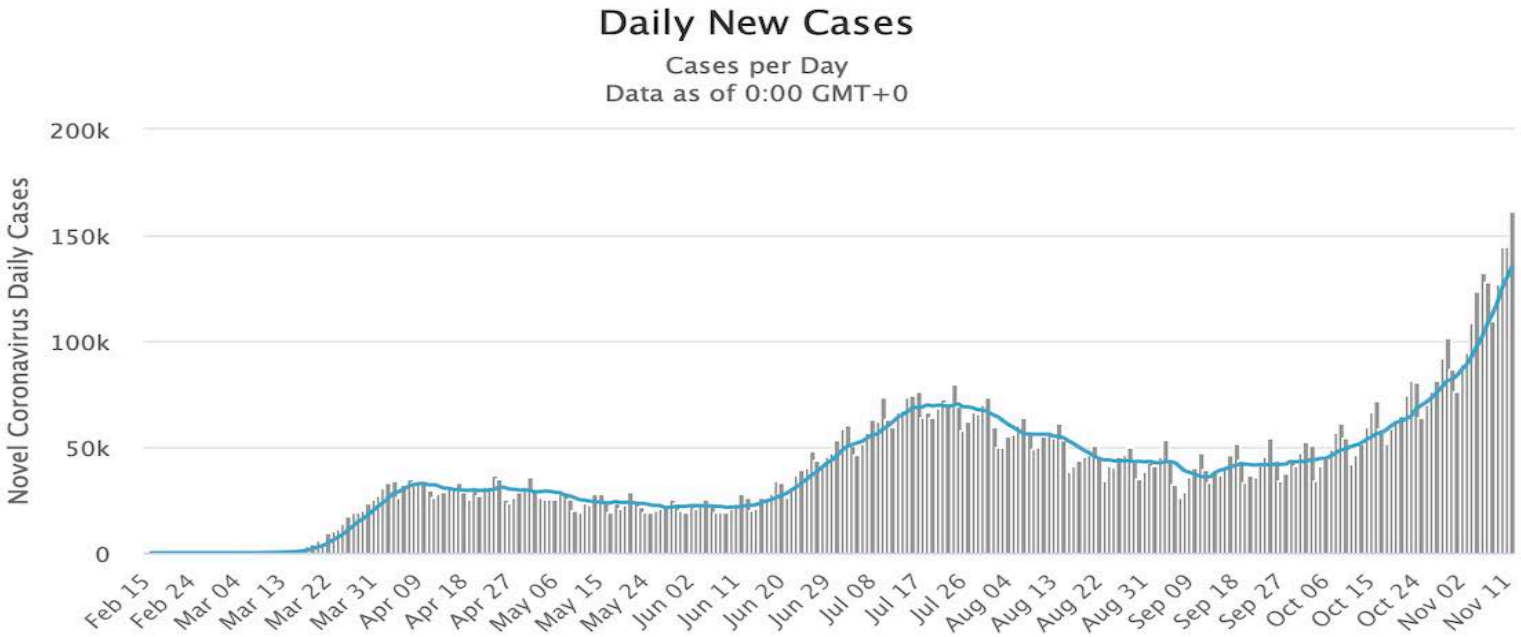
Spatial Issues: Where Did it Come From?



