Public Health England

Protecting and improving the nation's health

Modelling excess mortality across England during a national pandemic

Faculty of Clinical Informatics
Sharmani Barnard*, Sebastian Fox*, Paul Burton*^

*Public Health Data Science, Public Health England
^Population Health Sciences Institute, Newcastle University

Picture the scene

- It is mid-February 2027: Covid-26 looms (seemingly resistant to vaccines created for Covid-19)
- How should the Nation and Public Health England (PHE) respond?
- Set up robust strategic plans for:
  - Rapid case identification and contact tracking
  - Covid-resilient health and social care systems
  - Population availability and use of appropriate PPE
  - Financial and fiscal buffers to mitigate economic impact
  - Developing a modified vaccine
  - Systems to track progress of epidemic (in as close to real time as possible) and to monitor stress on the health and care systems during epidemic waves. These can be used to help National and Local politicians, public health experts, GPs and other clinicians to work together to rationally guide and monitor community, health service and social care system responses to the evolving epidemic
  - One key component of these systems is to monitor "excess mortality"

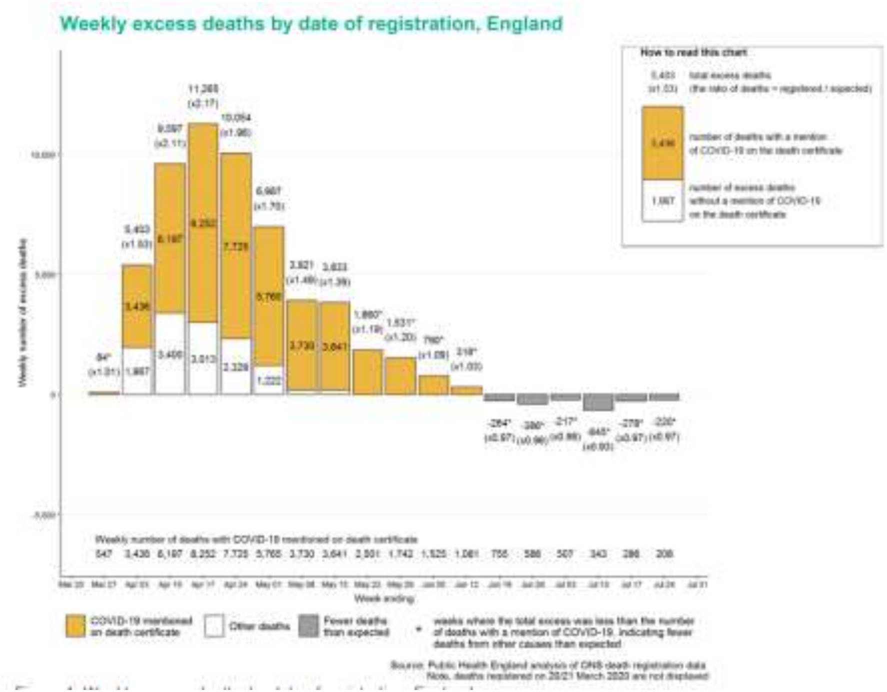
Covid-26: Global distribution

Excess Mortality

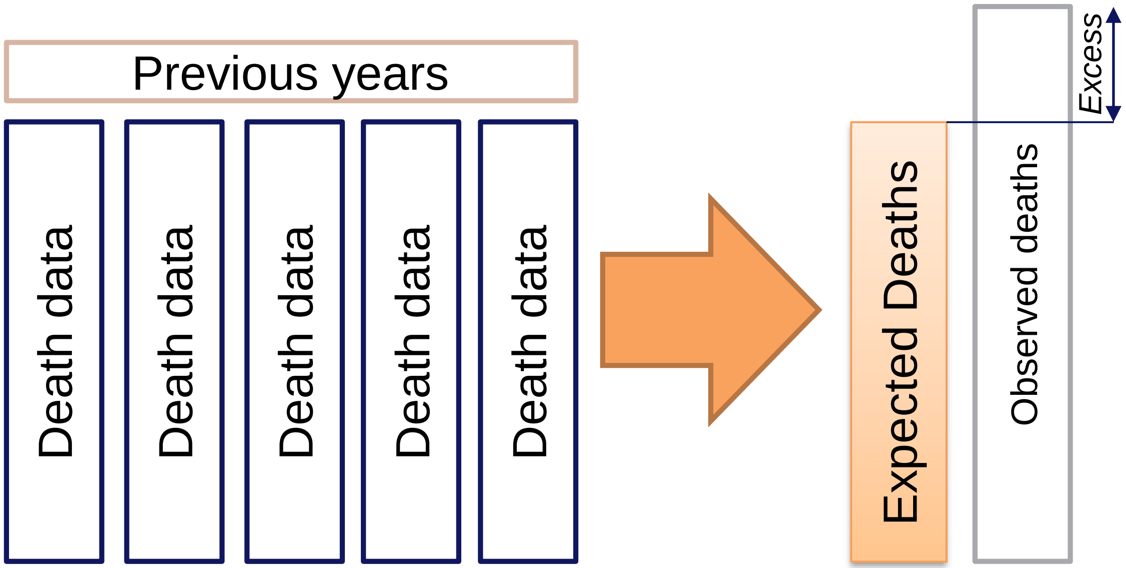
Provides an objective and comparable measure of the scale of the pandemic

