One year of Covid-19 in 5 major European countries: a comparative analysis of excess mortality

Gianluca Baio

Department of Statistical Science | University College London

g.baio@ucl.ac.uk
https://gianluca.statistica.it
https://egon.stats.ucl.ac.uk/research/statistics-health-economics/
https://github.com/giabaio
https://github.com/StatisticsHealthEconomics
@gianlubaio

GEOMED 2022, University of California, Irvine

14 October 2022

Follow our departmental social media accounts

check out our departmental podcast "Random Talks" on Soundcloud!

Statistical Science

Acknowledgements

This is joint work with

Konstantinoudis et al (2021)

Core team

- Marta Blangiardo (Imperial College London)
- Michela Cameletti (University of Bergamo, Italy)
- Monica Pirani (Imperial College London)
- Garyfallos Konstantinoudis (Imperial College London)
- Virgilio Gómez-Rubio (University of Castilla-La Mancha, Spain)

Other colleagues

- Amparo Larrauri, Imma León (Spain); Julien Riou, Matthias Egger (Switzerland); Paolo Vineis (UK/Italy)

Extra stuff...

- The datasets and code used in the analysis is available at https://github.com/gkonstantinoudis/ExcessDeathsCOVID
- The results are also provided in a ShinyApp available at http://atlasmortalidad.uclm.es/excess/

I'll take all the credit if you like this, but the blame is all on them if you don't... 😉

Nothing can be said to be certain, except death and taxes...

www.politico.eu

The trouble with coronavirus death tolls

David Spiegelhalter

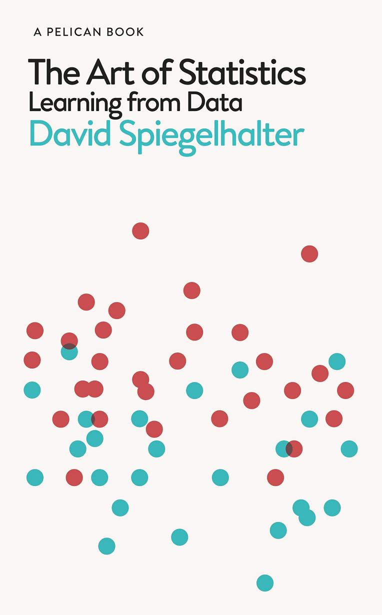
5 - 6 minutes

David Spiegelhalter is chair of the Winton Centre for Risk and Evidence Communication at the University of Cambridge and author of "The Art of Statistics: Learning from Data" (Penguin, 2019).

CAMBRIDGE, England — The British press has given a huge amount of attention to a grim milestone in the country: With the announcement on January 26 of an additional 1,631 COVID-19 fatalities, the total number of deaths caused by the pandemic has hit the 100,000 mark.

But there are problems with both of these headline numbers, and they are worth exploring — for reasons that go beyond statistical pedantry.

First, the daily COVID death count refers to reported deaths, which may have occurred some days previously, and tends to oscillate wildly due to the way reports come in.



> *"But in the US each state can have its own legal definition of death, and although the Uniform Declaration of Death Act was introduced in 1981 to try to establish a common model, some small differences remain. Someone who had been declared dead in Alabama could, at least in principle, cease to be legally dead were they across the border in Florida, where the registration must be made by two qualified doctors"*

Why excess mortality?

- The total impact of the COVID-19 pandemic on mortality should be the least controversial outcome to measure. But this is complicated by
  - Lack of real time cause specific data
  - Quality of coding on death certificate
- Excess mortality during the COVID-19 pandemic is the combination of deaths caused, or contributed to, by infection with SARS-CoV-2 plus deaths that resulted from the widespread behavioural, social and healthcare changes that accompanied national responses to the emergency.

Focus

- Much of the existing literature focus on national data
- In order to understand the dynamics of the pandemic there is the need to move to a sub-national level
  - Differences in the socio-demographics/environmental characteristics/healthcare provision
- Limited contribution, mainly for large regions

All cause mortality

- Data for all-cause deaths and population counts from official sources in the 5 countries
  - Italy
  - England
  - Spain
  - Switzerland
  - Greece
- Geographical resolution defined at *Nomenclature of Territorial Units for Statistics* (NUTS)
  - Specifically NUTS 3 level (small regions for specific diagnoses)
  - Some results aggregated back at NUTS2 level (basic regions for the application of regional policies)
  - National level
- Grouped by
  - Sex, age, week and NUTS3 region defined as areas with a population varying from 150,000 to 800,000, for 2015-2020
- Use of *International Organization for Standardization* (ISO) week calendar
  - Seven consecutive days beginning with a Monday and ending with a Sunday
- Mortality and population data by age groups
  - <40, 40-59, 60-69, 70-79 and 80 years and above

