Excess mortality statistics in the pandemic and country comparisons

Janine Aron and John Muellbauer, INET@Oxford

Seminar
Federal Reserve Bank Dallas
28 July 2020

Outline

- Why is it important to examine excess mortality data?
- How is excess mortality measured? Who measures it?
- Key issues for comparing rates of excess mortality across/within nations.
- Considering the comparability of statistical measures of excess mortality.
- Why the age distribution matters.
- What can we learn comparing P-scores from the European ‘all ages’ data?
- Excess mortality for other European age groups: ‘15-64’ and ‘85+’.
- Data and models
- ***Our own work in this area:***

https://ourworldindata.org/covid-excess-mortality

or pdf version on the INET, Oxford website:

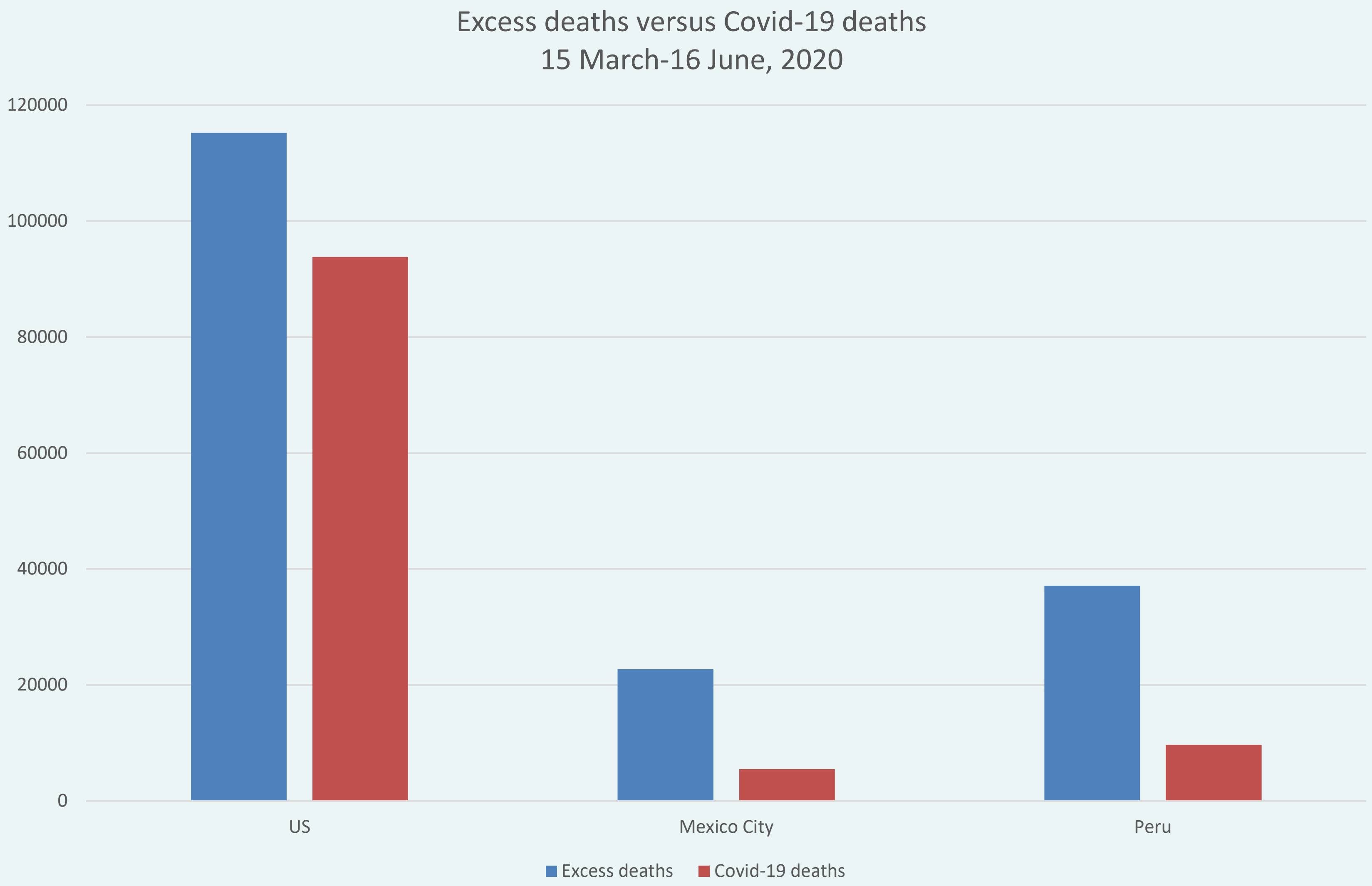
https://www.inet.ox.ac.uk/publications/a-pandemic-primer-on-excess- mortality-statistics-and-their-comparability-across-countries/

Why is it important to examine excess mortality data?

- Excess mortality: a count of deaths from all causes relative to what would *normally have been expected*.
    - In a pandemic, deaths rise sharply, but causes are often inaccurately recorded, particularly when reliable tests are not widely available.
    - The death count attributed to Covid-19 may thus be significantly undercounted.

- Excess mortality data *overcome 2 problems* in reporting Covid-19-related deaths.
    - Miscounting from misdiagnosis or under-reporting of Covid-19-related deaths is avoided.
    - Excess mortality data include ‘collateral damage’ from other health conditions left untreated, or by actions that prioritise patients with Covid-19 over other symptoms.

- In a pandemic, measures taken by governments and by individuals also influence death rates. Excess mortality captures *the net outcome* of all these factors.
    - E.g., deaths from traffic accidents may decline but suicide rates may rise.


Excess deaths versus Covid-19 deaths
15 March-16 June, 2020
120000
100000
80000
60000
40000
20000
0
US
Mexico City
Peru
Excess deaths
Covid-19 deaths

Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020

Journal: JAMA Intern Med.
Daniel M. Weinberger et al., July 1, 2020
https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2767980

- Estimates that from 1 March – 30 May 2020, the percentage of excess deaths reported as COVID-19 deaths was:

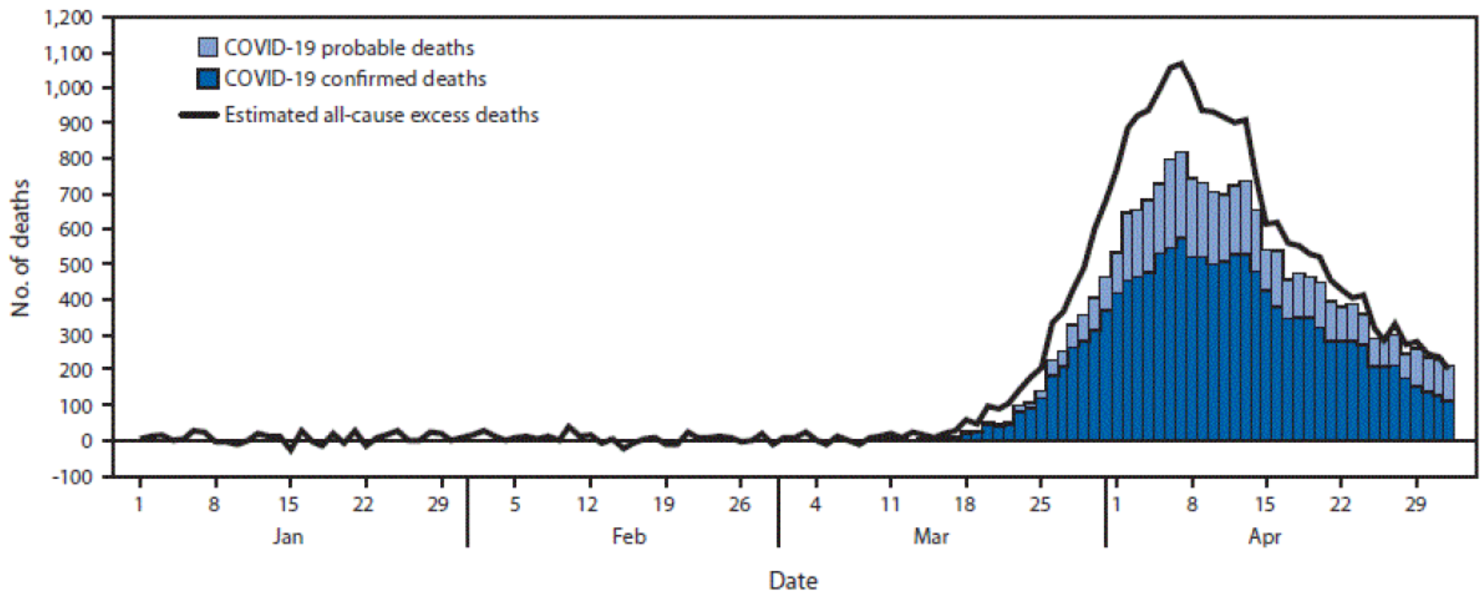
U.S.	78%	
Texas	45%	(*testing began far later*)
NY City	74%	
Mass	90%	

- Estimates depend on time interval and (somewhat) on method of estimating 'normal' (i.e. expected) deaths.

