International Journal of Epidemiology, 2024, **53(6)**, dyae145
https://doi.org/10.1093/ije/dyae145
Original article

IEA International Epidemiological Association
OXFORD

Original article

Development of a registration interval correction model for enhancing excess all-cause mortality surveillance during the COVID-19 pandemic

Anna A Sordo,[1,2,*] Anna A Do,[1] Melissa J Irwin[1] and David J Muscatello [3]

[1]Rapid Surveillance, Centre for Epidemiology and Evidence, NSW Ministry of Health, St Leonards, NSW, Australia, [2]NSW Biostatistics Training Program, NSW Ministry of Health, Centre for Epidemiology and Evidence, St Leonards, NSW, Australia and [3]University of New South Wales, School of Population Health, Kensington, NSW, Australia

*Corresponding author. Rapid Surveillance, Centre for Epidemiology and Evidence, NSW Ministry of Health, 1 Reserve Road, St Leonards, NSW 2065, Australia. E-mail: anna.sordo@health.nsw.gov.au

Abstract

Background: Estimates of excess deaths provide critical intelligence on the impact of population health threats including seasonal respiratory infections, pandemics and environmental hazards. Timely estimates of excess deaths can inform the response to COVID-19. However, access to timely mortality data is challenging due to the time interval between the death occurring and the date the death is registered and available for analysis ('registration interval').

Development: Using data from the New South Wales, Australia, Births Deaths and Marriages Registry, we developed a Poisson regression model that estimated near-complete weekly counts, for a given week of death, from partially-complete death registration counts. A 10-weeks lag was considered, and a 2-year baseline of historical registration intervals was used to correct lag weeks.

Application: Validation of estimated counts found that the root-mean-square error (as a percentage of mean observed near-complete registrations) was less than 7% for lag week 3, and <5% for lag weeks 4–9. We incorporated this method utilizing an existing rapid weekly mortality surveillance system. Counts corrected for registration interval replaced observed values for the most recent weeks. Excess death estimates, based on corrected counts, were within 1.2% of near-complete counts available 9 weeks from the end of the analysis period.

Conclusions: This study demonstrates a method for estimating recent death counts to correct for registration intervals. Estimates obtained at a 3-week lag were acceptable, while those at greater than 3 weeks were optimal.

Keywords: Mortality, surveillance, delay adjustment, delay correction, time lag, excess deaths, pandemic, COVID-19, SARS-COV-2.

Key Messages

- Registration interval correction of all-cause deaths can provide accurate, more timely assessment of pandemic-related mortality.
- Estimates obtained with a 3-week death registration interval were acceptable, while those at greater than 3 weeks were optimal.
- This method provides timely intelligence on excess mortality, making it an important tool for meeting the policy demands of a pandemic response.

Background

Estimating excess mortality is an internationally recognized approach for understanding the impact of the coronavirus disease 2019 (COVID-19) pandemic.[1–3] Excess mortality describes the increase in all-cause mortality associated with the occurrence of an epidemic. It is calculated by the difference between the observed and expected number of deaths during the epidemic period. Because epidemic-attributable deaths are often under-ascertained, approaches to estimating excess mortality have long been used to estimate the population burden of seasonal and pandemic influenza, and heatwaves.[4–6]

Since the COVID-19 pandemic began, many jurisdictions have used laboratory confirmed or suspected COVID-19 deaths to monitor and track recent disease activity and severity.[7,8] While knowing the number of COVID-19 deaths is important, this may only partially reflect the impact on population mortality.[9] Estimating excess mortality is a more comprehensive measure, because it captures deaths indirectly related, or not recognised as related, to COVID-19. A feature of excess mortality methods is that, in temperate countries, deaths are seasonal in the absence of influenza epidemics. Statistical models can be used to estimate the underlying seasonal pattern of expected deaths. They can be compared to observed death counts or population rates to estimate excess mortality.[10]

The Ministry of Health in New South Wales (NSW), Australia, has conducted rapid mortality surveillance since

Received: 16 July 2023. **Editorial Decision:** 16 September 2024. **Accepted:** 24 October 2024

2008, using methods similar to those of the Centers for Disease Control and Prevention.[11] The current surveillance approach developed in NSW uses a seasonal harmonic, robust regression to estimate the expected population rate of deaths.[5,10] Data are drawn from death registration information recorded by the state registry of births, deaths and marriages. The date of death is used to form the time series for excess mortality analysis. The NSW rapid mortality surveillance system is automated, with reports refreshed weekly with the latest death registration information.