Population at risk

- Official **yearly** estimates for the population for 2014-2020
  - Directly available for Greece, Spain and Italy at 1 January of every year
  - For Switzerland, available at 31 December
  - England a bit different... (more on this later)

- Two-steps linear interpolation to predict the population at 1 January 2021
  1. Use data for 1 January 2015 to 1 January 2020 to predict population counts by age, sex and NUTS3 region at 1 January 2021
  2. Calculate **weekly** 2015-2020 population by linear interpolation of estimates on 1 January 2020 and 1 January 2021, by age, sex and NUTS3

For England, only mid-year figures are available

- But in 2020 these are affected by COVID-19 deaths during the first wave (March-May)
- Use yearly data for 2015-2019 to estimate the mid-year population in 2020 using linear interpolation
- Then use the estimated population at 1 January 2020 and linear interpolation to obtain the **weekly** population for 2015-2019
- Use weekly data for 2019 as a proxy for 2020

Ambient temperature

- Typically affects death rates $\Rightarrow$ use data on temperature from the ERA5 reanalysis dataset of the Copernicus climate data
  - Global in situ and satellite measurements
  - Provides hourly estimates
  - Available measurements compatible with spatial ***and*** temporal resolution for our analysis
- For each centroid of the grid cells (at $0.25^\circ \times 0.25^\circ$ resolution) that fall into the NUTS3 regions, calculate the daily mean temperature during 2015-2020 and then the weekly mean to align temperature and mortality data
- Additionally, as mortality from all causes can be different during national holidays, we also included a binary variable taking the value 1 if the week contains a public holiday and 0 otherwise

Main objective

Predict deaths in 2020 in the hypothetical scenario of no pandemic

Notation

For each country, define separately

- y_{jtsk} = number of all-cause deaths in week j of year t for NUTS3 area s and age-sex group k
  $k = 1, \ldots, K = 10$ = age-sex group (male/female and $<$40, 40-59, 60-69, 70-79, $\geq$80)
- P_{jtsk} = population at risk in week j of year t for NUTS3 area s and age-sex group k
- ρ_{jtsk} = the risk of death (mortality rate) in week j of year t for NUTS3 area s and age-sex group k
- z_j = dummy variable for public holiday
- x_{jts} = average weekly temperature in each area

NB: Main analysis excludes younger groups

- Based on cross-validation and poor predictive performance
- Not too surprising – less affected in the earlier waves of the pandemic...

$$y_{jtsk} \sim \text{Poisson}\left(\rho_{jtsk} P_{jtsk}\right) \qquad \log\left(\rho_{jtsk}\right) = \beta_{0t} + \beta_1 z_j + f(x_{jts}) + b_s + w_j$$

- $\beta_{0t} = \beta_0 + \varepsilon_t$: year-specific intercept
  - $\beta_0 \sim$ Normal$(0, 10^3)$: global intercept
  - $\varepsilon_t \sim$ Normal$(0, \tau_\varepsilon^{-1})$: unstructured random effect
- $\beta_1 \sim$ Normal$(0, 10^3)$: effect of public holidays

Non-linear effect of average weekly temperature $f(x_{jts})$

- RW2 model: $x_{jts} \mid x_{(j-1)ts}, x_{(j-2)ts}, \tau_x \sim \text{Normal}\left(2x_{(j-1)ts} + x_{(j-2)ts}, \tau_x^{-1}\right)$

Spatial component

- $b_s = \frac{1}{\sqrt{\tau_b}}\left(\sqrt{1-\phi}\tau_v^{0.5} v_s + \sqrt{\phi}\tau_u^{0.5} u_s\right)$: Besag-York-Mollié (BYM)-type model
  - $v_s \sim$ Normal$(0, \tau_v^{-1})$: unstructured random effect
  - $u_s \mid \boldsymbol{u}_{-s} \sim \text{Normal}\left(\frac{\sum_{r=1}^R n_{rs} u_r}{\sum_{r=1}^R n_{rs}}, \frac{1}{\tau_u \sum_{r=1}^R n_{rs}}\right)$: spatially structured random effect
  - $\phi \in [0,1]$: mixing parameter (measures proportion of variance explained by the structured effect)

Temporal component (non-linear weekly effect)