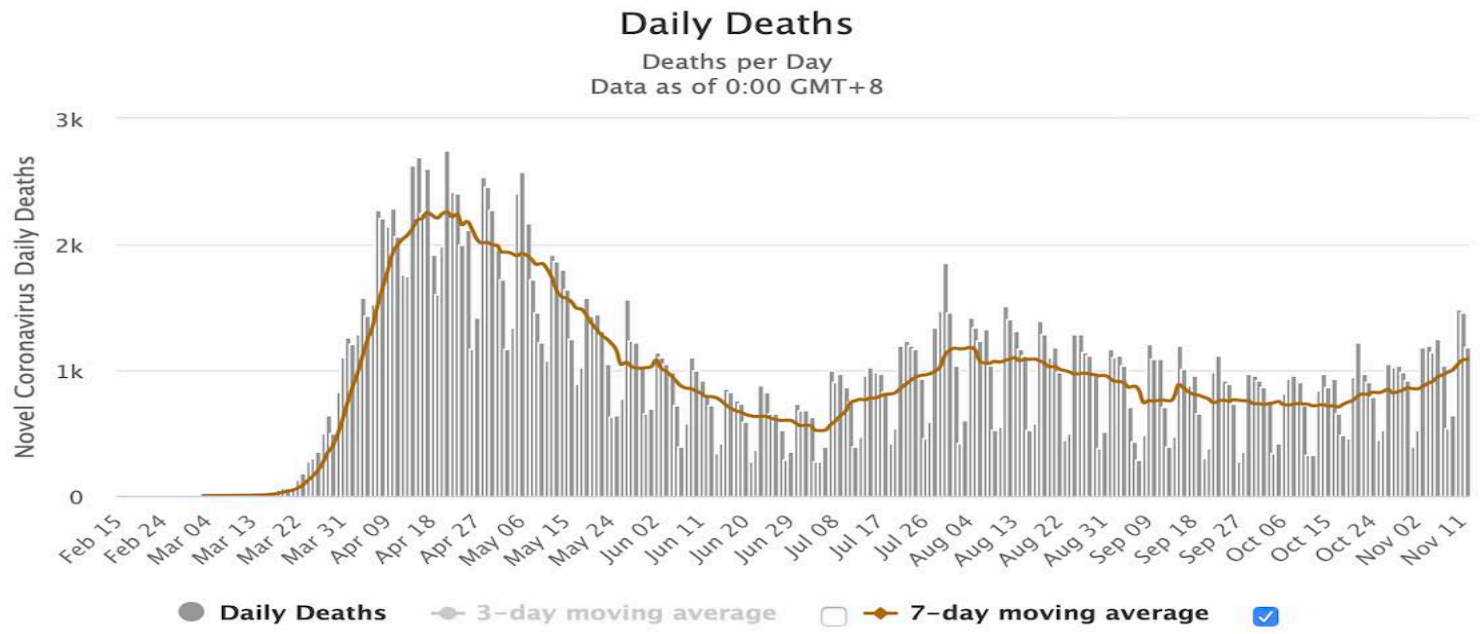
International flights from China in January 2020.