A challenge of rapid mortality surveillance in monitoring death trends is the under-reported of recent death counts, primarily due to the time interval between the death and its registration. We call this the 'registration interval'. For most deaths in NSW, funeral directors are required under the NSW Births, Deaths and Marriages Registration Act, 1995 to register the fact of death within 7 days of body disposal.[12] As the date of burial or cremation is often determined by the deceased person's family, variation in counts can be influenced by family requirements for timing of the funeral for religious or other reasons and by the registration delivery method; electronic or postal.

To account for under-reporting, NSW all-cause rapid mortality surveillance discarded deaths occurring in the most recent 6-week period. However, the demand for more timely situational intelligence during the COVID-19 pandemic led to registration interval correction being explored. This methods article describes the development and application of a method for registration interval correction to improve the timeliness and short-term accuracy of an existing rapid mortality system in NSW. The existing system aligned, where possible, with the approach taken by the Australian Bureau of Statistics to enable national comparisons.

Development and validation of the registration interval correction

Our method for estimating near-complete weekly all-cause death counts to correct for registration intervals was adapted from an approach implemented by EuroMOMO.[13,14] This involved modifying the EuroMOMO registration interval correction model to improve its performance in NSW. We assessed the model's performance by applying it to historical registration data, and calculating differences ('errors') between final registration and predicted counts at varying reporting lags in weeks.

Data source

Data from the Death Registrations Unit Record File held by NSW Health, was sourced daily from NSW Registry of Birth Deaths and Marriages Lifelink system. This register includes all doctor and coroner certified deaths from any cause registered in NSW including deaths of non-residents and overseas visitors.

Correcting for registration interval

A preliminary analysis confirmed that approximately 99% of deaths would be registered within a 10-week interval (Supplementary Figure S1). Guided by the EuroMOMO Poisson registration interval correction model,[13] we developed a Poisson generalised linear model that predicted near-complete weekly counts from partially-complete counts available in the most recent weeks prior to the surveillance reporting week. The registration interval was considered for "lag week" (k), with $k = 0, \dots, 9$. In this context, $k = 0$ includes deaths registered in the same week of death, and $k = 9$ being a death registered 9 weeks after death.

Like the EuroMOMO algorithm,[14] the model uses a dataset of observations with 10 rows per week ending date of death (for $k = 0, \dots, 9$) with each row relating to counts available at each lag week. Near-complete registrations were defined as the count available at $k = 10$. The dataset contained the following variables: week ending date of death, lag week (k), month of year of death, cumulative number of death registrations from week 0, …, k, and near-complete registrations, that is, the count at $k = 10$.

To estimate the corrections for each weekly surveillance reporting period, the dataset included a 2-year modelling reference period of historical data (104 weeks), that ended 10 weeks prior to the latest week for which near-complete counts were to be estimated. This period was chosen because surveillance is conducted weekly and it ensures enough coverage of usual but recent registration behaviour. For example, if N represents the week ending the most recent Sunday (surveillance date), then the model would be fit to a reference period from weeks N-114 to N-10 and predictions would be made for weeks N-9, N-8, … . N. The variables week ending date of death was included to account for linear (secular) trend and month of year of death for seasonality in registration interval. See Supplementary Tables S1 and S2 for sample datasets.

While the EuroMOMO model[14] included a variable for registry closure days, such as public holidays and weekends, we excluded this because a public holiday variable was not found to improve the model and our state registry adjusts weekend staff resourcing to meet registration demand. Our model also considers an interaction term between lag week and cumulative number of death registrations, where 'cumulative' refers to the number of registrations accumulated up to a given lag week k. Additionally, we accounted for month of year of death and an interaction between lag week and month of year of death, to minimise estimation error. The Poisson equation for this analysis is shown below:

$$\begin{aligned}\ln(\mathit{near-complete\ registrations}) = {} & \beta_0 + \beta_1(k) \\ & + \beta_2(\mathit{cumulative\ reg.}) \\ & + \beta_3(\mathit{week\ ending\ date\ of\ death}) \\ & + \beta_4(\mathit{month\ of\ year\ of\ death}) \\ & + \beta_5(\mathit{cumulative\ reg.} * k) \\ & + \beta_5(\mathit{month\ of\ year\ of\ death} * k)\end{aligned}$$

