Title

Tool for tracking all-cause mortality and estimating excess mortality to support the COVID-19 pandemic response: All-cause mortality calculator for COVID-19 response

Journal

Western Pacific Surveillance and Response, 13(2)

ISSN

2094-7321

Authors

Mengjuan, Duan
Handcock, Mark S
Blackburn, Bart
et al.

Publication Date

2022-08-04

DOI

10.5365/wpsar.2022.13.2.921

Tool for tracking all-cause mortality and estimating excess mortality to support the COVID-19 pandemic response

Duan Mengjuan,[a] Mark S. Handcock,[b] Bart Blackburn,[b] Fiona Kee,[a] Viema Biaukula,[c] Tamano Matsui[c] and Babatunde Olowokure[c]

Correspondence to Babatunde Olowokure (email: olowokureb@who.int)

Problem: Quantifying mortality from coronavirus disease (COVID-19) is difficult, especially in countries with limited resources. Comparing mortality data between countries is also challenging, owing to differences in methods for reporting mortality.

Context: Tracking all-cause mortality (ACM) and comparing it with expected ACM from pre-pandemic data can provide an estimate of the overall burden of mortality related to the COVID-19 pandemic and support public health decision-making. This study validated an ACM calculator to estimate excess mortality during the COVID-19 pandemic.

Action: The ACM calculator was developed as a tool for computing expected ACM and excess mortality at national and subnational levels. It was developed using R statistical software, was based on a previously described model that used non-parametric negative binomial regression and was piloted in several countries. Goodness-of-fit was validated by forecasting 2019 mortality from 2015–2018 data.

Outcome: Three key lessons were identified from piloting the tool: using the calculator to compare reported provisional ACM with expected ACM can avoid potential false conclusions from comparing with historical averages alone; using disaggregated data at the subnational level can detect excess mortality by avoiding dilution of total numbers at the national level; and interpretation of results should consider system-related performance indicators.

Discussion: Timely tracking of ACM to estimate excess mortality is important for the response to COVID-19. The calculator can provide countries with a way to analyse and visualize ACM and excess mortality at national and subnational levels.

PROBLEM

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified in late December 2019 and declared a pandemic by the World Health Organization (WHO) on 11 March 2020.[1] In the WHO Western Pacific Region, by the end of November 2021, there were 10 221 280 confirmed COVID-19 cases and 141 864 deaths.[2] Although the COVID-19 death count is essential to understanding the epidemiology of COVID-19, the attributable mortality due to COVID-19 remains unclear. In any given country, official statistics may not reflect the actual number of lives lost to the disease.[3]

Identifying deaths from COVID-19 is difficult, especially in low-resource settings.[4] Many countries have limited capacity for COVID-19 testing at national and subnational levels, and therefore no capability to track the spread of COVID-19. Even where cases are adequately detected, some deaths may not be reported promptly or even at all.[4] Also, reporting of cause of death may be inaccurate because the quality of death certification depends on the knowledge and skills of physicians, on the characteristics of the deceased person (older people are the most difficult to certify correctly), on errors in coding the death event and on the format of certification.[5] There can also be a long lag between the death occurring and being certified, especially for deaths outside hospitals or other health-care facilities, or those that require an autopsy. Service interruptions due to the pandemic may further delay the death certification process.

According to an internal rapid assessment in the WHO Western Pacific Region, most Member States have two to four death reporting systems. Most systems are

[a] Division of Data, Strategy & Innovation Team, World Health Organization Regional Office for the Western Pacific, Manila, Philippines.
[b] Department of Statistics, University of California, Los Angeles, CA, United States of America.
[c] Division of Health Security and Emergencies, World Health Organization Regional Office for the Western Pacific, Manila, Philippines.
Published: 25 May 2022
doi: 10.5365/wpsar.2022.13.2.921

electronic or partially electronic, and although some are well-integrated within civil registration and vital statistics systems, others are disjointed. The United Nations Statistics Division estimated that death registration coverage is over 80% in 15 of the 27 Western Pacific Regional Member States with data available.[6] Total death counts, reported either weekly or monthly, are publicly available from at least six Member States, and data are available internally from at least four. Thus, it may be feasible for several Member States in the WHO Western Pacific Region to track all-cause mortality (ACM) to provide timely information on COVID-19 deaths. Ideally, deaths would be reported as soon as possible, with more detailed information (e.g. cause of death) reported later when death certificates become available.

