UCLA College
Statistics

Department of Statistics
8125 Mathematical Sciences Building
Los Angeles, California 90095-1554

April 7, 2021

Methodology used in

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

These notes provide technical details on the statistical methods used to estimate excess deaths in WPR Member States during the COVID-19 pandemic using all-cause of mortality data.

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. For some states we will have only monthly data, for which much of the description below maintains with natural changes. We consider the case where we have separate *reported* counts for each sex and age group (typically, five-year age groups).

The primary objective is to estimate the *expected all-cause mortality* counts for each week starting at January 1, 2020 onward in the counter-factual situation where there had not been a pandemic. The *excess mortality* is defined to be the difference between the reported counts and expected counts for that week.

Current Model

To fix ideas, consider the case of females, aged 65-74 years in Australia. Let y_t be the count for week $t = 1, \ldots, T$ with $t = 1, \ldots, 260$ being the period January 1, 2015 to December 31st, 2020. We model y_t as a random variable following a 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

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

where $c(t)$ represents the annual cycle in all-cause mortality and $trend(t)$ is the curvilinear trend of all-cause mortality over time. The annual cycle $c(t)$ is modeled as a cyclic cubic spline function (Wegman and Wright 1983) of time with a period of 52 weeks (that is, $c(t) = c(t + 52)$). A spline is a piecewise polynomial. Conceptually, one can imagine a high-degree polynomial capable of crossing through every data point. Such a polynomial would likely overfit the observed data, meaning it may not predict well using new data. Splines allow many low-degree (degree three, in this case) polynomials to fit the data in pieces. This achieves a good fit to the data without the risk of overfitting.

Specifically, c_t is modeled as a piecewise cubic polynomial that has a continuous second derivative, is continuous, has continuous 1^{st} and 2^{nd} derivatives at 52 week cycles and best fits the recorded all-cause mortality while being smooth. The specific criterion for the last feature is to choose c_t to minimize the penalized square error (PSE):

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

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). Hence, choosing the function $c[s]$ that minimizes $PSE_\tau(c)$ provides a balanced representation of the annual cycle. It prioritizes smoothness of $c[s]$ over the closeness of fit of $c[s]$ to the recorded all-cause mortality. Note that the traditional estimator, $c[s]$, is the minimizer with $\tau = 0$, that is, with no penalty for lack of smoothness. The choice of τ is subjective. In this work we choose to maximize the ability to predict unrecorded all-cause mortality counts. Specifically, we use Generalized Cross Validation (GCV) (Craven and Wahba 1979) to choose, and the R package mgcv by Simon Wood for analysis (Wood 2004, Wood 2017). The annual cycle so obtained is the optimal smoothest annual cycle chosen to maximize the likelihood of the observed all-causes mortality.

A similar approach is taken to the curvilinear trend $trend(t)$. It is modeled as a (non-cyclic) cubic spline function, specifically, as a piecewise cubic polynomial that has a continuous second derivative, is continuous, and best fits the recorded all-cause mortality while being smooth. The specific criterion for the last feature is to choose $trend(t)$ to minimize the penalized square error (PSE):

$$PSE_\gamma(trend) = \text{log-restricted-likelihood}\ (y, X, t = 1, \dots, T) + \gamma \int_0^{260} trend''[t]^2\, dt \qquad \gamma > 0$$

where $trend''[t]$ is the 2^{nd} derivative of $trend(t)$ and γ is a smoothing parameter, chosen to balance the closeness of fit to the recorded counts (the first term) with the smoothness of $trend(t)$ (the second term). Hence, choosing the function $trend(t)$ that minimizes $PSE_\gamma(trend)$ provides a balanced representation of the trend. It prioritizes smoothness of $trend(t)$ over the closeness of fit of $trend(t)$ to the recorded all-cause mortality. Note that the traditional estimator, $trend(t)$, is the minimizer with $\gamma = 0$, that is, with no penalty for lack of smoothness. Like τ, the choice of γ is subjective. As for the annual cycle, we choose to maximize the ability to predict unrecorded all-cause mortality counts by using the Generalized Cross Validation criterion. The model allows for arbitrary time-varying covariates, X_t. Including both the date and period allows for the model to detect trends across years and within years.

~~Poisson~~ Negative-binomial regression is a natural choice in that we are seeking to estimate the death count during any time frame. Negative-binomial is preferred to Poisson regression because it allows for overdispersion, and it can also account for instances of low or zero counts without issue.

This particular negative-binomial regression model is a generalized additive model (GAM) in that it uses smoothing functions for the predictor variables. Since the date and period are input as discrete values, they are smoothed using cubic splines, a common smoothing technique. The parameters β and the splines themselves are found through restricted maximum likelihood estimation (REML). GAMs are a type of generalized linear model, which are generalizations of ordinary linear regression that allow for the response variable to have error distributions other than the normal distribution (in this case, the negative-binomial distribution).