To validate the count estimation, datasets and models were generated to retrospectively mimic weekly death reporting from week ending 1 January 2017 to 10 July 2022. Daily death counts were aggregated into surveillance reporting weeks ending Sunday. For each mimicked reporting week, the model was automatically refreshed for each new week N. Each model was fit to registrations recorded in a 2-year reference dataset for week numbers N-114 to N-10, with weeks N-9 to N being used in model fitting and registration count correction. Model fit and registration interval correction accuracy was assessed using root-mean-square error (RMSE). The initial model showed unacceptably large errors when including week numbers N-2, N-1 and N in the model estimation (Supplementary Table S3). In a sensitivity analysis, we found that a model that excluded weeks N-2, … , N provided a lower RMSE in estimated near-complete counts than the initial model in all lag weeks except lag week 3. Consequently, weeks N-2, … , N were excluded from the modelling, and predictions were only made for weeks N-3 to N-9. Our validation demonstrates that the largest overall error was seen in lag week 3 (6.7%) and there was less than 5% error for lag weeks 4 or more, improving at each subsequent lag week (Supplementary Table S4).

Further analysis which included registration interval differences by year, age and month of year of death found considerable variability in the proportion of registrations received in those relative weeks, but only minor variation was found by sex (Supplementary Figures S2–S4). We also assessed the impact of coronial deaths, which are typically registered later, and determined no significant impact on the model. We conducted a sensitivity analysis across different modelling periods and found that a 2-year period yielded the lowest percentage RMSE for the most recent lag weeks. Additionally, a 17-week pilot test comparing lag-adjusted counts to more complete counts available as of 31 January 2023, revealed the greatest percentage differences in lag weeks 3 (13.7%) and lag week 4 (5.2%) (Supplementary Figure S9). For further details on validation of the registration interval correction model (Supplementary Section S4).

Application to weekly rapid mortality surveillance and estimation of excess deaths

The current rapid mortality surveillance approach uses a robust linear regression model[5] to estimate the expected seasonal baseline of weekly all-cause death rates that would occur in the absence of seasonal influenza and other causes of substantial extra-seasonal mortality. The general methodology for background estimation is described in Muscatello et al,[5] see Supplementary Section S5 for a description of the current implementation.

Impact of registration interval correction on estimation of excess deaths

We compared excess deaths from corrected and uncorrected counts for the current surveillance application. Excess all-cause deaths were estimated by calculating the accumulated sum of the weekly differences between the observed and expected weekly rates per 100 000 population.[15] For the registration interval corrected application, the weekly rates incorporating estimated near-complete counts were used. Excess weekly rates were converted to a count by multiplying the weekly rate by the estimated resident population and dividing by 100 000.

Figure 1 demonstrated the result of registration interval correction for surveillance week ending 6 November 2022 for surveillance analysis conducted on 27 November 2022. The red line shows modelled estimates of the seasonal baseline of weekly all-cause death rates. The blue line shows weekly observed rates, while the green line shows corrected values from 25 September 2022 (lag week 9) to 6 November 2022 (lag week 3). Excess deaths during winter 2017 were attributed to a severe influenza season.[16] In 2020, lower mortality could be attributed to the use of non-pharmaceutical public health measures that effectively controlled SARS-CoV-2 transmission and substantially reduced circulation of other respiratory viruses.[17] Periods of excess mortality from July in 2021 were associated with community transmission of the SARS-CoV-2 Delta variant, which was partially controlled using a combination of non-pharmaceutical public health measures and the rapid rollout of a COVID-19 vaccination program that achieved high population uptake. In late 2021, despite high vaccination coverage, the emergence of the Omicron variant, with its combination of immune escape and high transmissibility led to increased SARS-CoV-2 virus transmission.[18–20] Weekly excess deaths across the three years were, on average, lower in Australia than in most comparable countries.[21] The observed weekly death rates decline towards the end of the time series reflecting incomplete death registration data. The corrected weekly death rates are higher than the observed rates across all lag weeks. For example, at lag week 3, the observed weekly death rate was 10.4 per 100 000 persons, compared with 12.3 per 100 000 persons in the lag-corrected model. Without the lag correction, the data observed at the time of analysis would only account for 84% of the deaths registered in that week.