Observed Heterogeneity

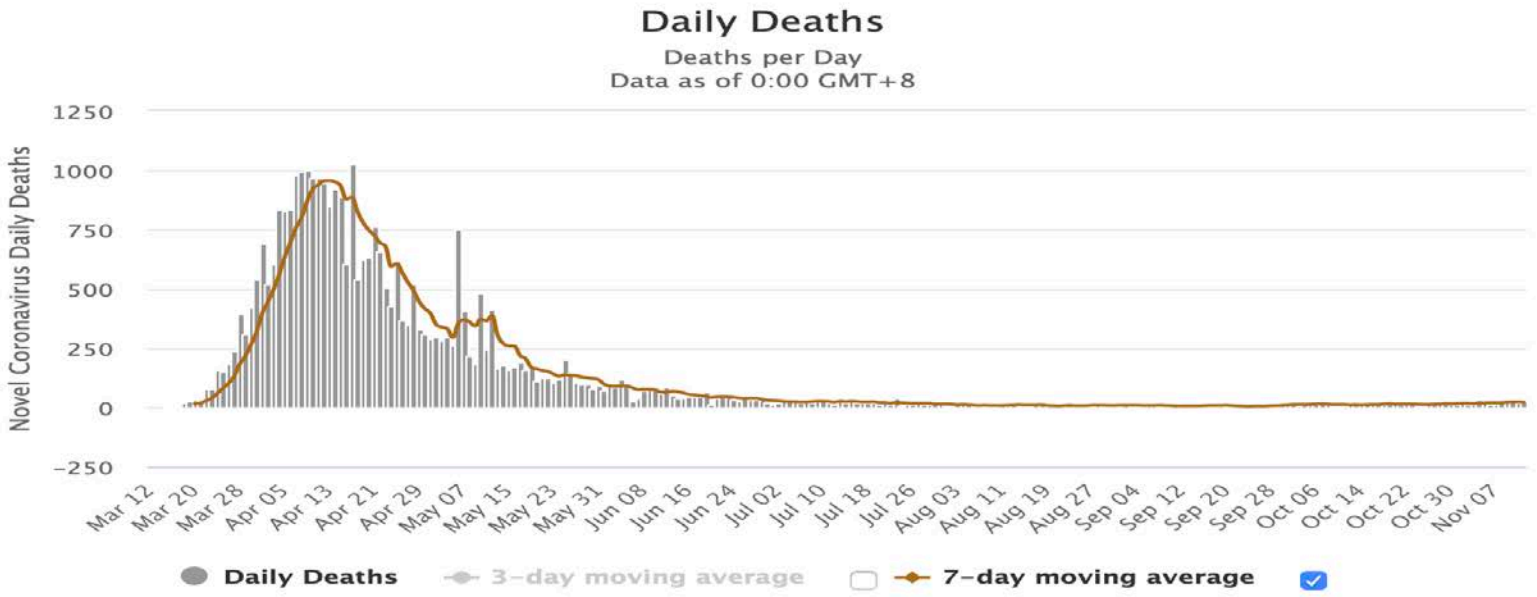


Daily New Deaths in the United States



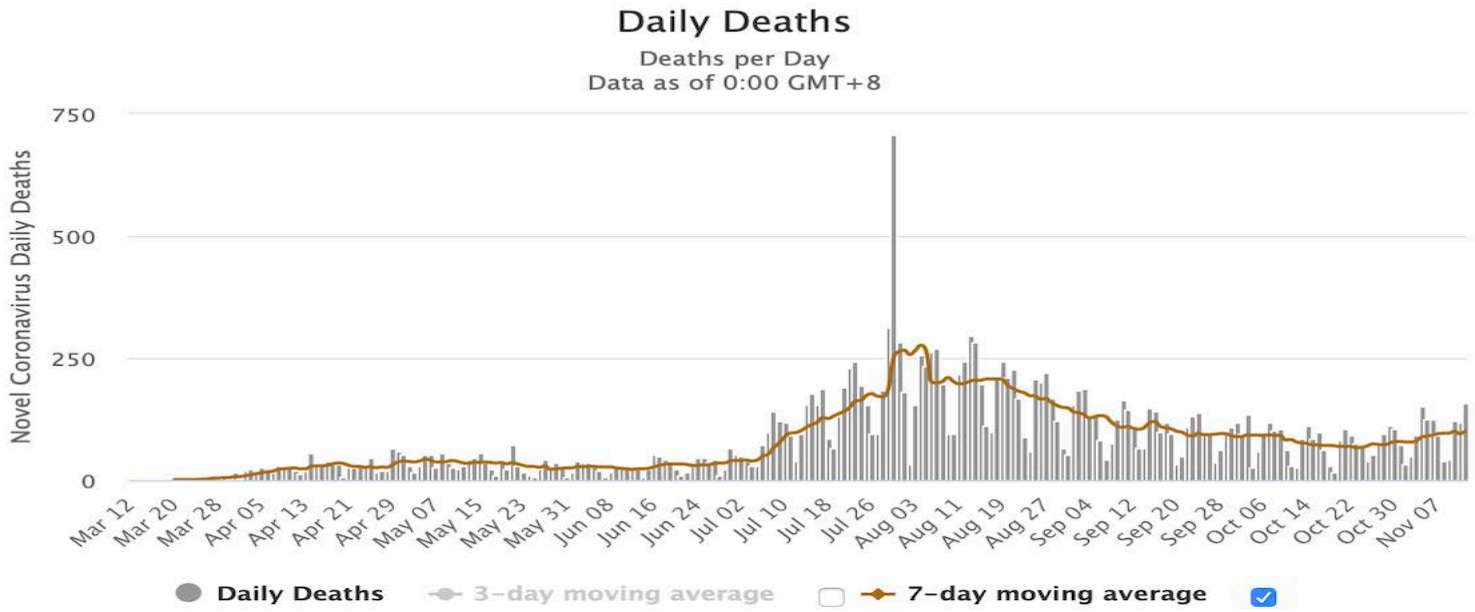
WHOLE USA

Daily New Deaths in New York



NEW YORK

Daily New Deaths in Texas



TEXAS

Endogeneity & How Do We Think About Policies?