CONTEXT

Tracking current ACM and comparing it with expected ACM from pre-pandemic data can provide an estimation of the overall burden of mortality potentially related to the COVID-19 pandemic.[4] This method requires first estimating the number of deaths that would be expected if the COVID-19 pandemic had not occurred (i.e. expected deaths) using historical data and a sophisticated statistical model.[7] Excess mortality is then estimated by comparing the current reported provisional deaths with the expected deaths.[8]

The excess mortality may be directly or indirectly due to COVID-19. Indirect deaths due to COVID-19 include those linked to conditions that were present before the pandemic and have resulted in death because health systems were overwhelmed, those due to patients avoiding health-care facilities and those linked to routine service delivery interruption for non-COVID-19 disease. These indirect deaths due to COVID-19 are not captured in the COVID-19 death numbers reported to WHO.[9] Given that COVID-19 deaths can influence national and subnational response measures, additional effort is required to ensure that this information is readily available and quickly tracked.

A common method to estimate the expected ACM is to use the average death count for each week over a 5-year period. However, this method does not account for the seasonality of mortality, or for the trend and smoothness of expected mortality from week to week or month to month. Additionally, if a trend is present over time, using historical averages will not capture the trend or allow it to be projected into the future. A more sophisticated method by Weinberger et al.[10] fits Poisson regression models that adjust for seasonality, year-to-year baseline variation, influenza epidemics and reporting delays. Our statistical model, the WHO Western Pacific Regional Office ACM calculator (hereafter, the ACM calculator), is based on this method.

ACTION

The WHO Western Pacific Regional Office ACM calculator

The ACM calculator was developed to assist Member States in the WHO Western Pacific Region in tracking and analysing ACM.[11] The user enters the relevant ACM data into the designated template in the calculator, and the expected ACM and excess mortality are calculated.

The calculator can be used online or installed onto a local machine. The input data are never stored offline and are only accessible to the user. Depending on the amount of data entered, the calculator will finish computing within seconds or minutes. Various outputs are available, including disaggregated results; for example, the calculator can provide expected ACM by age group and sex if the data entered are disaggregated by these factors. The results can be displayed in a variety of formats, including tables and graphs.[11]

Statistical methods

The ACM calculator is based on the model of Weinberger et al.,[10] but uses non-parametric negative binomial regression. This approach was preferred to Poisson regression because it allows for overdispersion and can account for instances of low or zero counts.[10] The mean function includes a smooth trend and a smooth non-parametric annual cycle in mortality over time. These terms were specified using cubic smoothing splines, including a cyclical one for the annual cycle. The model allows for arbitrary time-varying covariates, and the parameters were estimated through restricted maximum likelihood estimation. The methodology does not currently adjust for influenza epidemics and reporting delays because this information is not consistently reported.

The expected ACM deaths are forecast stochastically, to represent uncertainty in the estimate of the expected. Thus, statistical significance in observed data can be determined (i.e. a substantial increase or decrease from the baseline). The forecast is an average over the sampling distribution of the parameter estimates, which is a simple way to account for uncertainty in the expected deaths, in addition to the sampling variation of the counts for given model parameters. This approach is preferred to a formal Bayesian model because of its simplicity. The model goodness-of-fit was validated by forecasting 2019 mortality from 2015–2018 data (see **Appendix** for details). The validation indicated that the statistical coverage of the procedure is close to its nominal rate and that the prediction interval lengths are smaller than those based on the historical average model. The intervals based on the historical average are misleading and their actual coverage is far below their nominal coverage.