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. There are also other ways to enhance the model, such as considering negative-binomial regression for the case of overdispersion or using hierarchical models for sharing information across groupings. As such, this preliminary approach should serve as a strong starting point.

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). 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.

For the moment, models are fit separately to each sex and each age-group and each state. It is possible to improve the estimation by using information from both sexes and multiple age groups simultaneously. However, this is a bias-variance trade-off that can be explored.

For countries with missing (pandemic) weeks we can stochastically interpolate using simple time-series models. If there are significant missing weeks we will use a negative-binomial model like the above to stochastically interpolate.

An issue that may be important is to adjust for reporting delay (mainly an issue for recent weeks). To do this information is needed on the reporting delay. In the US, the NCHS reports deaths as they are received from the states and processed; counts of deaths from recent weeks are highly incomplete, reflecting delays in reporting. These "provisional" counts are updated regularly for past weeks, and the counts are not finalized until more than a year after the deaths occur. The estimate of completeness is based on the number of weeks that passed between the week in which the data set was obtained and the week in which the death occurred. We can model this relationship and use it to adjust the estimates, if necessary.

Validation of the statistical method for estimating the all-cause mortality without a pandemic
One may ask why it is not better to simply compare the observed all-cause mortality counts to historical averages of recent years. As we will show, doing so offers less robust prediction intervals than using the model described above. The following validation metrics detailed here also justify using this model to gauge the significance of current all-cause mortality counts relative to pre-pandemic times.
The model attempts to forecast all-cause mortality counts for each week of 2020 and beyond, assuming no pandemic had occurred. Since the discrepancy between actual counts and expected counts is the sought after estimate of excess mortality in 2020, it is vital the model makes accurate predictions. One way to validate the accuracy of the model is to use it to predict the number of all-cause deaths during 2019, a year in which there would be no 'excess' mortality. The model is trained on data from January 1, 2015 through December 31, 2018, and then predictions are made on a weekly/monthly basis for 2019. The closer the predicted counts are to the observed counts, the better the model is performing.
The model has been validated across all age groups, genders, and countries, but to keep with the example used earlier of females aged 65-74 years in Australia, we present those results. Figure 1 below shows the 95% prediction intervals for the model ("spline") as well as for the weekly average. The actual weekly counts are denoted by the black dots, showing that the spline model fails to capture the true count just three times out of 52 periods (95% accurate). The weekly average fairs far worse. As is evident from Figure 1, the lengths of the spline intervals are typically smaller than the lengths of the weekly average intervals, meaning the higher accuracy of the spline model is because it is better rather than just larger. More importantly, the weekly average intervals are misleadingly and their actual coverage is much below their nominal coverage.
The accuracy of the spline model is not solely for the females aged 65-74. Table 1 shows percent accuracy (i.e., how often the prediction interval contains the actual value) for each demographic breakdown. The spline model significantly outperforms the weekly average across all genders and age categories.