CDC: Preliminary Estimate of Excess Mortality During the COVID-19 Outbreak — New York City

UNIVERSITY OF OXFORD | OXFORD MARTIN SCHOOL | Institute for New Economic Thinking

FIGURE. Number of laboratory-confirmed* and probable† COVID-19–associated deaths and total estimated excess deaths§ — New York City, March 11–May 2, 2020



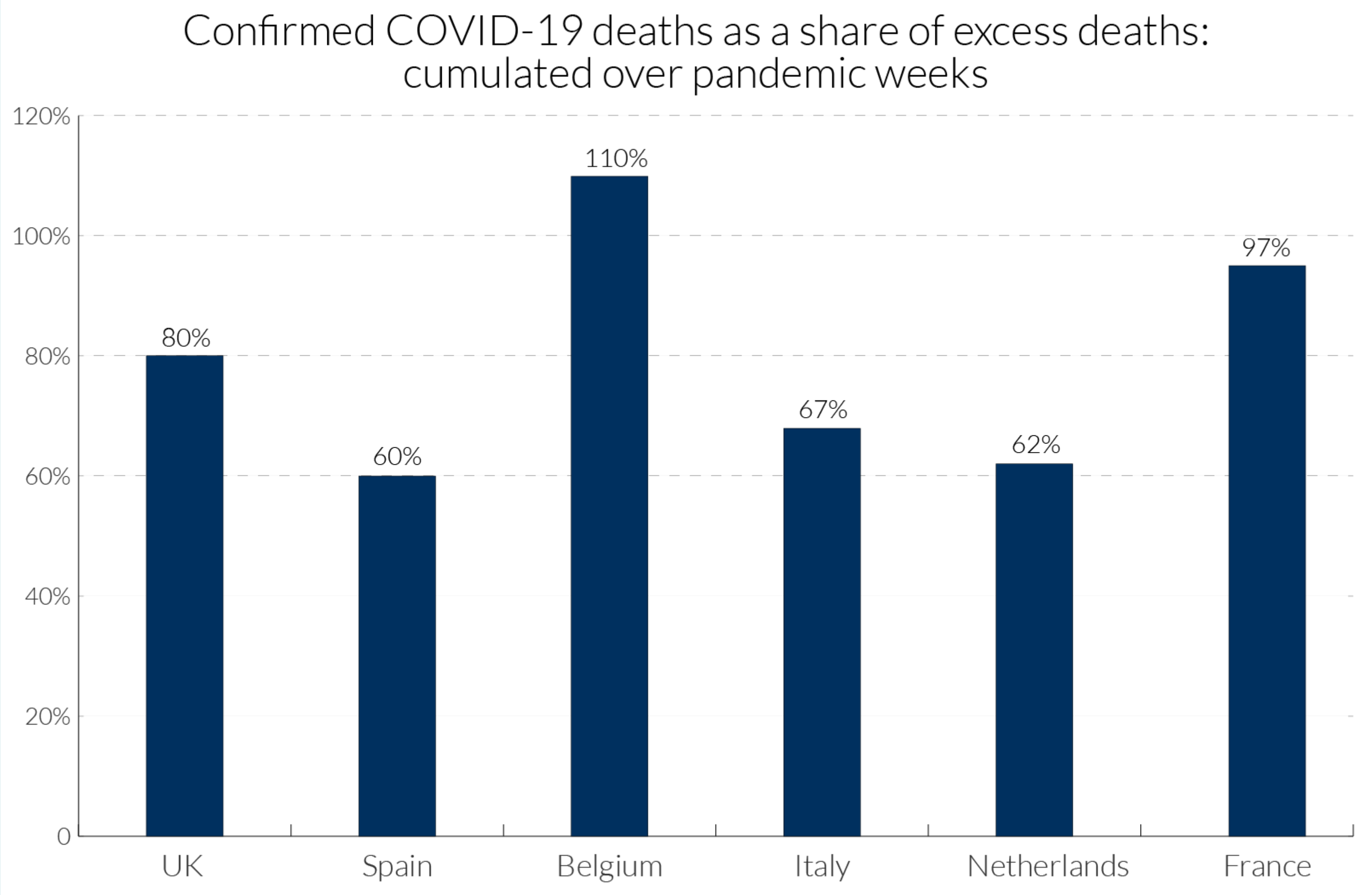
* Death in a person with a positive laboratory test for SARS-CoV-2 RNA.

† Death in a person without a positive test for SARS-CoV-2 RNA but for whom COVID-19, SARS-CoV-2, or a related term was listed as an immediate, underlying, or contributing cause of death on the death certificate.

§ Total excess all-cause deaths were calculated as observed deaths minus expected deaths as determined by a seasonal regression model using mortality data from the period January 1, 2015–May 2, 2020.

https://www.cdc.gov/mmwr/volumes/69/wr/mm6919e5.htm

Covid-19 deaths relative to excess deaths: Europe



Measures of excess mortality: Z-scores and P-scores and per capita excess mortality

Denote the number of weekly deaths by x.

- **The P-score** is defined:
  (x minus the *expected value* of x for the population),
  divided by the *expected value* of x for the population.

- **A variant P-score** (U.S. National Center of Health Statistics) is defined:
  (x minus the *upper threshold for the expected value* of x for the population),
  divided by the *upper threshold for the expected value* of x for the population.

  Upper threshold: the expected value plus the 2.5% confidence interval for this expected value (takes into account uncertainty created by the natural variability of x).

- **The Per Capita Excess Mortality** is defined:
  (x minus the *expected value* of x for the population),
  divided by the population.

- **The Z-score** used by EuroMOMO is defined:
  (x minus the *expected value* of x for the population),
  divided by the *standard deviation for the population* of x around its expected value, based on Poisson distribution (with over-dispersion).

  EuroMOMO *estimate* the expected values (i.e. ‘normal’ deaths) with data for previous 5 years, taking seasonal factors and trends into account, adjusting for delays in registration.

Data sources

- Raw data on *weekly deaths*:
  - Best global country source, 25 countries The Human Mortality Database
  - HMD has comprehensive, transparent metafile for data sources and coverage, by age and gender.
- *Europe*: Eurostat
  - transparent metafile for data sources and coverage by age, gender and region.
- *Excess mortality data*:
  - EuroMOMO (Z-scores), The Economist (P-scores), FT (P-scores), NY Times (per capita excess mortality) were pioneers.
  - More recently, the Health Foundation, BBC and Guardian newspaper.
- *US data by state:* variant P-scores from CDC
  - now provide raw expected deaths so that user can compute P-scores
  - like EuroMOMO, estimate of std. dev. based on Poisson distribution (with over-dispersion)
  - https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm

Motivations for comparisons of mortality

- Compare the death toll of the first wave of the pandemic.
  - Aggregate measures are useful such as the count of excess deaths relative to normal deaths, the P-score and excess deaths relative to population size.
  - The last overstates the incidence of the pandemic in older vs. younger populations.
- Evaluate the effectiveness of policy responses.
  - Countries may differ in the size of the initial source of infection, in their age structure, in the distribution of co-morbidities in the population and the prevalence of dense urban centres, making some countries more vulnerable.
  - Country differences need *interpretation*.
- Improve the scientific understanding of the dynamics of the spread of infections, their incidence and death rates.
  - Key is the production of granular data.
  - Disaggregation of excess deaths data by age, gender, region, and socio-economic categories.