- RW1 model (accounts for seasonality): $w_j \mid w_{j-1}, \tau_w \sim \text{Normal}(w_{j-1}, \tau_w^{-1})$

Priors (hyperparameters)

All the hyperparameters are modelled using Penalised Complexity (PC) priors

- Regularise inference while not forcing too strong information
- Penalise departure from a "base" model (eg parameter = some fixed value)
  - Prior tends to favour the base model $\Rightarrow$ need fairly strong evidence to move away from it
  - Distance between the **base** model $g(\xi)$ and an **alternative**, more complex model $f(\xi)$ is measured by

$$d(f,g) = \sqrt{2\text{kld}(f,g)} \qquad \text{with} \qquad \text{kld}(f,g) = \int f(\xi)\log\left(\frac{f(\xi)}{g(\xi)}\right)d\xi$$

- Penalisation done at a constant rate

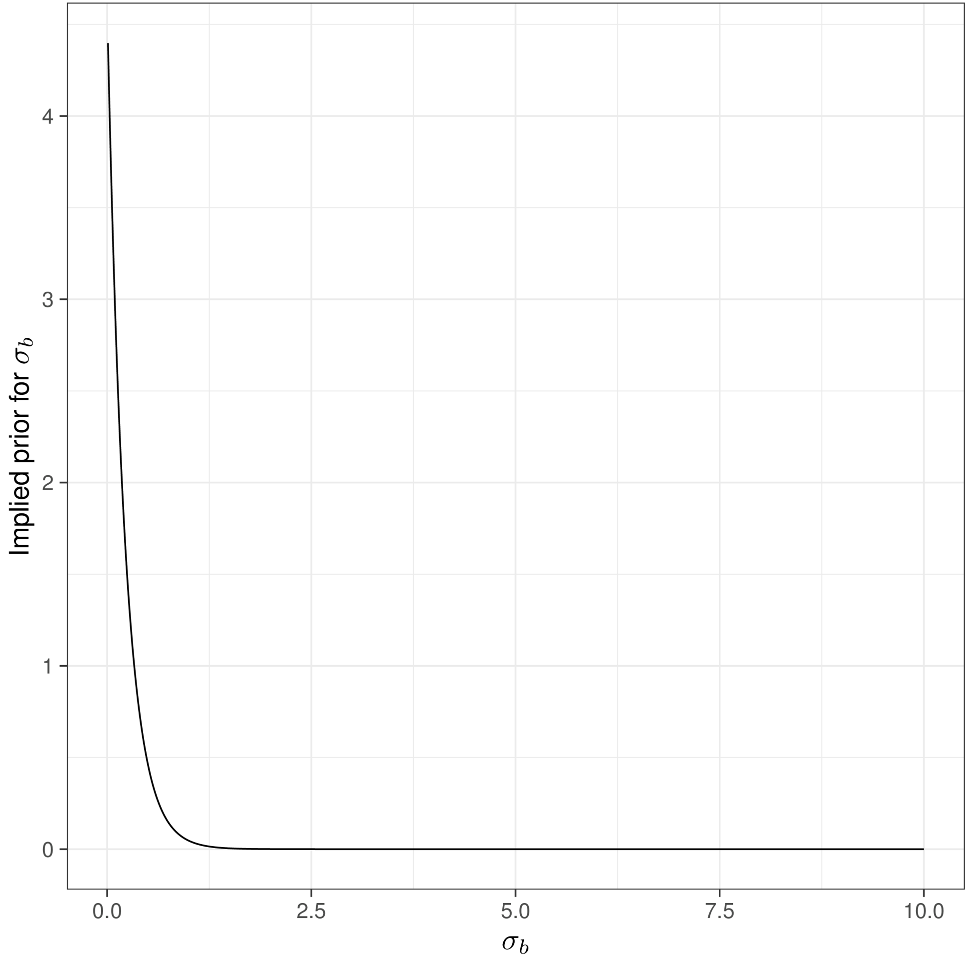
$$p(d) = \lambda\exp(-\lambda d) \sim \text{Exponential}(\lambda) \qquad \Rightarrow \qquad p(\xi) = \lambda e^{-\lambda d(\xi)}\left|\frac{\partial d(\xi)}{\partial \xi}\right|$$

- PC prior defined using probability statements on the model parameters (in the appropriate scale) to determine the value of λ using "reasonable" information

Priors (hyperparameters)