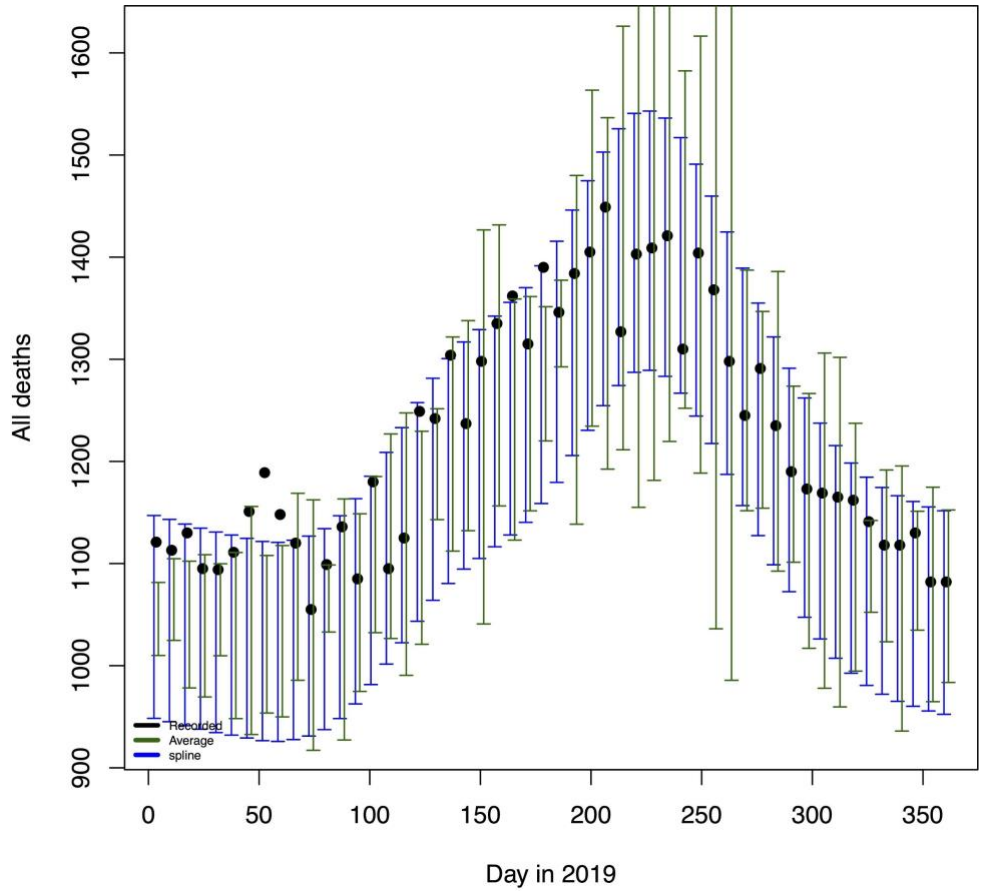

Figure 1: Prediction intervals for 2019 based on the deaths from 2015-2018. The black dots are the reported deaths for each week in 2019. The green error bars are those based on the weekly averages. The blue intervals are those based on the spline model. Those based on the weekly averages are incorrect and their actual coverage is well below their nominal coverage. The intervals based on the spline model are valid.

Age group	Average (PI %)	spline (PI %)
Female 0-44	85	94
Female 45-64	83	93
Female 65-74	81	91
Female 75-84	92	93
Female 85 and over	87	96
Female Total	88	95
Male 0-44	83	89
Male 45-64	81	97

Male 65-74	92	92
Male 75-84	87	91
Male 85 and over	81	87
Male Total	75	86
Total 0-44	87	89
Total 45-64	85	95
Total 65-74	88	90
Total 75-84	81	92
Total 85 and over	81	95
Total Total	83	91
median %	84	92
mean %	84	92

Table 1: Prediction interval accuracy for all age/gender groups. The intervals produced by the spline model have the correct coverage while those produced by the weekly average model are well below their nominal coverage.

Another way to check the validity of the model is to look at the length of the prediction intervals. The intervals should be long enough to capture the true values most of the time, but intervals too long leave too much uncertainty to be worthwhile. Table 2 shows the lengths of the prediction intervals for the spline, ETS, and weekly average. The spline intervals tend to be nearly the same length as those of the ETS for ages 0 - 74 (the weekly average has a short length but is woefully inaccurate). It is in ages 75+ (and when aggregating across all age groups) that the spline intervals are longer than its counterparts. The significant increase in the uncertainty surrounding the older age categories is something that will be looked into.

Country / Age group	Average (PI length)	ETS (PI length)	spline (PI length)
AUS: Female 0-44	8.9	20	20
AUS: Female 45-64	18.6	43	43

AUS: Female 65-74	24.9	52	53
AUS: Female 75-84	37.8	70	78
AUS: Female 85 and over	70.3	103	148
AUS: Female Total	106.5	143	225
AUS: Male 0-44	10.0	21	21
AUS: Male 45-64	23.6	50	53
AUS: Male 65-74	36.0	63	72
AUS: Male 75-84	45.1	78	93
AUS: Male 85 and over	60.2	87	118
AUS: Male Total	107.2	143	210
AUS: Total 0-44	13.8	29	29
AUS: Total 45-64	28.3	66	66
AUS: Total 65-74	50.8	81	98
AUS: Total 75-84	64.7	105	130
AUS: Total 85 and over	111.9	135	228
AUS: Total Total	191.0	202	386
median length	41.5	74	85
mean length	56.1	83	115

Table 2: Prediction interval length for all Australian age/gender groups.

References

Splines in Statistics (article)

Author
Edward J. Wegman and Ian W. Wright
Journal
Journal of the American Statistical Association
Year
1983
Volume
78
Number
382
Pages
351--365
Issn
01621459
Publisher
[American Statistical Association, Taylor & Francis, Ltd.]
Url
http://www.jstor.org/stable/2288640

Stable and efficient multiple smoothing parameter estimation for generalized additive models (article)

Author
S. N. Wood
Journal
Journal of the American Statistical Association
Year
2004
Volume
99
Number
467
Pages
673-686

Generalized Additive Models: An Introduction with R (book)

Publisher
Chapman and Hall/CRC
Year
2017
Author
S.N Wood
Edition
2