- Differences in infection rates in preceding weeks
  - 2 countries/regions with the same average Covid-19 case fatality risk: 1% of all adults are infected in A, and 5% are infected in B. P-score will be about 5x as large in B in weeks after infection.
  - Early lockdown, effective test, trace and isolate procedures, lower average infection rates and hence excess deaths.
  - Fraction of adults *in locations or occupations where virus spreads easily*.
  - Social distancing, availability and use of face masks, clarity of health messaging and cultural differences in the exercise of self-discipline and following of advice.

- Differences in case fatality risks for infected adults
  - *Co-morbidities* (obesity, diabetes, pre-existing lung conditions).
  - Part of reason why *income and ethnic groups* differ.
  - Prevalence of *inequality, obesity and smoking* will influence comparative excess mortality.

- Non-linear effect of differences in Covid-19-specific health capacity
  - *Capacity constraints* on numbers of hospital beds and staff, numbers of ventilators, PPE, testing, and logistical failures in delivery, e.g. to care homes.
  - Similar initial capacities? Country with higher average infection rate faces greater constraints.
  - Same high infection rate? Country with lower health capacity faces higher excess mortality.
  - This is why there is such a focus on '*flattening the pandemic curve*'.
  - Capacity constraints can have *different implications for different groups* e.g. the oldest individuals.

Comparability of statistical measures of excess mortality

- It is *hard to compare* similar time-spans*:*
  - for relatively *homogeneous countries* with moderate population sizes (e.g. European countries, Japan) with *large countries* spanning diverse regions with different timings and incidence *(e.g.* China and the U.S.)

- Normal death rates *already* reflect persistent factors:
  - age composition of the population, incidence of smoking and air pollution, prevalence of obesity, poverty and inequality, and usual quality of health service delivery.

- **P-scores:** attractive for this reason - even if age compositions & other persistent factors differ.
  - Measure the *percentage* deviation compared to what is 'normal'

- **Per capita excess mortality:** better than attempting to compare crude excess death counts for countries with vastly different populations.
  - But, countries with older populations will tend to measure higher normal death rates.

- **Z-scores:** in principle, should *not* be compared across countries of different sizes, though captures profile of weekly excess deaths for an individual country.
  - EuroMOMO suppress information on raw excess and normal deaths and on the standard deviations they use (and have limited metafile). Since Poisson assumption is problematic, there is no simple adjustment, e.g. for size of population, to make Z-scores comparable for large and small countries.

UNIVERSITY OF OXFORD | OXFORD MARTIN SCHOOL | Institute for New Economic Thinking

- EuroMOMO & CDC use Poisson (extra dispersion) for non-pandemic mortality data
  - model the std. deviation, hence the upper threshold for variant P-score.

- Evidence for *common systematic factors* driving mortality data:
  - *Examine the ratios of Z-scores to P-scores*.
    - With same concept of normal or expected deaths, the Z/P ratio would equal the inverse ratio of the standard deviation to normal deaths.
    - Under the constant mean Poisson assumption, proportional to the square root of the number of normal deaths. But the evidence is against this.
  - *Examine the correlations of Z-scores within the UK*.
    - If there are systematic sources of variation of death rates, as well as pure noise, these systematic factors for the UK regions are very likely to be *correlated*.
    - 100 observations, 2015-19, excl. winter & summer weeks (as do EuroMOMO), correlation matrix reveals quite high correlations: implies systematic factors common to regions.

	England	Wales	Scotland	N. Ireland
England	1			
Wales	0.345482	1		
Scotland	0.326606	0.205122	1	
N. Ireland	0.298233	0.106424	0.17243138	1

- Poisson may be appropriate for sparse count data, but less relevant for 100s or 1000s of weekly deaths.

- The accuracy of the basic data collected
  - Differences in efficiency of death registration systems, particularly if devolved to regional/local administrations.
    - National records can be slow to absorb regional information. Capacity of systems can be overwhelmed, e.g. in hotspots, often urban.
    - Sometimes observers resort to *data on burials* - Indonesian capital, Jakarta, burial data reported by the Financial Times.
    - *Striking recent revision* in Spain (May 27th): raw deaths from early March revised up by ~12,000, (mainly Catalonia & Madrid), the bulk for age group 75+

- Lag between occurrences data versus registration data on death counts
  - Difference between *death counts by week of registration of the death* and *week of actual occurrence of the death*. The registration data occur later.
  - EuroMOMO Z-scores use data by occurrence for all reporting countries.
  - HMD use occurrence data for most countries (exception is England & Wales).
  - Occurrence-data *prone to revision*; *increasing lags* of registration data during a pandemic.

- Measurement of ‘normal deaths’
  - The 5-year average is crude estimate of normal deaths. If mortality is on an improving trend, or weather is unusually kind, normal deaths would be over-estimated by the 5-year average.
  - Where populations are increasing or ageing, normal deaths could be rising.
    - EuroMOMO use statistical models to adjust for such trends but do not provide their estimates of ‘normal’ deaths. CDC do publish estimates.