Spatial field

- Set $\Pr(\tau_b^{-0.5} > 1) = 0.01 \Rightarrow \lambda = -\log(0.01) \approx 4.61$
  - Basically implies $\sigma_b \sim$ Exponential(4.61)
  - *Very* unlikely to have a relative risk > $\exp(2)$, based solely on spatial variation
- Set $\Pr(\phi < 0.5) = 0.5$
  - Reflect lack of knowledge about which spatial component dominates the field
  - **NB**: Resulting distribution is non-standard



Variance components

- Set $\Pr(\sigma_\varepsilon > 1) = \Pr(\sigma_x > 1) = \Pr(\sigma_w > 1) = 0.01$

Training and prediction

- Use data from 2015-2019 to "train" the model
- Predict area-level weekly mortality for 2020 ($t = 6$), **in the hypothetical scenario** in which the pandemic hadn't occurred

$$p(y_{j6sk} \mid \mathcal{D}) = \int p(y_{j6sk} \mid \boldsymbol\theta) p(\boldsymbol\theta \mid \mathcal{D}) \boldsymbol\theta$$

  - $\boldsymbol\theta =$ all model parameters
  - $\mathcal{D} =$ observed data in 2015-2019
  - Mortality rates applied to 2020 come from the model trained on 2015-2019
- Compare observed deaths in 2020 with model predictions

This assumes **exchangeability** between 2015-2019 and 2020...

- Which is **obviously** an unjustifiable assumption – the pandemic **did** change the underlying data generating process!
- **But**: it allows us to measure the excess mortality

Model validation

Based on cross-validation

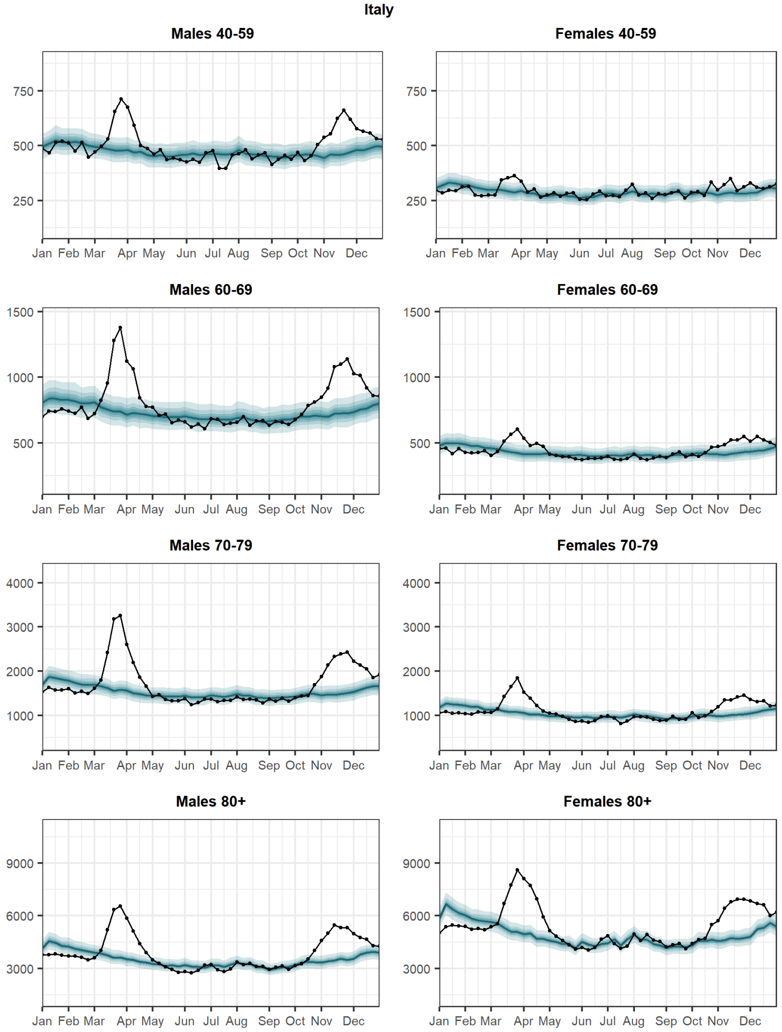
- Fit the model for 2015-2019 multiple times, leaving out one year at a time
- Predict the weekly number of deaths by NUTS3 region for the year left out
- Repeat for different age/sex groups and countries

Assess agreement based on

- Correlation between predicted and observed deaths
- 95% coveage = $\Pr\big($Observed deaths lie within 95% interval from the model$\big)$