CONTROLS

- School Closing
- Workplace Closing
- Cancel Public Events
- Restrictions of Gatherings
- Close Public Transport
- Stay at Home Requirements
- International Travel Controls

EXPENDITURE

- Income Support
- Debt/Contract Relief

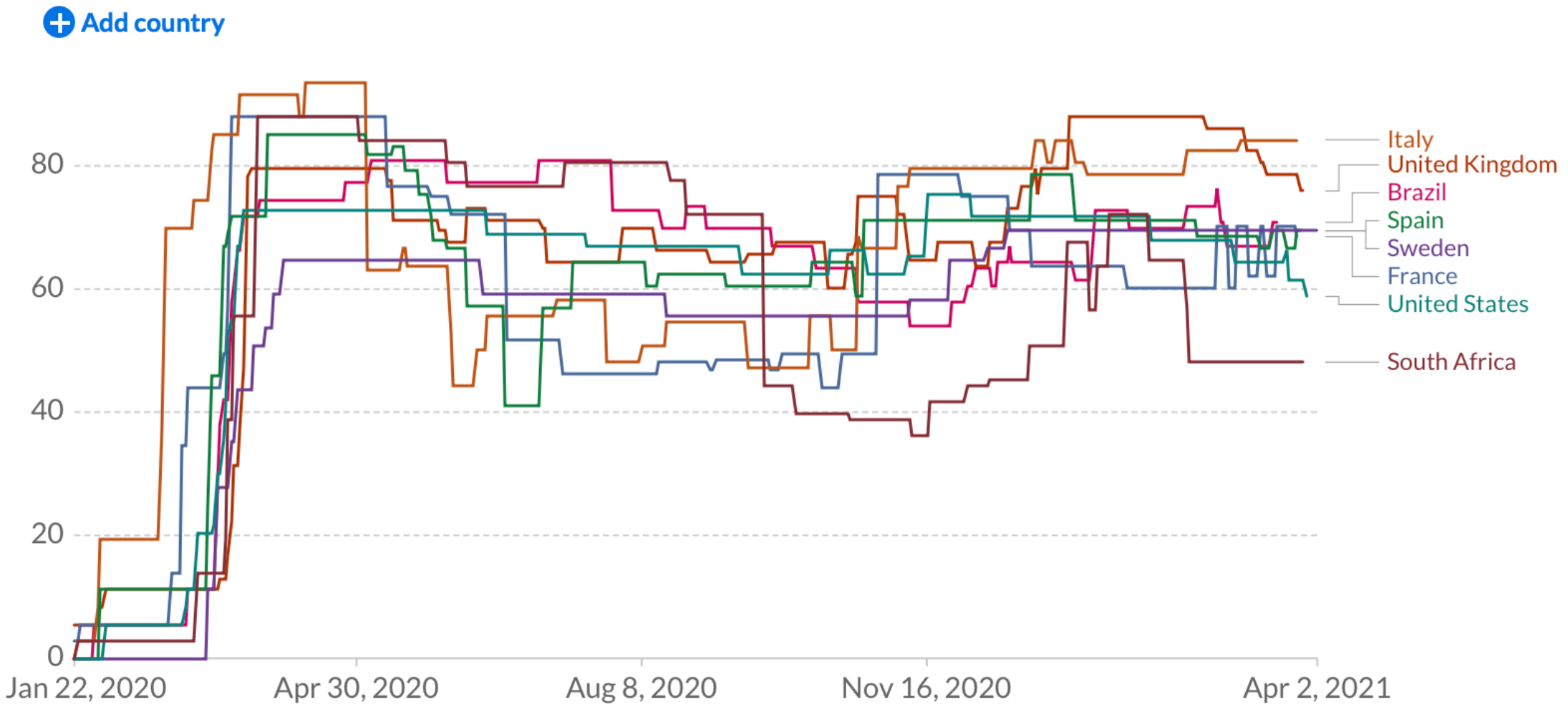
PUBLIC HEALTH

- Public Information Campaigns
- Testing Policy
- Contact tracing

COVID-19: Stringency Index

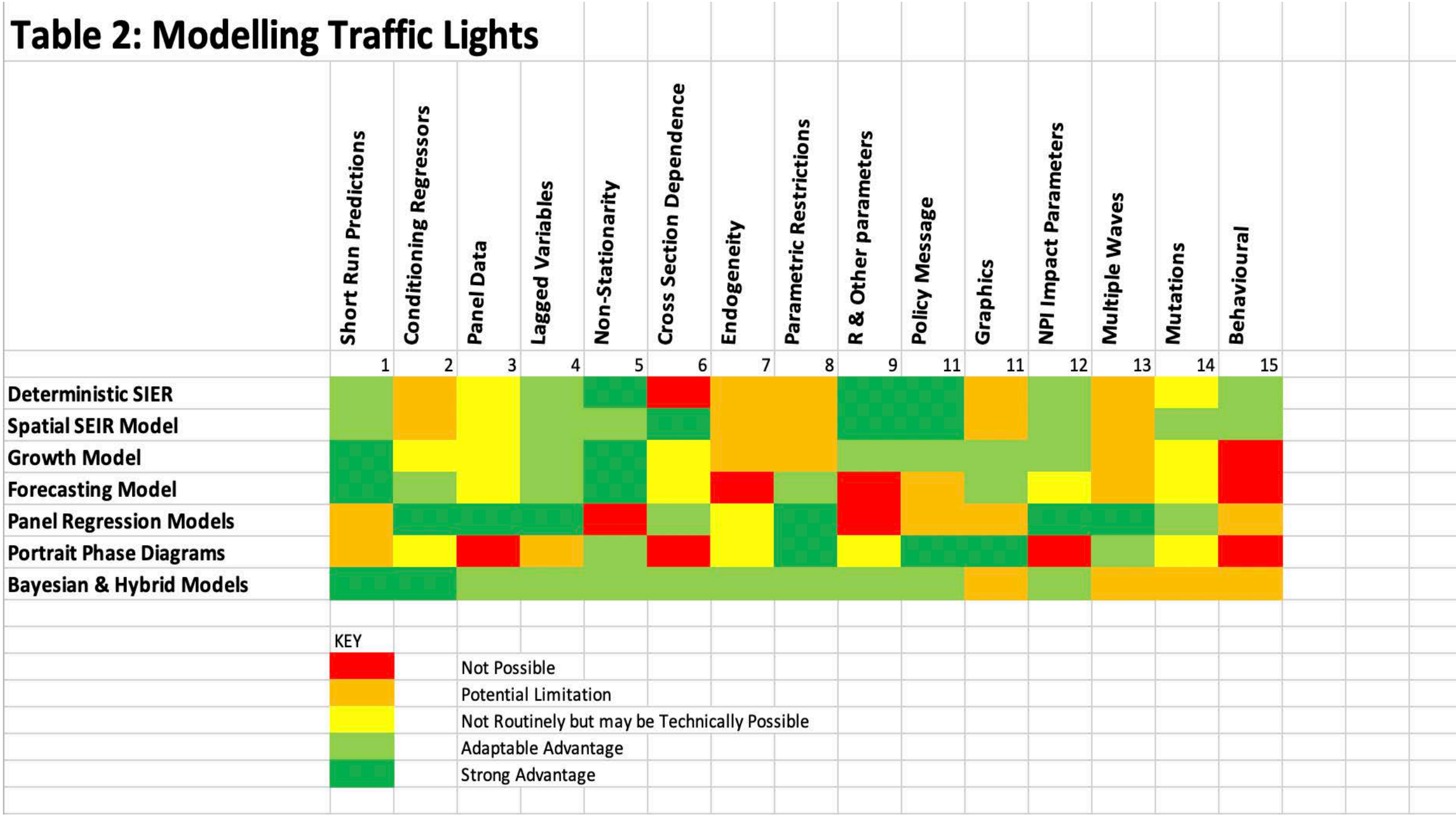
This is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest). If policies vary at the subnational level, the index is shown as the response level of the strictest sub-region.

Our World in Data



Source: Hale, Angrist, Goldszmidt, Kira, Petherick, Phillips, Webster, Cameron-Blake, Hallas, Majumdar, and Tatlow (2021). "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." Nature Human Behaviour. – Last updated 4 April, 00:00 (London time)
OurWorldInData.org/coronavirus • CC BY

Table 2: Modelling Traffic Lights



CONCLUSIONS

- Important Econometric Modelling Problems Remain:
  - Measurement Error
  - Endogeneity
  - Stationarity
  - Non-linearities & Multiple Waves
  - Cross Section Dependence & Networks & Spatial Aggregation & Modelling
  - Behaviourial Responses

- Econometric and other Statistical Models & Graphical Methods have a complementary role to play to SEIR models.
- Should be cross validation studies employing all different models on the same data.
- There should be econometricians, economists and other health professionals on SAGE.