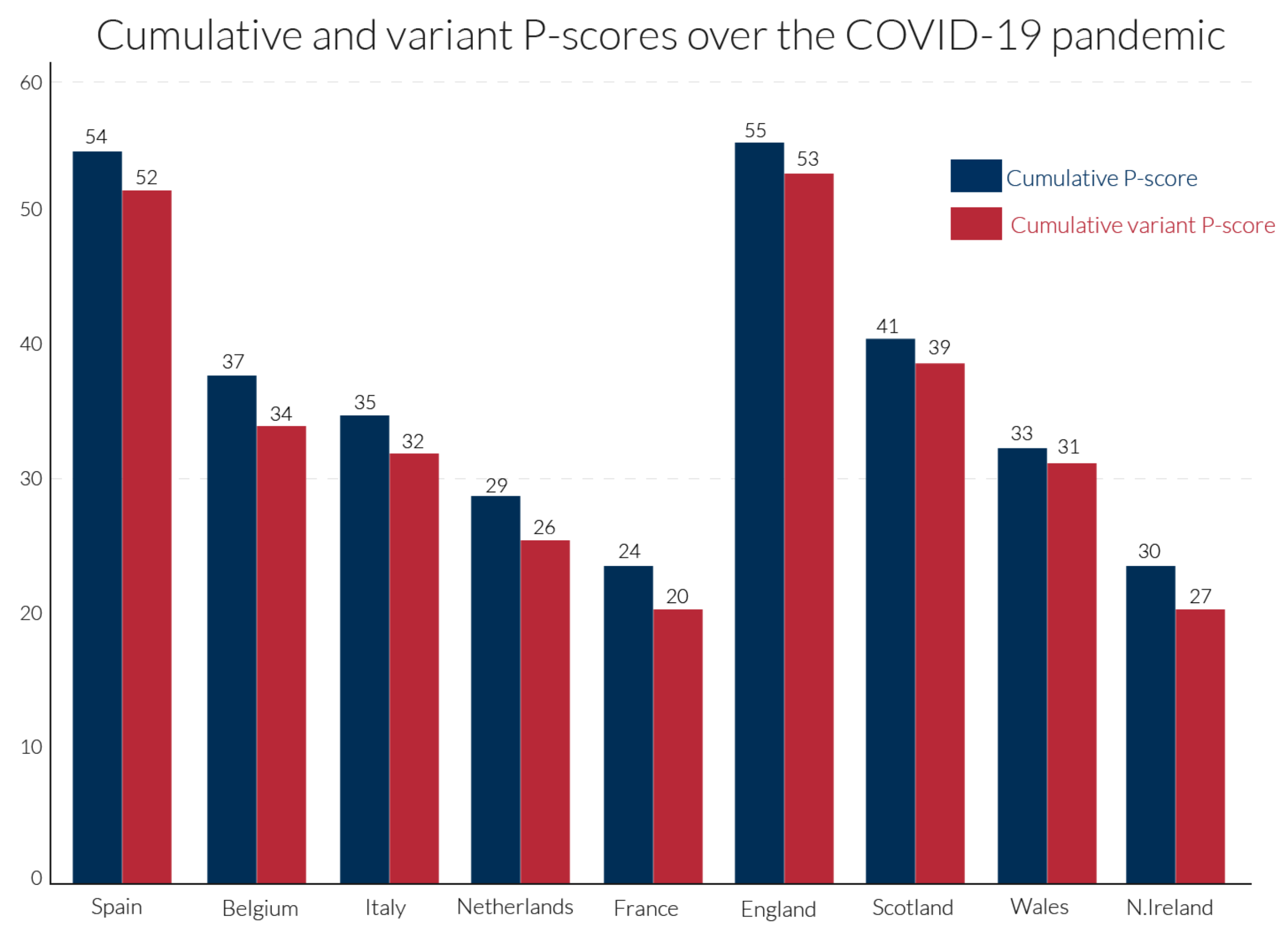
Why the age distribution matters

- There is strong empirical evidence *against* the hypothesis of a proportionate increase in mortality risk at all adult ages.
- Countries also differ in the age-profile of P-scores.
- Compare the ratio of the P-score for working age adults (age '15-64')to older adults, e.g. '65+' or '85+'.
- Ratio is less than 1 everywhere, but some countries have far lower ratios.
- *To see the implications*:
  - Simple example of 2 countries with same age-structure of young and old adults.
  - Suppose the P-score is 1 for the old in both countries, but that country A has a P-score of 0.1 for young adults while that for country B is 0.3.
  - The overall P-score for country B will clearly be higher than for country A.
  - However, if country B also has a higher fraction of young adults, that will attenuate the difference in the overall P-scores between the 2 countries.
  - Thus, *differences in age distributions between countries will affect the measured 'all-age' P-scores*.
  - This should be recognised when comparing P-scores.

- *One could envisage an 'age-standardized P-score',* adapting the 'age-standardized mortality rate', sometimes used to examine the impact of a pandemic.
- *The 'age-standardized mortality rate':*
  - weighted average of the age-specific mortality rates per 100 000 persons
  - weights are the proportions of persons in the corresponding age groups of a standard population.
    - The WHO explains the rationale*: "Two populations with the same age-specific mortality* rates for a particular cause of death will have different overall death rates if the age distributions of their populations are different. Age-standardized mortality rates adjust for differences in the age distribution of the population by applying the observed age-specific mortality rates for each population to a standard population."
- The 'age-standardized mortality rate' *unfortunately conflates* variations in 'normal' mortality risk with variations in risk of death during a pandemic.
  - If the age-standardised mortality rate in 2020 is higher in region A than in region B, this does not necessarily indicate that the Covid-19 mortality risk is higher in A.
  - *It may be that normal mortality risk is higher* in region A than in B.
- Age-standardisation removes that part of the difference due to differing *age structures* of the two populations; but it does not remove from 'normal' mortality risk the socio-economic differences, and differences in the incidence of obesity or smoking and in health provision.

- Be aware of the limitation of any single measure of comparability between countries.
  - Subsumed within the aggregates are implicit value judgements.
  - Focusing on simple counts of excess deaths in pandemic, has an implicit assumption that the toll of an older life lost is the same as that of a younger life.
  - However, when a younger life is lost, many more years of life expectancy are lost, and one might want to attach a larger weight to deaths of the young.
- An *important argument of the lockdown sceptics* is an extreme version of this last point:
  - "the virus is mainly killing off those that were on their way out anyway", see Kelly (2020), Financial Times.
- On the 11th June, cancer specialist Prof. Karol Sikora stated in the UK's Telegraph that at least half of those dying of Covid-19 would have died anyway by the end of the Summer of 2020.
  - Sikora was wrong: detailed insurance industry data suggests an obese male smoker aged 80, even with heart or pulmonary disorders, would still have a life expectancy of at least 5 years.
- This suggests that the pandemic had a huge impact not just on the death count but on life-years lost, properly measured.

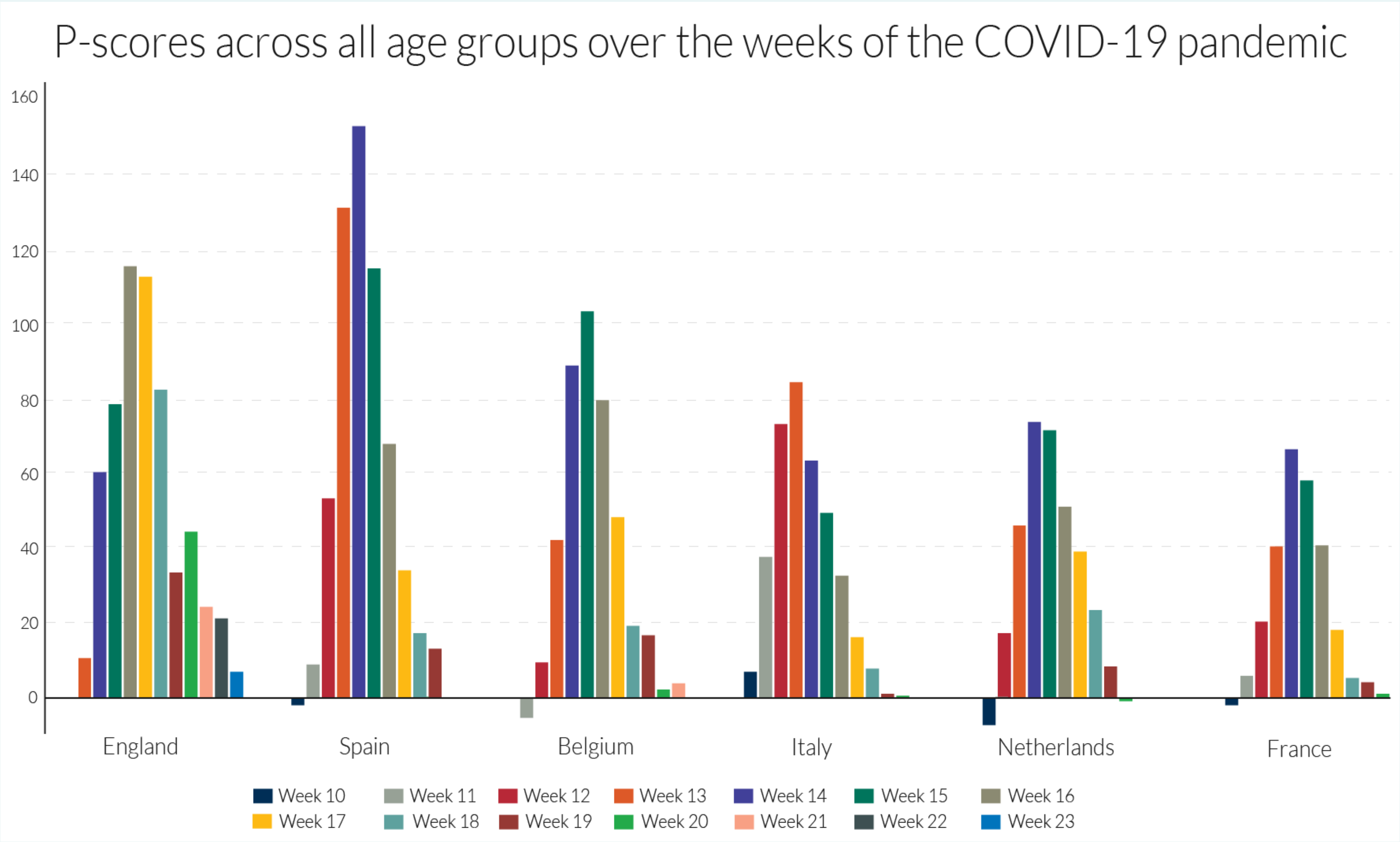
What can we learn comparing P-scores in Europe from the 'all ages' data?