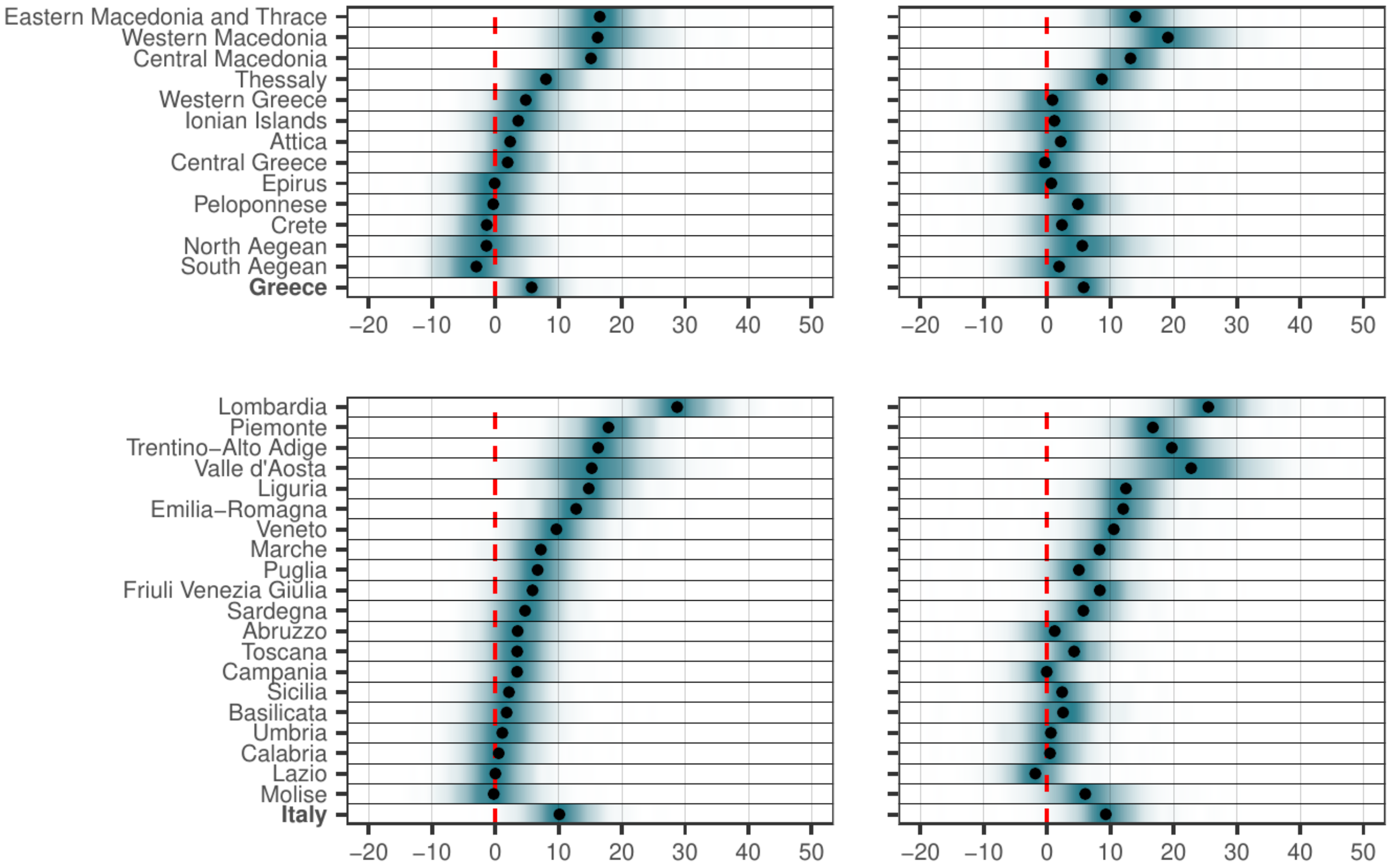
- Generally, models had good predictive ability
  - Highest correlation for >80 yo: 0.83 (0.82-0.84) for females/England to 0.97 (0.97-0.98) for males/Spain
  - Coverage range from 0.90 (females/Spain) to 0.95 (males/Switzerland)
- <40 yo had poorer performance
  - Coverage *close* to nominal 0.95, but correlation much lower $\Rightarrow$ excluded from base-case analysis