To assess the effect of the registration interval correction on estimates of excess deaths, we re-calculated the observed number of deaths for the surveillance period 2 January to 6 November 2022, allowing additional weeks for registration

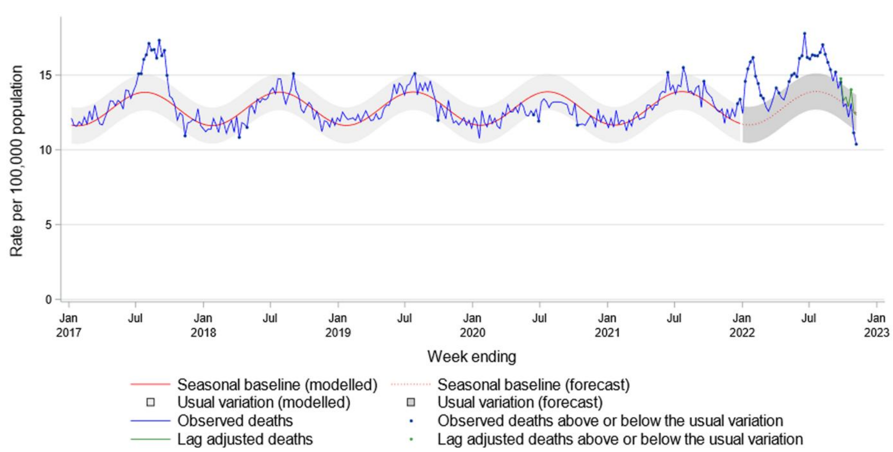

Figure 1. Corrected, observed all-cause weekly death rates, estimated and forecast seasonal baseline mortality, and limits of usual variation for one reporting week (data analysed week ending 27 November 2022), NSW, persons of all ages. Weeks included were 1 January 2017 to 6 November 2022. Black dots represent observed deaths outside the 95% confidence interval for background deaths in that week

Table 1. Percentage difference between corrected and observed excess death counts for deaths that occurred between week ending 2 January to 6 November 2022 and registered by 31 January 2023, all ages and by age group

Age group	Difference between corrected and observed excess counts(%)[a,b,c,d]
All ages	−1.2
< 65 years	−6.9
65–84 years	−2.9
85 Years and over	−0.1

[a] Comparing excess counts from deaths that occurred from week ending 2 January to 6 November 2022.
[b] Corrected excess counts were corrected between week ending 25 September to 6 November 2022 and registered by 27 November 2022.
[c] Recalculated observed excess counts are deaths registered by 31 January 2023.
[d] Excess from corrected counts is calculated from the mean.

to 31 Jan 2023. The rationale being that almost all deaths for this surveillance period would be registered by 31 January 2023 (Supplementary Figure S4). Table 1 shows the percentage difference between corrected excess counts from deaths registered by 27 November and observed excess counts for deaths registered, for the same period, by the 31 January 2023. For all-ages, corrected excess deaths counts were 1.2% lower than re-calculated observed excess counts, while differences ranged from 6.9% lower for those aged under 65 years to 2.9% lower for those aged 65–84 years. There was 0.1% difference between observed and corrected excess counts for persons aged 85 and over.

SAS Enterprise Guide 8.3[22] was used for all analysis. The SAS ROBUSTREG procedure with the 'M estimation' method was used for the seasonal baseline estimation.[23]

Discussion

This study demonstrates a method for estimating near-complete counts of death registrations from three weeks of date of death. This provided timely intelligence on excess mortality to meet the policy demands of a pandemic response. Our analysis shows that reasonably accurate interval-correction is possible with only limited registration data available in the most recent weeks. Although variation in the registration interval in the current and previous 2 weeks appears to prevent successful estimation for those weeks, reasonable estimates can still be made from 3 to 9 weeks before the analysis date. In the past, we excluded data six weeks before the analysis week and the earlier 3 weeks were used without correction.

By observing model performance over time, we showed that near-complete registrations could be predicted as few as 3 weeks from the date of analysis, with improvement in each subsequent lag week. The percentage of RMSE of estimated compared with observed near-complete death registrations showed that overall, the model performed well, particularly in lag weeks 4–9 (<5% error). We found that excess deaths estimated from near-complete registrations were 2% lower in all-ages and less than 3% lower for those aged over 65 years. Our finding, based on RMSE, showed that a shorter period of historical data provided better estimation of near-complete registrations than did a longer period. This finding could reflect changes over time in clerical procedures or information systems used for registrations. Further, a 17-week pilot to test for current surveillance validity showed similar results, with the highest percentage difference between corrected and observed counts seen at lag week 3, but mostly minor differences at subsequent lag weeks. Each week, the application of our model continues to deliver valuable information to policy decision makers by providing them with trends in mortality for the state.