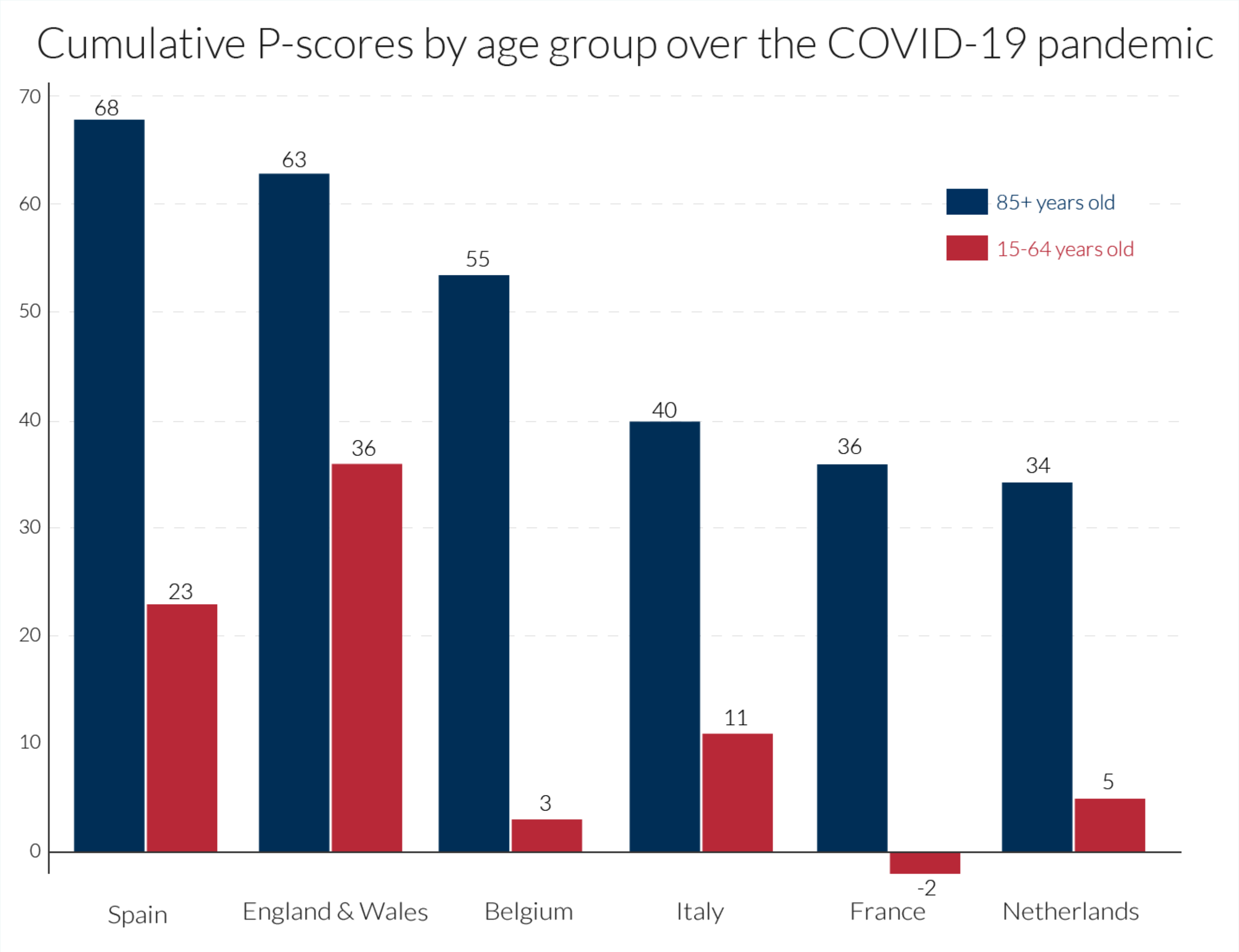
Sources: The P-scores and variant P-scores are calculated by the authors using the Human Mortality Database, see meta file: https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf, and the Office for National Statistics for the UK. Cumulative P-scores cover 11 main pandemic weeks, dates differ by country.

Recent weeks of P-scores for poor performers showing peak weeks of excess mortality for 'all ages'

UNIVERSITY OF OXFORD | OXFORD MARTIN SCHOOL | Institute for New Economic Thinking



Excess mortality for other age groups: '15-64' and '85+'.



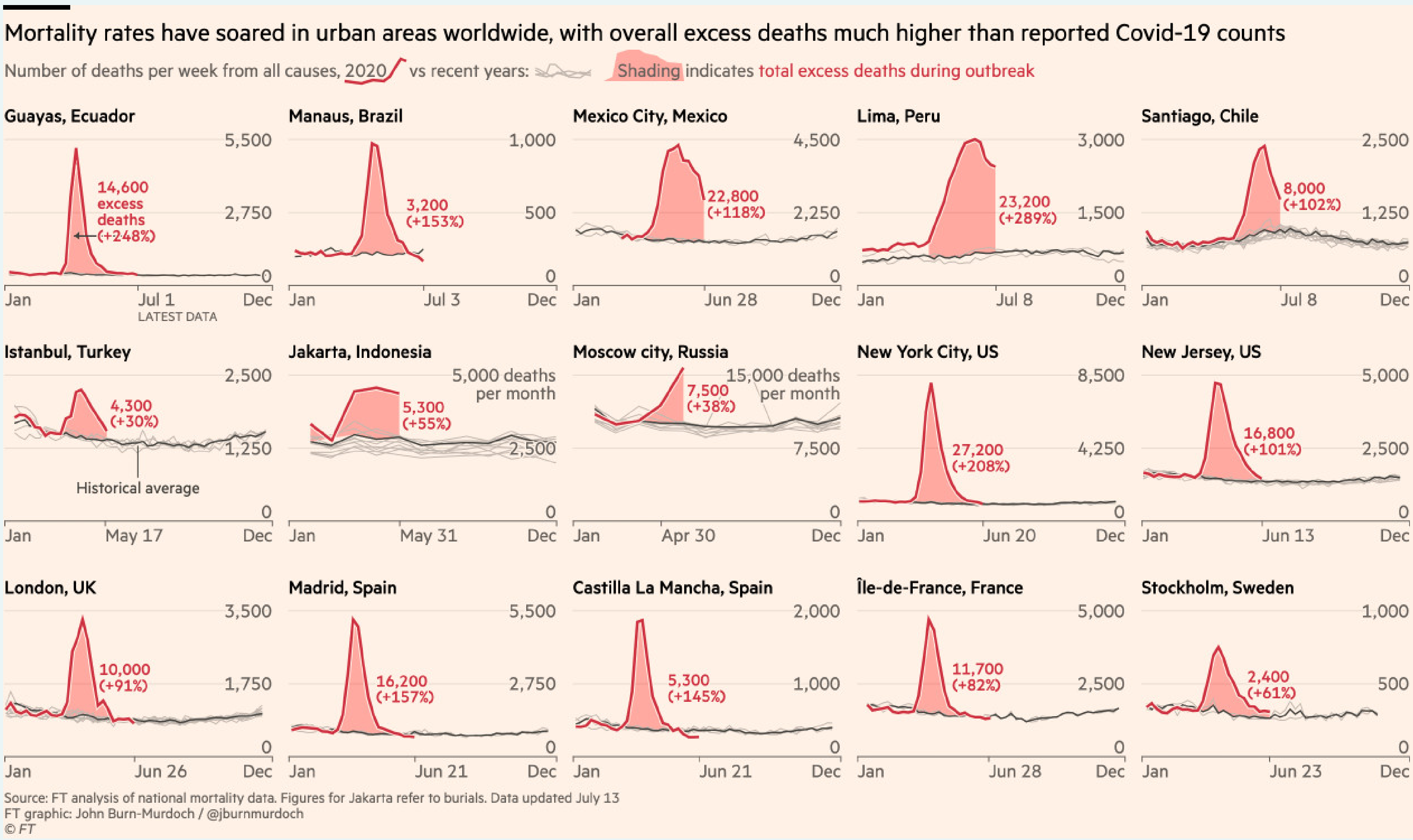
Sources: The P-scores are calculated by the authors using the Human Mortality Database, see meta file: https://www.mortality.org/Public/STMF_DOC/STMFmetadata.pdf, and the Office for National Statistics for the UK.