The calculator was developed using R statistical software (ver. 4.1.2), which includes the estimation of historical patterns and the computation of expected ACM. The software computes the excess mortality from 2020 to the present time; displays different visualizations of expected ACM and excess mortality and allows these visualizations and their raw data to be downloaded for further analysis and inclusion in reports; and includes interactive help and documentation of the methodology. The software is open-source. For reproducibility purposes, the exact code used for the analyses in this paper is in a static archive.[13]

OUTCOME

The ACM calculator has been tested using publicly available data from several Member States. Two examples are provided to highlight key lessons from implementing the calculator.

The first example from one country (January through September 2020) compares ACM plots using the calculator versus ACM plots based on historical averages only. The results from the calculator showed that the recorded counts were well within the 95% prediction interval generated (**Fig. 1A**). Although the reported counts were sometimes above the expected counts (most notably in August), the reported counts were always within the prediction interval. In contrast, the recorded counts based on the historical average only were well above the historical average (**Fig. 1B**) but confidence intervals and statistical increase were not calculated. The calculator values are above the historical average because of the presence of an upward trend in reported counts from 2015 to 2019; the calculator takes this into account whereas the historical average does not. Because historical averages do a poor job of predicting, comparison with the monthly average alone would lead to false conclusions.

The second example illustrates the ability of the calculator to show hidden excess mortality within subregions based on disaggregated data. Using data from another country, the national data indicate no excess mortality over a particular period (**Fig. 2A**), whereas the data for that period from a single local region show excess mortality during July and August that is outside the 95% prediction intervals for these months (**Fig. 2B**). Therefore, the excess mortality for July and August is statistically significantly different from zero (even after adjusting for multiple comparisons).[7] This example highlights the value of being able to analyse subregions, because excess mortality may not be identifiable at the national level in some cases.

Lessons identified

Three key lessons were identified from piloting the tool: using the calculator to compare reported provisional ACM with expected ACM can avoid potential false conclusions from comparing with historical averages alone; using disaggregated data at the subnational level (e.g. by region, sex and age) can detect excess mortality by avoiding dilution of total numbers at the national level; and interpretation of results should consider system-related performance indicators such as system coverage, completeness and reporting delays.

Suggestions for interpreting results

Given that the quality of mortality reporting varies greatly within and between Member States, the results of the ACM calculator should be interpreted with caution. Death coverage may differ if mortality reporting systems do not cover all death counts, with inconsistencies if a country has multiple systems, especially in low-resource settings. Civil registration of deaths is often below 20% in low- and middle-income countries.[4] There are also timeliness issues and reporting delays, so the death count may be

Fig. 1. **A) Monthly reported ACM compared with expected ACM for the first 9 months of 2020 using the calculator. The red zone is the 95% prediction interval. B) Monthly reported ACM compared with the expected ACM and the historical average ACM. The blue lines plot the recorded number of deaths, the orange the expected number of deaths under the model and the green the average number of deaths by month during 2015–2019.**