Measuring excess mortality from all-causes overcomes the issues of variation in testing and differential coding of cause of death between individuals/areas and over time

Direct and indirect deaths



Overarching methodology



Excess mortality is the number of observed deaths compared to the number of deaths that would have been expected based on previous data

There are a number of methodologies

Variation between methods is generally around how the *expected deaths* are calculated

The choice of method used really depends on who the stakeholder is and what the ultimate use of the measure is

Where to start?

Needs Assessment

Identify stakeholders

Identify what they intend to use the measure for

- Real time surveillance
- Surveillance among specific groups
- Quantification of total excess

<u>**Considerations**</u>

- Urgency - how quickly it can be produced (from a standing start)
- Population subgroups - National/subnational, Age/sex, Ethnicity, Deprivation, Cause of death, Place of death
- Timeliness - How up-to-date do the reports need to be

Modelling groups

Clinical Directors

Politicians

Medical officers

Regional directors of PH

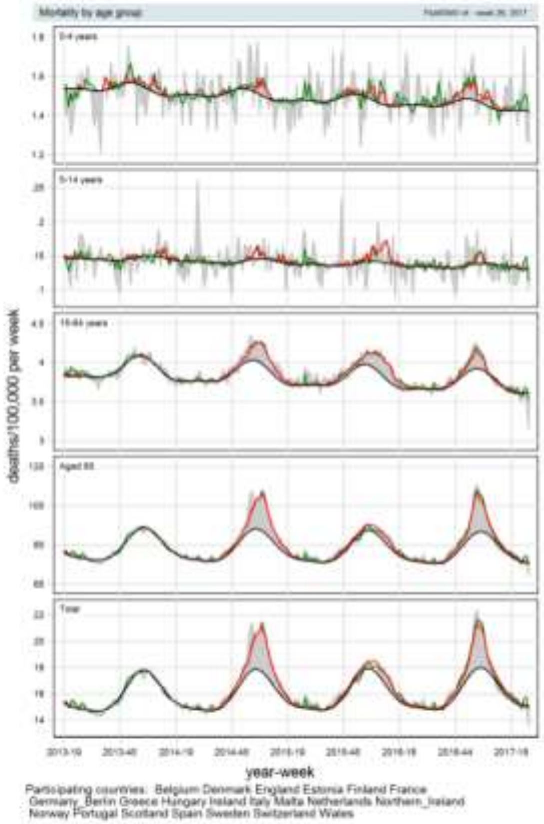
Academics

Groups interested in disparities

Existing methods

Surveillance activities

- FluMOMO and EuroMOMO
- Surveillance of flu
- Other surveillance activities



Note: The shaded grey areas represent deviations in expected deaths from the estimated baseline. The red curves signify mortality attributable to influenza activity, and the green curve signify effect of extreme temperatures, on top of the red curve.

Data Availability

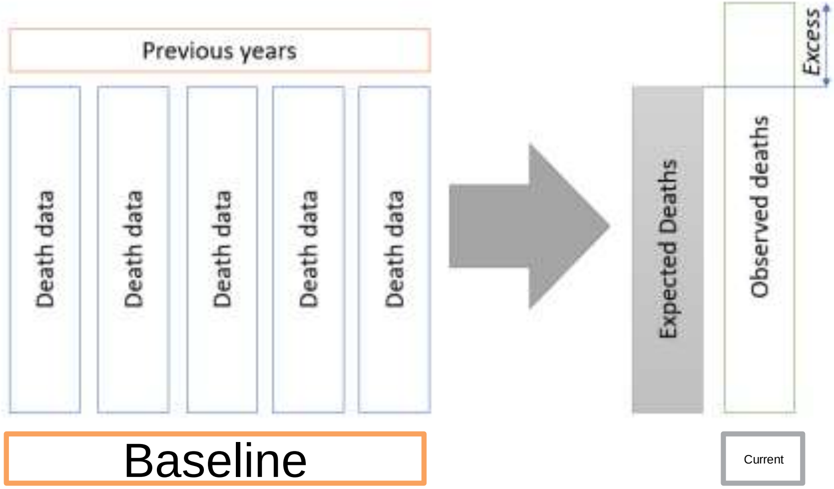
Identify availability of mortality data

- Current period
- Baseline

Establish frequency of data feed

Identify availability of denominator data

- Current period
- Baseline



Can subgroups be identified consistently across mortality data for the current and baseline period and, where relevant denominator data