- Average infection rates in preceding weeks.
- Average case fatality risk from Covid-19: not just function of age and co-morbidities, but also of exposure to viral load –depends on social context.
- Non-linear effect of constraints on Covid-19-specific health capacity.
- Factors – intrinsic and policy implementation – that may cause differences in excess mortality, including:
    - **'Big city factor'** – its connectedness, the timing of its infection outbreak, its scale and population density, the nature of its public transport system.
    - **Social distancing and lockdown** – late and initially unclear application of social distancing/delayed lock-down measures in response to initial London outbreak/s.
    - **Public health system preparedness** – collective failure in preparedness for testing capability and on supplies and distribution logistics of PPE for health workers. Differences in bed, ICU counts per capita and health insurance.

- **Care homes** – late recognition of need to provide care-homes with PPE and tests.
- **Occupational groups** – differing Covid-19 death rates in different occupational groups – relatively high amongst security guards, transport workers, care workers. https://www.nytimes.com/interactive/2020/03/15/business/economy/coronavirus-worker-risk.html
- **Areas of deprivation** – places with greatest economic deprivation have much higher Covid-19-related death rates; underlying health is worse; low-paid key workers more exposed to infection may reside there disproportionately; over-crowded housing.
- **Ethnic differences** – in the incidence of Covid-19-related deaths.
  - Why was Covid-19 death rate among black and ethnic minorities in Scotland 'lower' than wider population, but higher elsewhere?

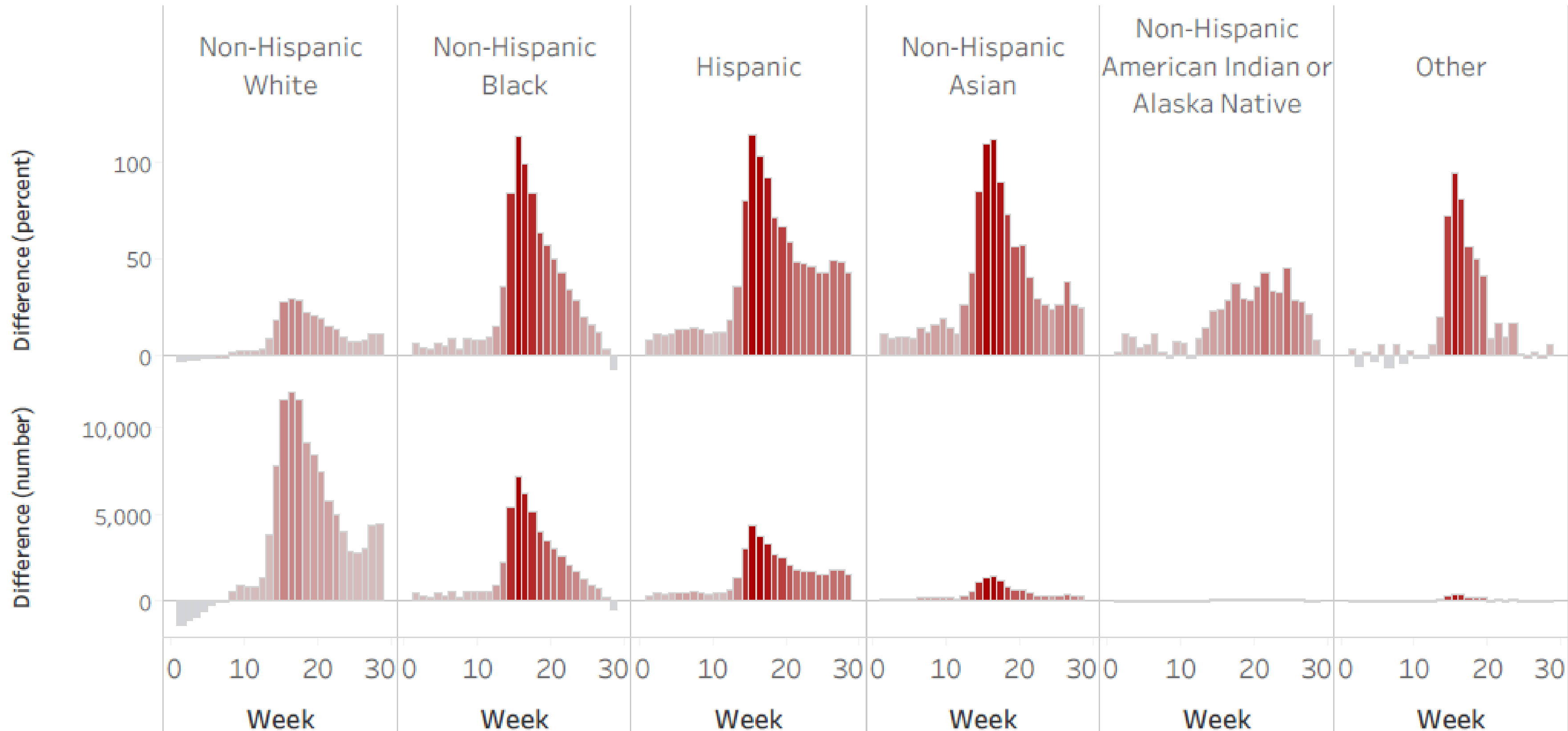
- England more adversely affected by this range of factors than rest of UK.
  - West Midlands & North West, the next largest conurbations in England after London area, had the next highest excess death rates.
  - Underlines the roles of timing and urban density.

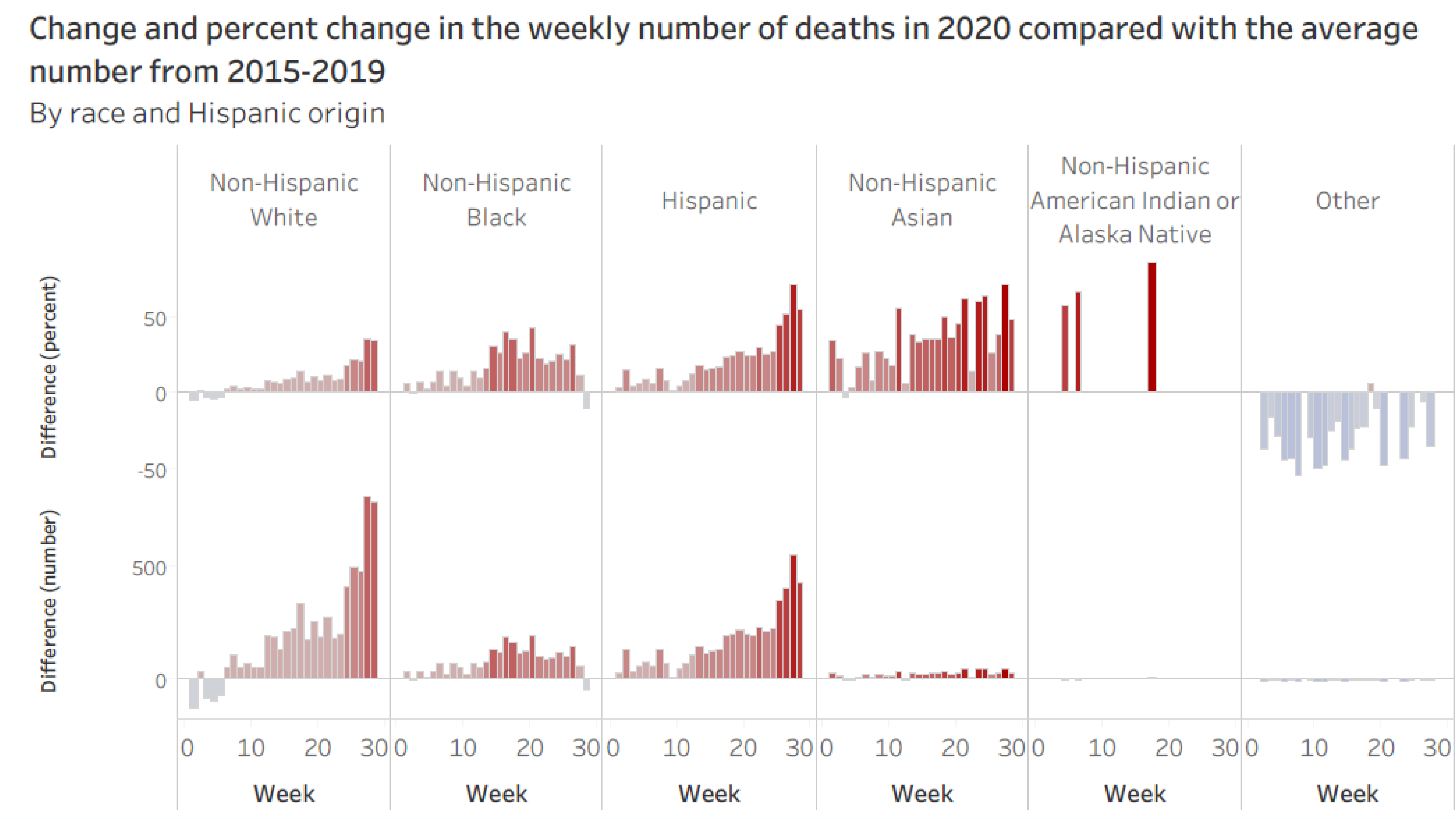
UNIVERSITY OF OXFORD | OXFORD MARTIN SCHOOL | Institute for New Economic Thinking