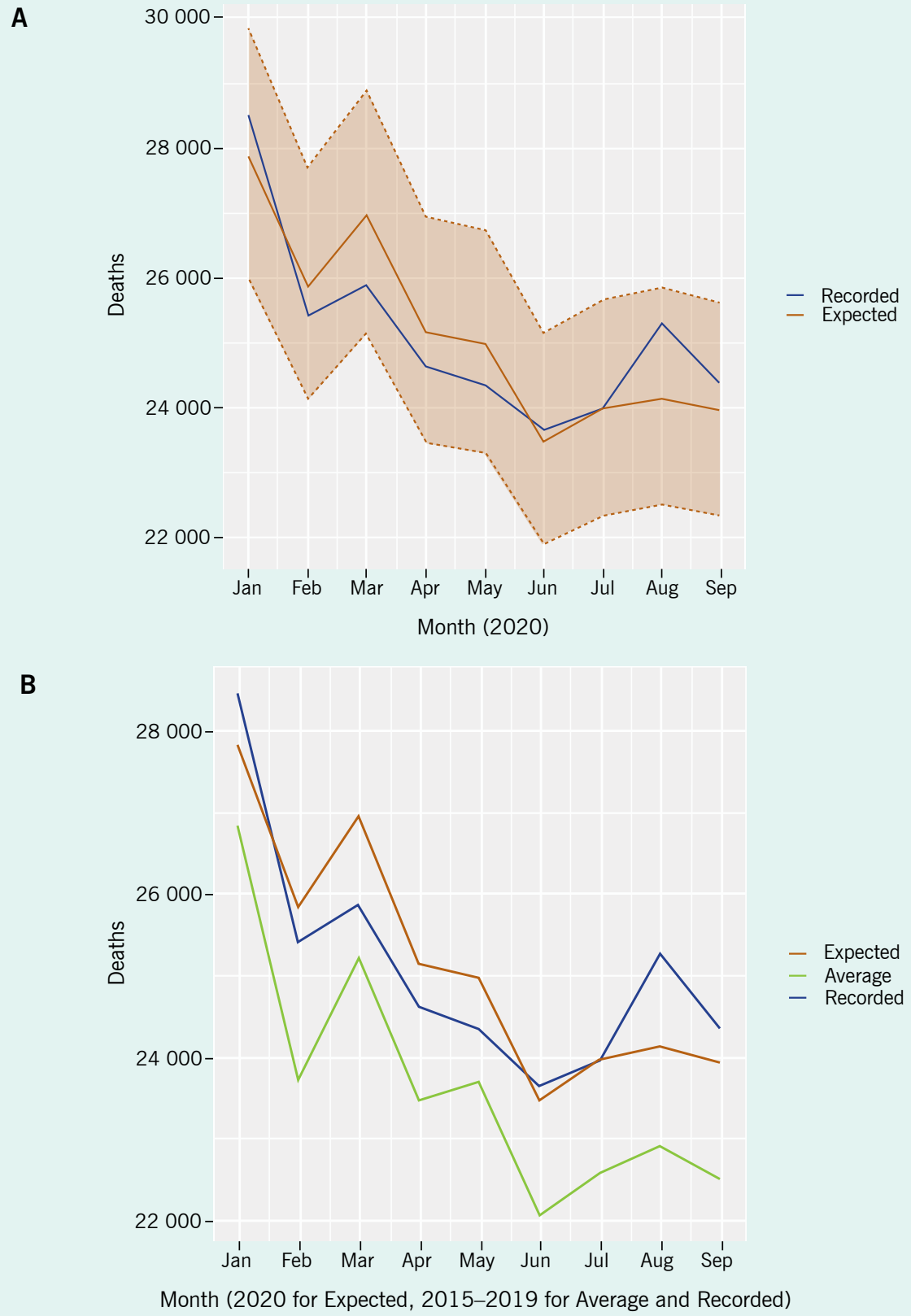


Fig. 2. **ACM at the national (A) and subregional level (B) within the same member state in the WHO Western Pacific Region. Looking at the aggregate would lead to a conclusion of no excess mortality present; however, by disaggregating the data into subregions we can identify areas where significant excess mortality is present.**

incomplete for certain periods (e.g. the latest week or month). It can take more than 12 months for mortality data to be finalized at the national level owing to deaths not being promptly reported or registered by subnational authorities, a long lag between a death and completion of the death certificate, a backlog at the subnational level that delays reporting to the national level and long processing times for the reporting systems. The use of disaggregated data to improve monitoring sensitivity may be affected by differences in the severity of COVID-19 transmission between subnational regions; also, the impact may vary among different population groups (e.g. by sex, age and occupation).

Proactively tracking ACM at the local level may help to capture more timely information, given that reporting and validation from the local to the national level may take several months to complete. Also, in both the short and long term, careful interpretation of the results is crucial to tailor specific actions based on conditions within each Member State.

For countries with existing systems that cover compulsory and universal mortality reporting, it is important to make use of the existing data to monitor weekly and monthly trends, to drive decision-making. For countries with low levels of mortality reporting coverage, it is still worth monitoring weekly and monthly trends based on available data; however, results should be interpreted with caution, as mentioned above. Additional resources or channels (e.g. burial or cemetery registration) can be employed to track total death counts. Community based mortality reporting should also be considered if necessary.