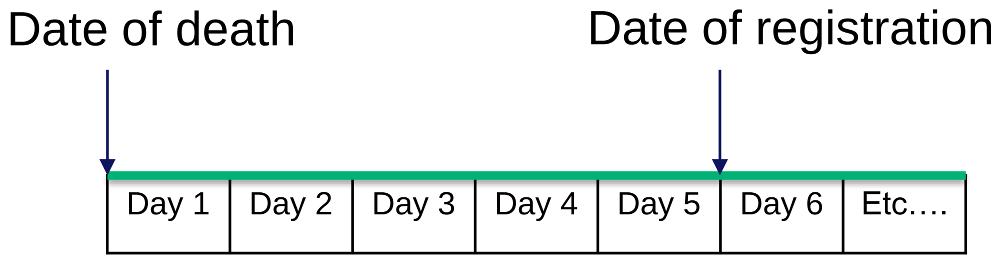
What other datasets are there that can be linked to obtain information on subgroups

Identify other information for parameters – weather, flu

Calculating the baseline

- Date of interest
- Choice of data
- Inclusion of baseline data
- Choice of model parameters
- Model type
- Prediction intervals
- Other considerations

Date of interest



(The point at which data becomes available)

Date of death
Ideal date for epidemic surveillance

However, you don't know a death has happened until it's been registered.

Requires a correction to the observed figure

Adds uncertainty to the estimate, particularly among subnational groups.

Requires at least 3 weeks before publication to estimate the delay between date of death and date of registration

Date of registration

Can produce robust estimates for subgroups quickly

Requires adjustments for bank holidays and possibly weekends

During the pandemic the delay between occurrence and registration may be different from non-pandemic times

Choice of data

Type of death data

- Counts
- Rates
- Calculate rates using counts and denominators
- Data availability (sub-groups)

Number of years in the baseline

- Previous years mortality rates
  - how many to include – were there any outliers?
- Sensitivity analysis

Model parameters

- Subgroups of interest
  - Age/sex/region/ethnicity/deprivation/place of death/cause of death
- Seasonality (day of week/time of year) or sine curves
- Dealing with year on year trend
  - Linear/non-linear
- Other factors
  - Bank holidays (date of reg)
  - Temperature/weather
  - Flu

Model type

Incorporating the baseline

- <u>Simple average</u> *e.g.* mean number of deaths in corresponding week (or day) over *e.g.* last 5 years
- Simple to understand ✓ (or [more correctly] *appears* simple X)
- No. of registrations severely biased around public holidays significant mortality events X

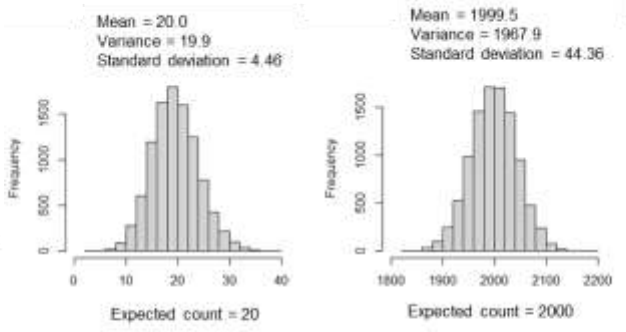
- <u>Mathematical Modelling</u>
- Build all key components into a suitable mathematical model, estimate the impact of each individual component using the historic data (*e.g.* over 5 years) then apply those estimated impacts that are relevant to the day/week of the current year for which expected number of registered deaths are to be calculated.
- Requires non-trivial modelling and careful explanation X. But once undertaken and method explained, the interpretation is truly straightforward ✓ Because underpinning method is basically "correct" it easily extends to other settings. ✓