Change and percent change in the weekly number of deaths in 2020 compared with the average number from 2015-2019

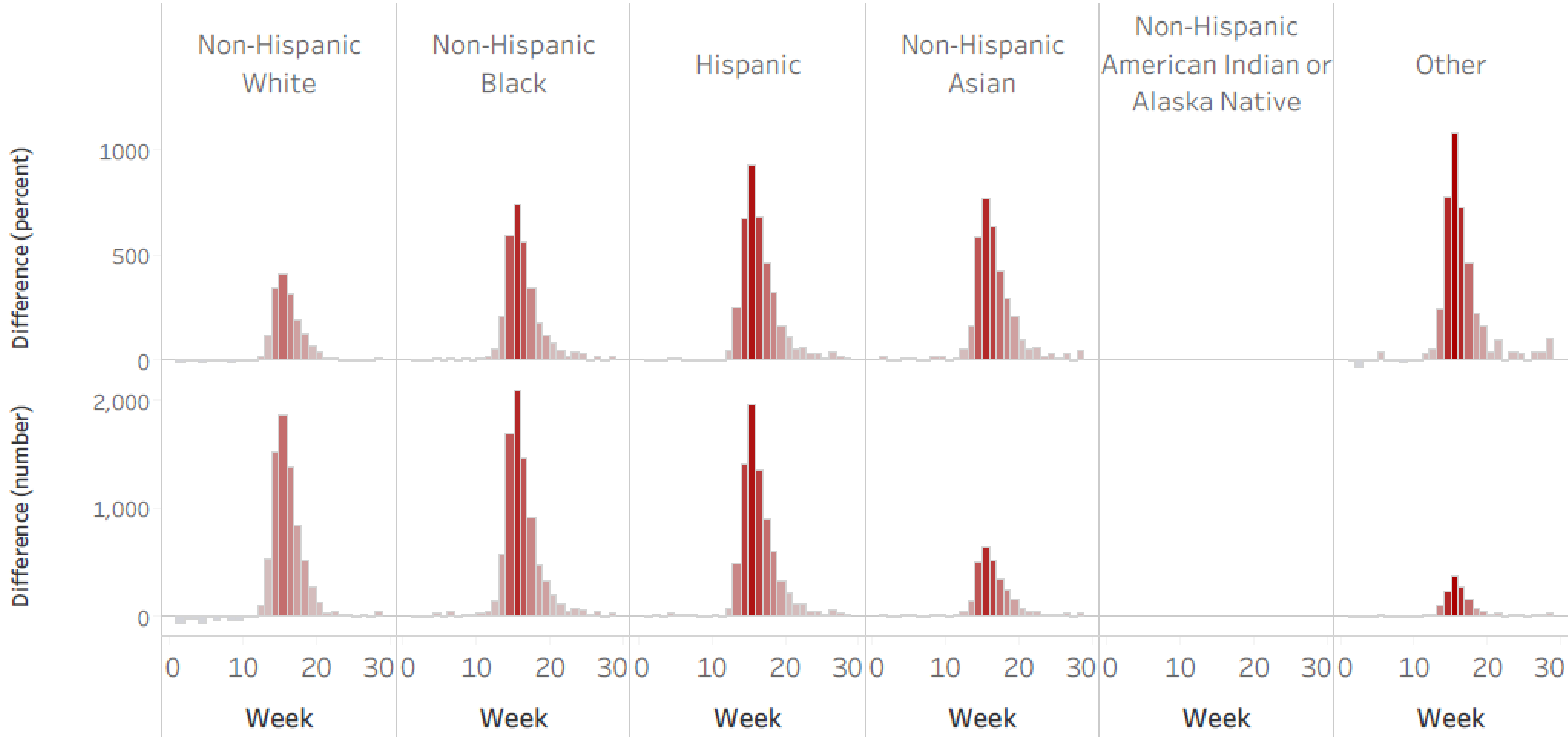
By race and Hispanic origin




Change and percent change in the weekly number of deaths in 2020 compared with the average number from 2015-2019
By race and Hispanic origin
Non-Hispanic White
Non-Hispanic Black
Hispanic
Non-Hispanic Asian
Non-Hispanic American Indian or Alaska Native
Other
Difference (percent)
Difference (number)
50
0
-50
500
0
0 10 20 30
Week

Change and percent change in the weekly number of deaths in 2020 compared with the average number from 2015-2019

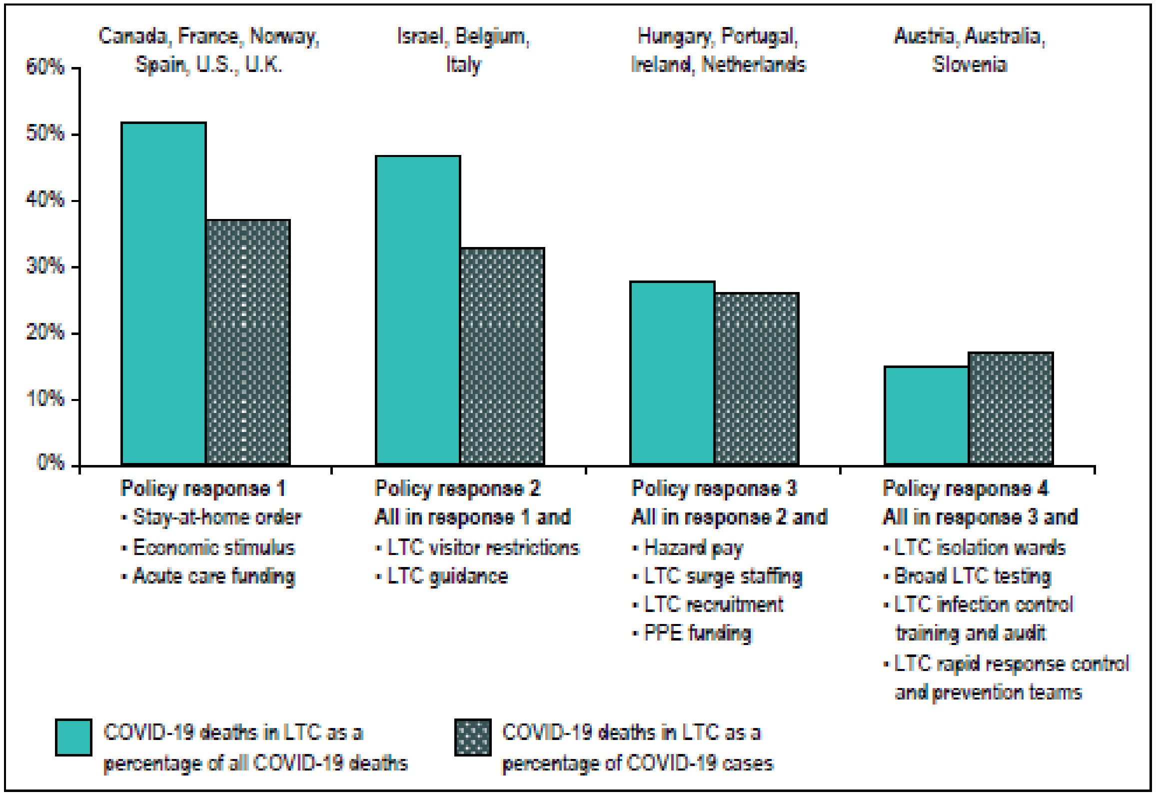
By race and Hispanic origin



Excess deaths in the '45-64' age group are far higher relative to '65+' (as in England) than in most of mainland Europe.