Country-level trends & excess mortality



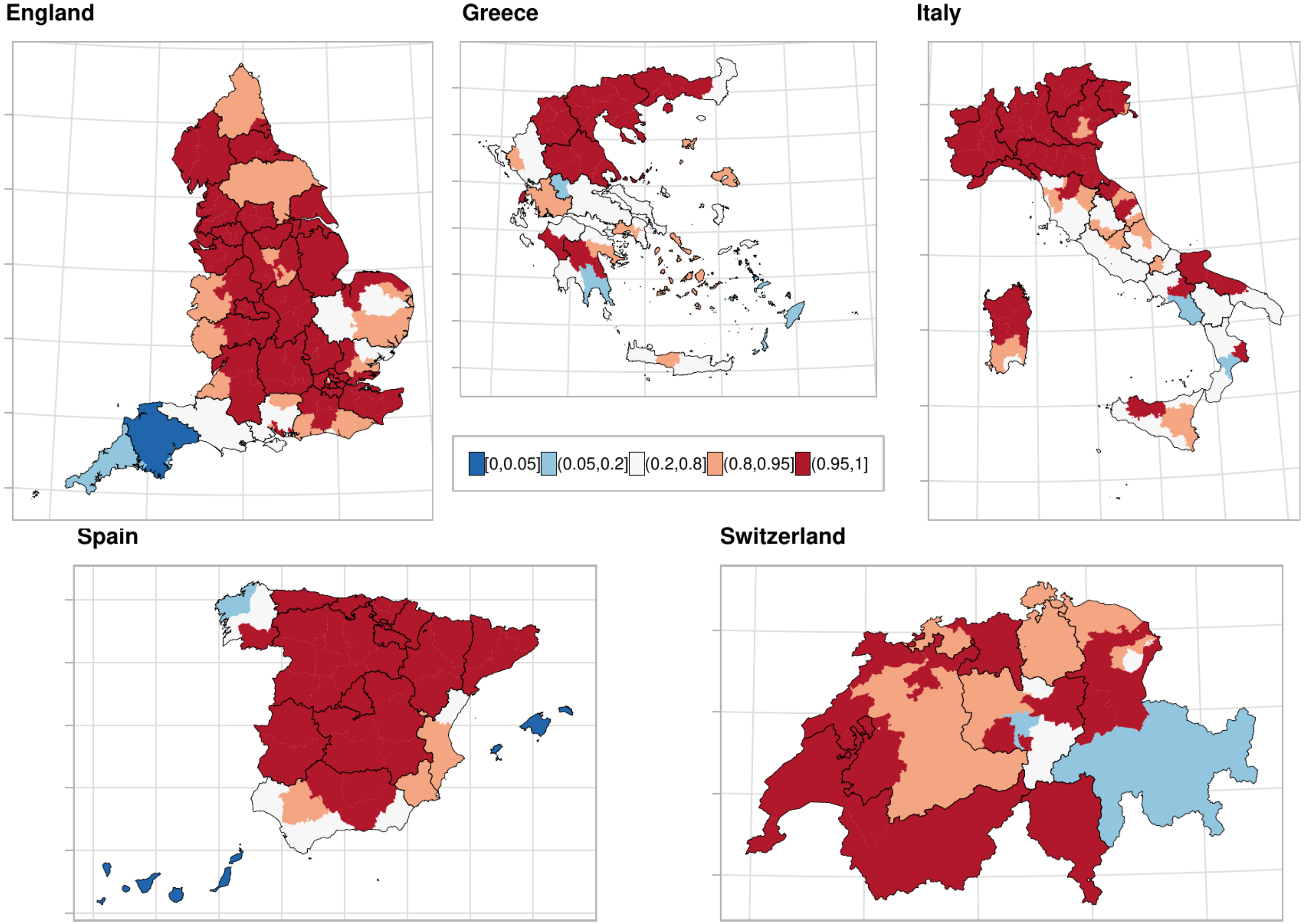
Sub-national level trends & excess mortality (NUTS2)

Relative excess death (%)



Sub-national level trends & excess mortality (NUTS3)

Probability that relative excess deaths is > 0%



Conclusions

- Wide variation in 2020 excess mortality both within and across countries
  - Spain seems to have experienced the largest excess mortality
  - Greece and Italy especially had a very strong spatial gradient
  - Temporal patterns in all countries (possibly less so for Greece)

- Results are generally in line with other findings in the literature
  - Slightly lower *point* estimates than **national** analyses for England (but intervals agree)
  - Consistent results for **national** estimates for Greece, Italy, Switzerland and Spain

- Overall, seem to suggest that a timely lockdown led to reduced community transmissions and, subsequently, lower excess mortality


thank