Limitations

There are two main limitations to the calculator. First, our methodology assumes that reported counts are the actual values and that reports are complete and accurate. However, provisional death counts are normally used for timely monitoring. Results should be compared with in-place systems, as mentioned above. Second, the fundamental assumption is that the statistical variation in ACM during the historical period (2015–2019) is the same as that from 1 January 2020 onward in the counter-factual situation where there was no pandemic. This is not directly testable because of confounding by the pandemic itself. In addition, it is assumed that the negative binomial regression model is adequate to capture this variation, and that counts are independent from period to period (conditional on the annual cycle and covariates). If these assumptions are incorrect, the estimates and prediction intervals will be inaccurate and probably overly optimistic.

DISCUSSION

During an epidemic or other public health emergency where mortality occurs, such as the COVID-19 pandemic, many countries experience disruption to routine health-care services and socio-behavioural changes in the population. For example, 90% of countries have reported disruptions to essential health services since the COVID-19 pandemic began.[12] These changes, together with a lack of reliable data and reporting systems, make the true burden of the pandemic difficult to quantify. ACM, when reported in a timely manner, can be used to estimate excess mortality, providing a rapid snapshot of the situation to support decision-makers to identify the extent and progression of the pandemic. Analysing and interpreting ACM data (including disaggregated data) can also provide important information about who is dying and where, which can then guide decisions on targeted surveillance and efficient use of health resources. The ACM calculator was developed to make it easy for Member States to analyse and visualize their ACM data. Users reported that the tool allowed them to analyse data on their own and easily generate results. Although the underlying statistical model is sophisticated, the use of complex algorithms in the background provides state-of-the-art summaries in the foreground. The model is standardized for a broad user base but could be customized for the needs of specific Member States. However, caution should be exercised when interpreting the results.

Acknowledgements

This article and the calculator described herein are the result of close collaboration between different units of the WHO Regional Office for the Western Pacific and all country offices in the region. We would like to thank all the emergency and health system focal points. We are also grateful for the great support from Gao Jun in the Health Information and Innovation Unit of the Division of Data, Strategy and Innovation in the WHO Western Pacific Regional Office.

Conflicts of interest

The authors declare no conflicts of interest.

Ethics statement

This work did not require ethics committee review because it did not involve human participants or active intervention. All data are publicly available.

Funding

None.

References

1. Timeline: WHO's COVID-19 response. Geneva: World Health Organization; 2021. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline, accessed 11 October 2021.
2. Regional COVID-19 dashboard. Manila: WHO Regional Office for the Western Pacific; 2021. Available from: https://who.maps.arcgis.com/apps/opsdashboard/index.html#/345dfdc82b5c4f6a815f1d54a05d18ec, accessed 11 October 2021.
3. Kaczorowski J, Del Grande C. Beyond the tip of the iceberg: direct and indirect effects of COVID-19. Lancet Digit Health. 2021;3(4):e205–6. doi:10.1016/S2589-7500(21)00024-8 pmid:33612431
4. Vital Strategies, World Health Organization. Revealing the toll of COVID-19: a technical package for rapid mortality surveillance and epidemic response. New York: Vital Strategies; 2020. Available from: https://preventepidemics.org/wp-content/uploads/2020/05/RMS_Report.pdf, accessed 8 April 2021.
5. Challenges in accurate medical certification of cause of death. Melbourne: CRVS Knowledge Gateway; 2018. Available from: https://crvsgateway.info/Challenges-in-accurate-medical-certification-of-cause-of-death~633, accessed 8 April 2021.
6. Coverage of birth and death registration. New York: United Nations, Department of Economic and Social Affairs, Statistics Division; 2021. Available from: https://unstats.un.org/unsd/demographic-social/crvs/index.cshtml, accessed 30 November 2021.
7. Hsu JC. Multiple comparisons: theory and methods. London: Chapman & Hall: CRC Press; 1996.
8. Humanitarian health action definitions: emergencies. Geneva: World Health Organization; 2020. Available from: https://www.who.int/hac/about/definitions/en, accessed 8 April 2021.
9. WHO COVID-19 excess mortality estimation methodology. Geneva: World Health Organization; 2021. Available from: https://www.who.int/publications/i/item/who-covid-19-excess-mortality-estimation-methodology, accessed 11 October 2021.
10. Weinberger DM, Chen J, Cohen T, Crawford FW, Mostashari F, Olson D, et al. Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. JAMA Intern Med. 2020;180(10):1336–44. doi:10.1001/jamainternmed.2020.3391 pmid:32609310
11. WPRO all-cause-of-mortality and excess death calculator. Manila: WHO Regional Office for the Western Pacific; 2021. Available from: https://worldhealthorg.shinyapps.io/WPRO-all-cause-of-mortality-and-excess-death-calculator/, accessed 20 April 2021.
12. Pulse survey on continuity of essential health services during the COVID-19 pandemic: interim report, 27 August 2020. Geneva: World Health Organization; 2020. Available from: https://apps.who.int/iris/handle/10665/334048, accessed 16 April 2021.
13. Handcock MS, Blackburn B. Handcock/WPROACM: Methodology and code used in the WHO WPRO ACM calculator (v1.0.0). Geneva: Zenodo; 2021. doi:10.5281/zenodo.5746071