Model components

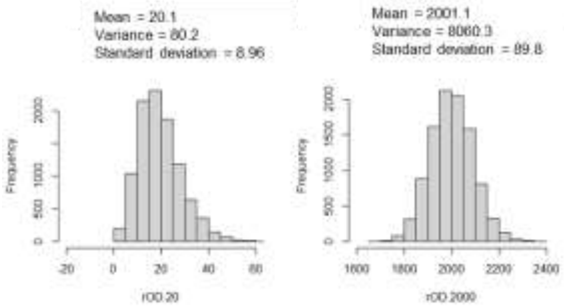
- Subgroups of interest
  - Age/sex/region/ethnicity /deprivation/place of death/cause of death
- Seasonality
  - Day of week/time of year; or direct seasonal effect, *e.g.* fourier analysis - sin/cos
- Year-on-year trend
  - Linear/non-linear
- Other factors
  - Bank holidays (date of reg)
  - Temperature/weather
  - Influenza outbreaks

Quasi-Poisson modelling

- Simple counts and rates are naturally “Poisson” distributed
- Link (prediction) function $\log_e(\text{deaths}) = \ldots. \; - 0.0245 \text{ (if Mon)} + 0.1217 \text{ (if Tues)} + 0.0376 \text{ (if Wed)} - 0.0052 \text{ (if Thurs)} \ldots.$
- Log scale so multiplicative — All else being equal if it is Tuesday, predicted deaths increase by $e^{0.1217} = 1.129$
- Uncertainty around the predicted number of deaths



- Over-dispersion and the Quasi-Poisson distribution

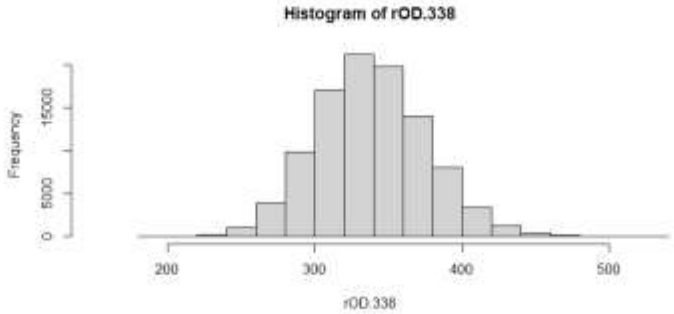


Prediction intervals vs confidence intervals

- Tuesday effect = 0.1217 SE=0.000786
- 95%CI = 0.1217 +/- 1.96*0.000786
- Log-scale: 0.1217 [0.1202 to 0.1232]
- Multiplicative on natural scale: 1.129 [1.128-1.131] (*i.e.* 12.8% -13.1% increase)

- Appears ***very precise*** but if the rest of the model predicts 300 in a subgroup the 95% uncertainty having taken account of the Tuesday effect is not 300*1.128 to 300*1.131 *i.e.* 338.4 to 339.3 deaths, but:

- 95% prediction interval = 270-414 deaths



Other model type issues

- Standard multiple (Gaussian) regression X
- Negative binomial models ✓
- Bayesian approaches ✓

Other considerations

Data linkage

- Is there data within other sources that can be linked to? E.g. ethnicity

Missing data

- How to deal with missing data (HES-ONS linked file)

Avoiding competing risk

- COD and POD analyses

Summary methodology for the PHE weekly reports

Date of interest: date of registration

Type of data: Rates, using counts and population denominators

Baseline: 5 years of historical data (sensitivity analysis was carried out)

Parameters: day of week (weekdays only), time of year for seasonality and (proximity to) bank holidays

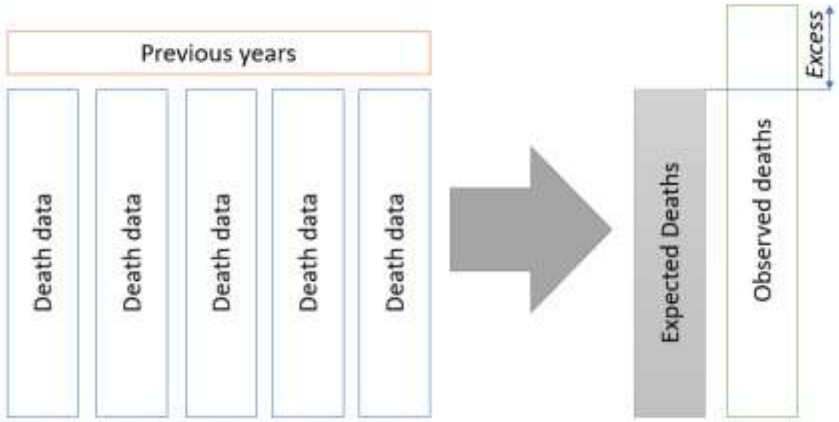
Subgroups: Age:sex, UTLA, ethnicity, deprivation, cause of death* and place of death*

