Overview of Statistical Methodology for HQ TAG

“Tracking all cause of death and estimating excess mortality during the COVID-19 pandemic: statistical and computational tools”

World Health Organization
Western Pacific Region

HEALTH
EMERGENCIES
programme

Modeling frame

- We consider the case where we have multiple time-series of all-cause mortality counts from each member state for each week from January 1, 2015 to a recent date.
- To fix ideas, consider the case of females, aged 65-74 years in Australia.
- Let y_t be the mortality count for week $t = 1, \ldots, T$

  with $t = 1, \ldots, 260$ being the period January 1, 2015 to now

Model

- We model y_t as negative-binomial distribution with mean parameter λ_t.
- We make this choice rather than a Poisson distribution to account for overdispersion in the counts.
- The overdispersion parameter is itself estimated from the data and the mean parameters λ_t are modelled as

$$\log \lambda_t \; = c(t) + trend(t) + X_t \beta$$

Model for the mean

$$\log \lambda_t \; = c(t) + trend(t) + X_t\beta$$

where $c(t)$ represents the annual cycle in all-cause mortality

$trend(t)$ is the curvilinear trend of all-cause

The annual cycle $c(t)$ is modeled as a cyclic cubic spline function of time with a period of 52 weeks (that is, $c(t) = \; c(t+52)$).

Model for the mean

Choose c_t to minimize the penalized square error (PSE):

$$PSE_\tau(c) = \text{log–restricted–likelihood}\,(y, X, t = 1, \dots, T) - \tau \int_0^{52} c''[s]^2\, ds \qquad \tau > 0$$

where $c''[s]$ is the 2^{nd} derivative of $c[s]$ and τ is a smoothing parameter, chosen to balance the closeness of fit to the recorded counts (the first term) with the smoothness of $c[s]$ (the second term).

Specifically, we use Generalized Cross Validation (GCV) (Craven and Wahba 1979) to choose, and the R package **mgcv**

Model for the mean

Choose $trend(t)$ to minimize the penalized square error (PSE):

$$PSE_{\gamma}(trend) = \text{log–restricted–likelihood}\ (y, X, t = 1, \ldots, T) + \gamma \int_{0}^{260} trend''[t]^2\, dt \qquad \gamma > 0$$

However, there is great extrapolation uncertainty for $t > T$, so the trend penalty $\gamma \gg 0$ and we use a linear extrapolation. There is an argument to make the slope zero (i.e., remove the trend), but this depends on an assessment of each country. In our out-of-sample forecasting validation studies the linear trend was better for WPRO counties, but an assessment should be made for each country separately.

Expected death estimation

- The expected is then forecast stochastically to represent the uncertainty in the estimate of the expected.
- Thus, the statistical significance of the observed can be determined (i.e., if it is a substantial increase or decrease from the baseline).

Expected death estimation

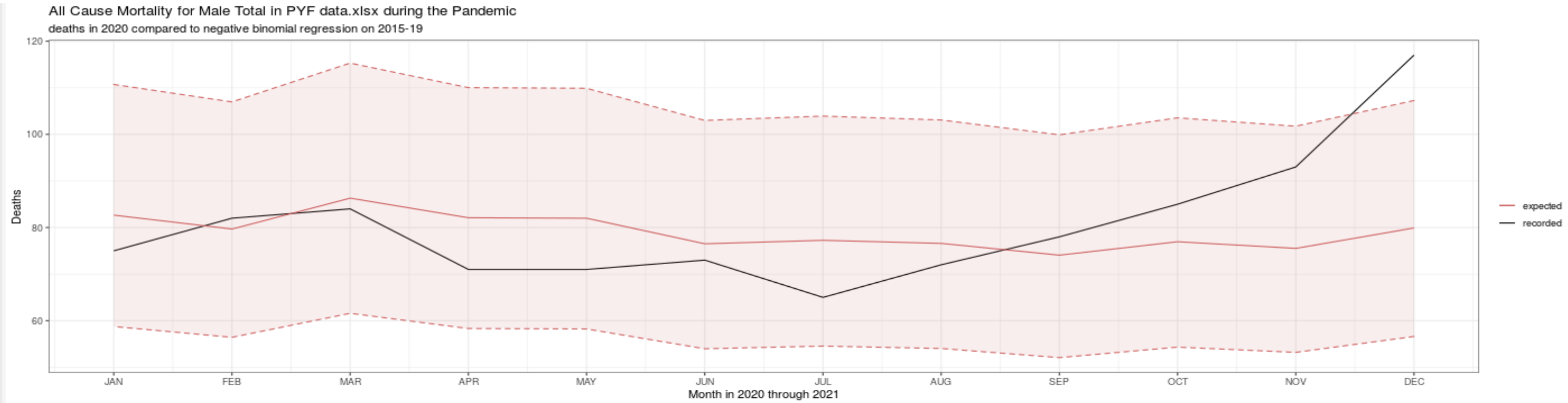
- One detail of the forecast is that it is an average over the sampling distribution of the parameter estimates.
- This is a simple way to account for uncertainty in our model for the expected deaths in addition to the sampling variation of the counts for given model parameters.
- We prefer this to a formal Bayesian model due to its simplicity.

Expected death modeling

In summary, we use a negative binomial generalized additive model (GAM)

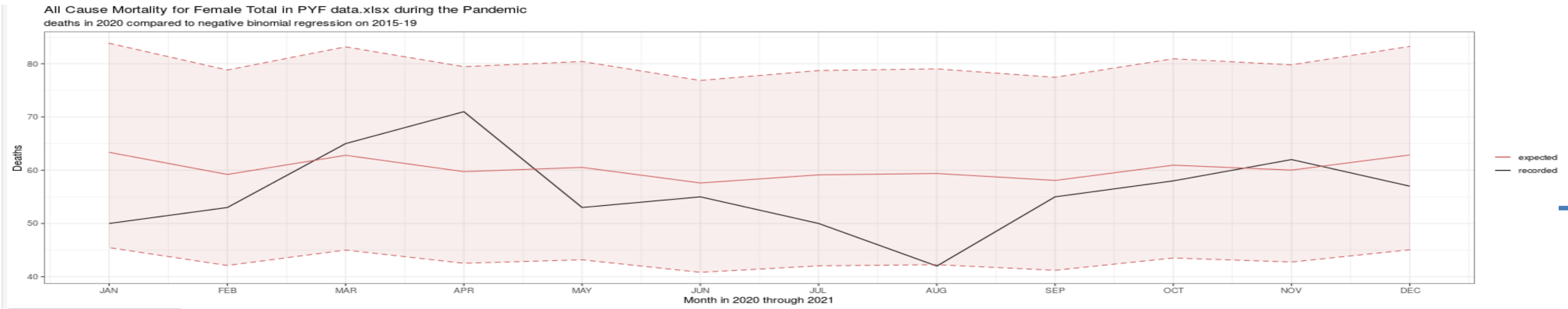
- At the moment, this model is very simple in that it uses no other information outside of sex, age-group, and time/date.
- Once more data becomes readily available, such as flu counts, the model can easily be extended to incorporate it.
- Can easily add hierarchical models for sharing information across groupings
- We do not have a model for the *reporting delay* as member states rarely can provide that information

Example output from Calculator



All Cause Mortality for Male Total in PYF data.xlsx during the Pandemic

deaths in 2020 compared to negative binomial regression on 2015-19



All Cause Mortality for Female Total in PYF data.xlsx during the Pandemic

deaths in 2020 compared to negative binomial regression on 2015-19

Western Pacific Region

HEALTH
EMERGENCIES
programme

Prediction Interval modeling age continuously