APPENDIX

This appendix provides technical details on the statistical methods used to estimate excess mortality during the coronavirus disease (COVID-19) pandemic using data on all-cause mortality (ACM). It also provides a simulation study to assess the validity of the methodology. The exact code used in the analyses in this paper is in a static archive.[1]

We consider the case where we have multiple time-series of ACM counts from each Member State for each week between 1 January 2015 and a recent date. For some Member States we have only monthly data; for such cases, the description below is also suitable. We consider the case where we have separate reported counts for each sex and age group (typically, 5-year age groups).

The primary objective is to estimate the expected ACM counts for each week from 1 January 2020 onward assuming no pandemic had occurred. The excess mortality is defined as the difference between the reported counts and expected counts for each week.

Model

To illustrate, let us consider the case of females aged 65–74 years in Australia. Let y_t be the count for week t=1,…,T with t=1,…,260 being the period 1 January 2015 to 31 December 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 using 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 ACM and $trend(t)$ is the curvilinear trend of ACM over time. The annual cycle $c(t)$ is modelled as a cyclic cubic spline function[2] of time with a period of 52 weeks (i.e. $c(t) = c(t + 52)$), where 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 probably overfit the observed data, meaning it may not predict well using new data. Splines allow many low-degree (in this case, degree three) polynomials to fit the data in pieces, achieving a good fit to the data without the risk of overfitting.

Specifically, c_t is modelled as a piecewise cubic polynomial that has a continuous second derivative, is continuous, has continuous first and second derivatives at 52-week cycles and best fits the recorded ACM while being smooth. The specific criterion for the last feature is to choose c_t to minimize the penalized square error (PSE):

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

where $c''[s]$ is the second 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 ACM. The traditional estimator, $c[s]$, is the minimizer with τ=0; that is, there is no penalty for lack of smoothness. The choice of τ is subjective. In this work we chose to maximize the ability to predict unrecorded ACM counts. Specifically, we used generalized cross validation (GCV)[3] to choose, and the R package 'mgcv' (created by Simon Wood) for analysis.[4,5] The annual cycle obtained in this way is the optimal smoothest annual cycle chosen to maximize the likelihood of the observed ACM.

A similar approach is taken to the curvilinear trend $trend(t)$. It is modelled 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 ACM while being smooth. The specific criterion for the last feature is to choose $trend(t)$ to minimize the PSE:

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

where $trend''[t]$ is the second 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 ACM. The traditional estimator, $trend(t)$, is the minimizer with $\gamma=0$; that is, there is no penalty for lack of smoothness. Like τ, the choice of γ is subjective. Also, as with the annual cycle, we chose to maximize the ability to predict unrecorded ACM counts by using the GCV criterion. The model allows for arbitrary time-varying covariates, X_t. Including both the date and period allows for the model to detect trends both across and within years.