Deaths of those in long-term care (LTC)

Figure 3 Impact of COVID-19 on LTC residents by level of policy response at the time of 1,000 COVID-19 cases



Notes

LTC: Long-term care.

PPE: Personal protective equipment.

Countries are grouped according to which policy interventions were announced as mandatory at the time of a country's first reported 1,000 COVID-19 cases.

https://www.cihi.ca/sites/default/files/document/covid-19-rapid-response-long-term-care-snapshot-en.pdf

- International comparisons of % of cumulative Covid-19 deaths that occurred **in** care homes can be misleading.
- Ideally, should compare % of cumulative excess deaths among care home residents (even if they die away from the care home).

- **England & Wales** - comprehensive data on mortality **in** care homes *Comas-Herrera et al. (2020), Comas-Herrera & Fernandez (2020)).*
  - England & Wales, ~45% of total excess deaths took place *in* care homes (early March -12 June 2020).
  - 84% of total deaths from all causes of care home **residents** took place ***in*** care homes in the same period.
  - Scaling up the figure of 45% of total excess deaths that took place ***in*** care homes for England & Wales, by the 84% figure, i.e. 45.2/0.84=54%.
  - But 84% includes normal deaths that would have occurred in the absence of the pandemic, as well as the deaths induced by the pandemic. Hence 54% may under-estimate % of excess deaths among care home residents.

- Germany:
  - Developed the first coronavirus test in mid-Jan.-20 and ensured that capacity to manufacture large quantities of the test kits was rapidly rolled out.
  - Early large-scale testing and contact tracing.
  - Built up stocks of personal protective equipment (PPE) for health workers and care home staff earlier than other countries.
  - Trust in the government's social distancing guidelines and earlier lock-down.
  - More than 3 times as many hospital beds and intensive care beds than the UK, and many more ventilators.
  - Better protocol to protect care homes. UK sent infected residents from hospital back into care homes.
    - https://www.politicshome.com/news/article/government-threw-care-homes-to-the-wolves-amid-coronavirus-pandemic-mps-say
  - Many cities introduced 'coronavirus taxis', enabling medical staff to visit patients at home and take a test. If positive, the doc would then be in touch every other day.
  - Patients experiencing breathing problems, sent straight to hospital where beds and, if need be, ventilators, were available. Typical UK advice: stay at home 'to protect NHS'.
  - Patients admitted into hospitals if unable to shake off symptoms within a week.

- The *interaction of multiple factors* helps explain Germany's better record.
  - If health workers are effectively protected, fewer become sick, affecting the delivery capability of the health system, and many fewer patients are infected.
  - People are then less fearful of the risk of catching the virus from medical staff when seeking medical attention for other conditions.

Conclusions: need granular data in multiple classifications

- Many countries lack good death records.
- Big recent improvements at Eurostat and CDC:
  - Production of historical (last 5 years) as well as current deaths counts by age group, gender, region.
- Regional data:
  - NUTS1-3 is a common regional classification system in Europe but not used outside Europe.
  - US States. Census Bureau: the Northeast, Midwest, South (SE & SW) and West are broadly comparable to larger European countries.
- Other classifications are more problematic, e.g. residents **of** & **in** care homes.
- Race and ethnic group:
  - In U.S. for data years 2018 – 2020, race and Hispanic-origin categories are based on the 1997 Office of Management and Budget (OMB) standards. Not all states used these in 2015-17.
  - In UK, ONS study matched 2020 death certificates to census record for ethnicity. Would need to repeat for 2015-19 to find 'normal' death counts.
  - France does not record ethnicity in census.

Conclusions: need granular data in multiple classifications

- UK and US death certificates record occupation.
  - https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/whichoccupationshavethehighestpotentialexposuretothecoronaviruscovid19/2020-05-11
  - Study translates occupational codes into common ISCO to compare US and UK: locate occupations in ‘exposure to disease’ and ‘physical proximity’ space
  - Repeat for excess mortality? Standard international framework for comparison?
- Income:
  - US census recorded income. So could use last census record matched to death record to compute excess mortality by income group.
  - In many countries, hard to match the two. Could impute using occupation and location but probably not suitable for routine production of data.
  - https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand30june2020 this compares covid-19 deaths in high & low deprivation areas.
  - Article also gives age-standardised all-cause mortality rates, but not excess mortality.

Conclusions: models

- Models using granular data:
  - Forecasting P-scores from epidemiological and other models for different scenarios on ending lockdown measures should aid policy formulation.
  - Granular Italian death registry data, Ciminelli and Garcia-Mandicó (2020).
  - A study which forecasts the one year ahead mortality, Denaxas et al. (2020). Analyse daily death registry data for over 1000 Italian municipalities, which suggest that deaths registered as Covid capture only about half of excess deaths. They find strong evidence that locations where mass testing, contact tracing, and at-home care provision was introduced experienced lower numbers of excess deaths.

- Economic impact: is there a trade-off with health?

- New paper on Scandinavian nations, Kaiser et al. (2020) concludes:
  - Precautionary approach can be lowest cost, while still expensive;
  - Detection and monitoring (e.g. testing & tracking) are integral to a successful precautionary approach;
  - Economic activity and health are complements not substitutes.
    - https://cepr.org/content/covid-economics issue 39

- *Denmark, Iceland and Norway*:
  - chose a precautionary approach that formally shut down schools and businesses to protect human health.
- *Sweden*:
  - took a Business-as-Usual approach aimed at maintaining normal economic and social activities.
- *Iceland and Denmark*:
  - have further invested in testing, tracking and containing the disease.
- Similar economic costs **so far** of the pandemic and government fiscal & monetary interventions,
  - but *Sweden* has had the most severe loss of life.
- They use a panel from the *4 countries* from the beginning of pandemic:
  - to predict Covid-19 deaths, conditional on Covid-19-cases;
  - to model Covid-19 deaths as functions of behaviour and government interventions;
  - to estimate the additional lives lost if these interventions did not occur.

- Daily new cases depend on:
  - previous day's cumulative total cases, days since middle schools closed, and mobility data (from Google, classified into 6 locations e.g. retail, transit).

- Daily new recorded Covid-19 deaths depend on:
  - lagged cumulative total of cases, total number of tests, days since middle schools closed, and Google mobility data (6 locations).

- Simulate counterfactual outcomes, comparing 'value of statistical life lost' per capita with predicted loss of per capita GDP: find net benefit for Norway, Denmark & Iceland.

- **Measurement problems:** Covid-19 deaths may undercount excess deaths; death count by day recorded (vs. by day of occurrence) can obscure dynamics; case numbers depend on testing regime;
- **Specification issues:** mobility could be endogenous; how well and when care homes were protected should be included; a distributed lag of new cases would be better than the cumulative case number in their Covid-19 death equation

- **An improvement: re-run these models using better-measured data on deaths**
  - use *granular excess mortality data* for the dependent variable - disaggregated by age, and by region (big city vs. hinterland locations).