Trend: Linear

Model: Quasi-Poisson model on the log scale

Linkage: linked to HES-ONS linked mortality for ethnicity data

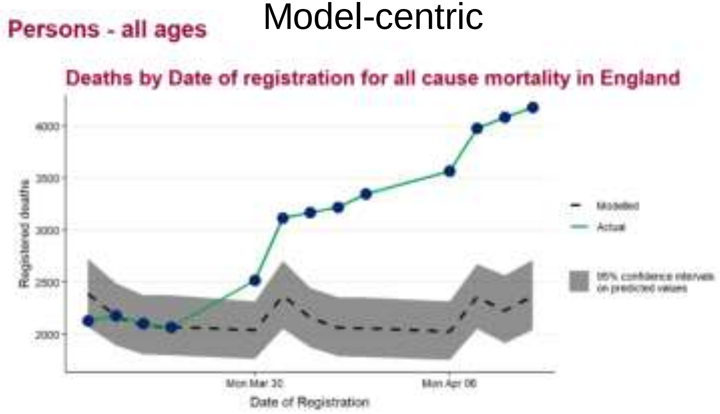
Missing data – uplift



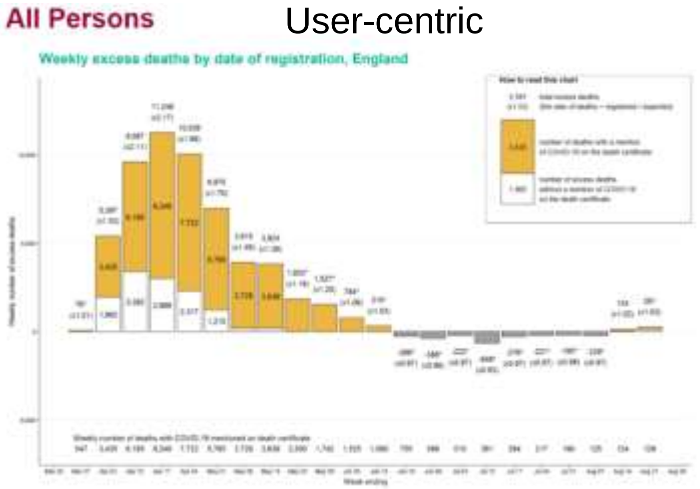
Data Visualisation

- What's important?
  - Total excess deaths
  - Registered / excess ratio
  - Time periods?
  - Weekly / cumulative
- Model-centric vs user-centric

Model-centric



User-centric



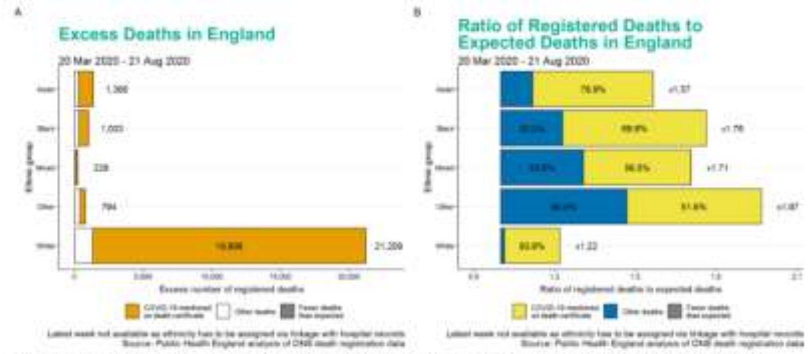


Figure 10: Cumulative excess deaths (A) and the ratio of registered deaths to expected deaths (B) by ethnic group, females, England.

Code development and ways of working

Rapidly evolving scope of project.

Coding principles:

- Robust to changes
- Reproducible analytical pipelines[1]
    - Open source
    - Version control
    - Functional programming
    - Testing

```
deaths_data %>%
  chain_start() %>%
  verify(has_all_names("UTLAApr19CD", "POD_out", "Sex", "Age_Group", "Reg_Date", "deaths_total")) %>%
  verify(nrow(.) == expected_records) %>%
  assert_rows(col_concat, is_uniq, UTLAApr19CD, POD_out, Sex, Age_Group, Reg_Date) %>%
  assert(not_na, everything()) %>%
  assert(within_bounds(0, Inf), deaths_total) %>%
  chain_end()
```

[1]See RAP Companion: https://ukgovdatascience.github.io/rap_companion/

Example report

[Example report]