Negative-binomial regression is a natural choice given 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; also, it can account for instances of low or zero counts without issue.

This particular negative-binomial regression model is a generalized additive model (GAM) that 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. 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).

Currently, this model is simple in that it uses only information on sex, age group and time/date. When more data become readily available (e.g. influenza counts), the model can easily be extended to incorporate that data. 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. Hence, this preliminary approach should serve as a strong starting point.

The next step is to stochastically forecast the expected to represent the uncertainty in the estimate of the expected. Thus, the statistical significance of the observed can be determined (i.e. if it represents 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 mortality in addition to the sampling variation of the counts for given model parameters. We prefer this to a formal Bayesian model owing to its simplicity.

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

For Member States with missing (pandemic) weeks, we can stochastically interpolate using simple time-series models. If the number of missing weeks is significantly high, we use a negative-binomial model such as the one described above to stochastically interpolate.

An issue that may be important to adjust for is reporting delay (this is mainly an issue for recent weeks). To do this, information is needed on the reporting delay. In the United States, the National Center for Health Statistics reports mortality as the serial provisional data from the states are received and processed – counts of deaths from recent weeks are highly incomplete, reflecting delays in reporting. These "provisional" counts are updated regularly over the following weeks, and the counts are not finalized until more than a year later. The estimate of completeness is based on the number of weeks that have passed between when the death occurred and when the data set was obtained. We can model this relationship and use it to adjust the estimates, if necessary.

Validation of the statistical method for estimating ACM without a pandemic

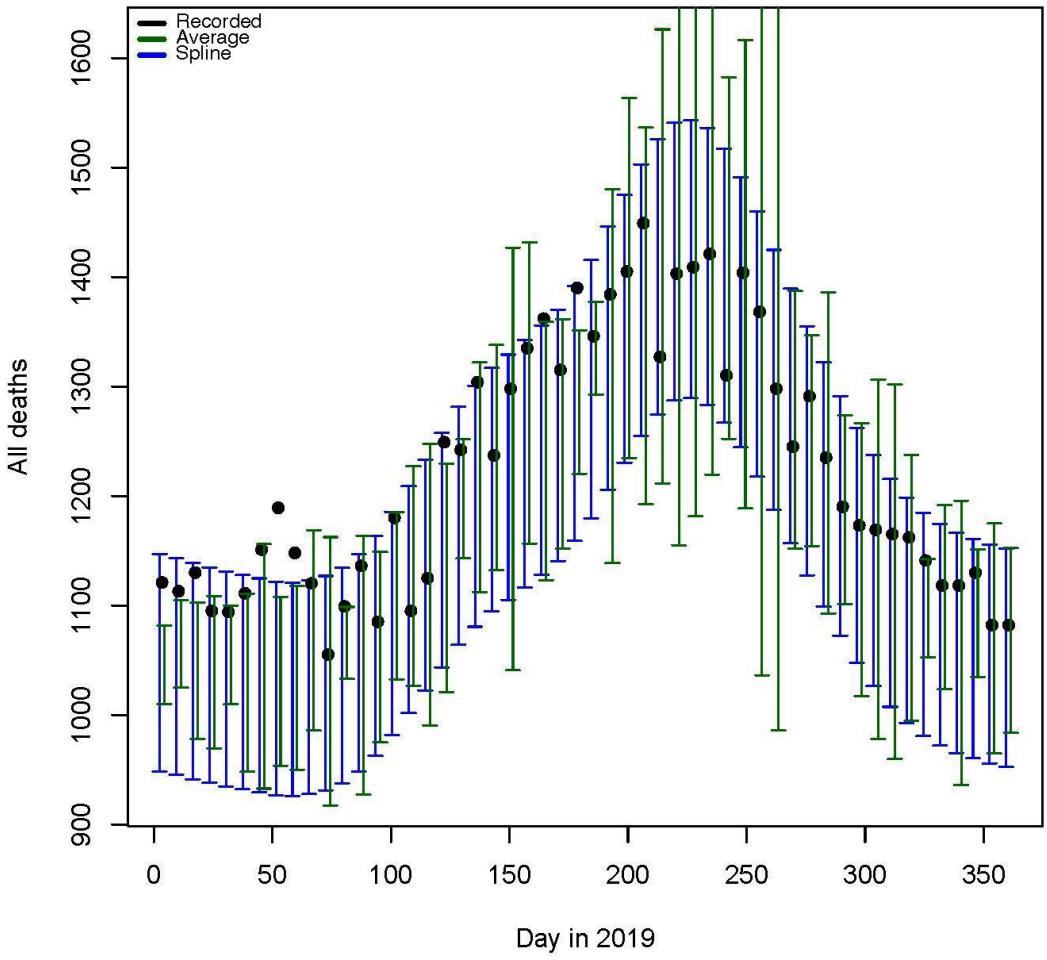
One may ask why it is not better to simply compare the observed ACM counts to historical averages of recent years. As we will show, doing so offers less robust pre-

diction intervals than using the model described above. The following validation metrics also justify using this model to gauge the significance of current ACM counts relative to pre-pandemic times.

The model attempts to forecast ACM 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 that the model makes accurate predictions. One way to validate the accuracy of the model is to use it to predict ACM during 2019, a year in which there would have been no "excess" mortality. The model is trained using data from 1 January 2015 through to 31 December 2018, then predictions are made on a weekly or 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, sexes and Member States, but to continue with the example used earlier (i.e. of females aged 65–74 years in Australia), we present those results for that example. **Appendix Fig. 1** below shows the 95% prediction intervals for the model ("spline") and 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 **Appendix Fig. 1**, the lengths of the spline intervals are typically smaller than the lengths of the weekly average intervals, meaning that the spline

Appendix Fig. 1. **Prediction intervals for 2019 based on deaths in 2015–2018. The black dots are the reported deaths for each week in 2019. The green error bars are based on the weekly averages. The blue intervals are 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.**



Appendix Table 1. **Prediction interval accuracy for all age and sex groups. The intervals produced by the spline model have the correct coverage whereas those produced by the weekly average model are well below their nominal coverage.**

Sex and 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	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	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	81	95
Overall total	83	91
Median %	84	92
Mean %	84	92

ETS: exponential triple smoothing; PI: prediction interval.

model has higher accuracy because it is a better model rather than just because it is larger. More importantly, the weekly average intervals are misleading and their actual coverage is far below their nominal coverage.

The accuracy of the spline model is not solely for females aged 65–74. **Appendix Table 1** shows per cent 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 sex and age groups.

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; however, intervals that are too long create too much uncertainty to be worthwhile. **Appendix Table 2** shows the lengths of the prediction intervals for the spline, exponential triple smoothing (ETS) and weekly average. The spline intervals tend to be nearly the same length as those of the ETS for those aged 0–74 (the weekly average has a short length but is highly inaccurate). It is in those aged 75+ (and when aggregating across all age groups) that the spline intervals are longer than their counterparts. The significant increase in the uncertainty surrounding the older age categories is something that will be investigated.

Appendix Table 2. **Prediction interval length for all Australian age and sex groups.**

Member state: Sex and age group	Average from 2015 to 2018 (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	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	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	111.9	135	228
AUS: Overall total	191.0	202	386
Median length	41.5	74	85
Mean length	56.1	83	115

AUS: Australia; ETS: exponential triple smoothing; PI: prediction interval.

References

1. Handcock MS, Blackburn B. Handcock/WPROACM: Methodology and code used in the WHO WPRO ACM Calculator (v1.0.0). Geneva: Zenodo; 2021. doi:10.5281/zenodo.5746071
2. Wegman EJ, Wright IW. Splines in statistics. J Am Stat Assoc. 1983;78(382):351–65. doi:10.2307/2288640
3. Craven P, Wahba G. Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math. 1979;31:377–403.
4. Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc. 2004;99(467):673–86. doi:10.1198/016214504000000980
5. Wood SN. Generalized additive models: an introduction with R. 2nd ed. Boca Raton: Chapman and Hall/CRC Press; 2017.