The strength of this methodology is that it may be adaptable for use in other similar sized or larger jurisdictions that have access to regular death registration feeds. The minimum requirement is a weekly death feed with a relatively stable registration interval distribution. Different jurisdictions will need to assess what registration interval is appropriate to model their near-complete deaths. For NSW, 99% of registrations were complete by 10 weeks, however this will vary for each jurisdiction, and a longer registration interval may require more lag weeks to achieve near-complete estimates. Further, the registration correction method presented in this paper was useful over the winter period, particularly at the beginning, when attempting to understand the combined impact of COVID-19, influenza and other respiratory viruses. It became clear, that NSW was unlikely to have a short sharp winter mortality season in 2022, but rather, a prolonged season as the interval-corrected death rates at the time were suggesting.

This approach has some limitations. First, this method relies on a stable registration interval distribution for good predictions of near-complete registrations in each lag week. Variation in the registration interval for deaths registered in NSW has improved over time; however, this may not be the case for other jurisdictions, where a more severe epidemic may have resulted in longer registration delays. Secondly, the pilot suggests that in the later part of 2022, which has seen sustained high mortality unlike previous years, that lag corrections for week 3 are conservative, producing smaller estimates of excess for those weeks. However, for public health decision making, a conservative estimate in lag week 3, which is subsequently corrected for in the following week, is unlikely to delay critical public health action. Thirdly, the seasonal baseline is based on cyclical robust regression estimates and in some years, these may not appear to closely follow background death rates. In retrospective applications, alternative modelling strategies may provide a better fit. Live, continuous surveillance strategies require models that are simple, comparable across jurisdictions and that can provide a plausible one-year-ahead baseline forecast when prospectively applied. Lastly, excess death calculations only consider uncertainty in the estimate of the seasonal baseline and do not consider uncertainty around the estimated near-complete counts. Nevertheless, the RMSE analysis showed that the point estimates of near-complete counts were quite accurate.

Timely excess deaths information is crucial for population health risk assessment. The methods we have demonstrated here may be useful in international settings where death registration operates in a similar way and where death registration statistics can be made available to population health protection analysts. Registration interval correction of all-cause deaths can provide more accurate assessment of all-hazard related mortality, offering policy makers valuable insights to respond promptly to public health challenges.

Ethics approval

This work was conducted to enhance an existing surveillance system and did not require ethics approval. The NSW *Public*

Health Act 2010[24] provided the legislative basis for data use and disclosure. This work was conducted in accordance with the National Health and Medical Research Council (NHMRC) ethical principles.

Data availability

The unit record data in this study were used under NSW government legislation and is not approved for public release. However, an aggregate dataset to support the lag adjustment analysis has been included in the supplementary material.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

A.A.S., A.A.D., D.J.M designed the analytical strategy. M.J. I., D.J.M guided interpretation and provided subject matter expertise. A.A.S., A.A.D., ran the analysis and visualization, and A.A.S. wrote the original draft. All authors reviewed and edited the draft manuscript and approved the final version as submitted. D.J.M provided supervision for this work. D.J.M and is the guarantor for the paper.

Use of artificial intelligence (AI) tools

None.

Funding

David Muscatello is supported by an NHMRC Investigator Grant (APP1194109).

Acknowledgements

The contents of the published material are solely the responsibility of the Administering Institution, a Participating Institution or individual authors and do not reflect the views of the NHMRC. We would also like to acknowledge Dr Jens Nielsen on behalf of the EuroMOMO network for helpful advice on the implementation of the EuroMOMO registration interval correction procedure. Lastly, we like to acknowledge our colleagues within the Centre for Epidemiology and Evidence at the NSW Ministry of Health for their advice and peer review.

Conflict of interest

None declared.

References

1. Wang H, Paulson KR, Pease SA *et al.* Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* 2022;**399**:1513–36.
2. Msemburi W, Karlinsky A, Knutson V, Aleshin-Guendel S, Chatterji S, Wakefield J. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* 2023;**613**:130–7.
3. Barnard S, Chiavenna C, Fox S *et al.* Methods for modelling excess mortality across England during the COVID-19 pandemic. *Stat Methods Med Res* 2022;**31**:1790–802.
4. Coates L, van Leeuwen J, Browning S, Gissing A, Bratchell J, Avci A. Heatwave Fatalities in Australia, 2001–2018: an analysis of coronial records. *Int J Disaster Risk Reduct* 2022;**67**:102671.
5. Muscatello DJ, Morton PM, Evans I, Gilmour R. Prospective surveillance of excess mortality due to influenza in New South Wales: feasibility and statistical approach. *Commun Dis Intell Q Rep* 2008;**32**:435–42.
6. Zhao Q, Guo Y, Ye T *et al.* Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. *Lancet Planet Health* 2021;**5**:e415-25–e425.
7. Mathieu E, Richie H, Rodés-Guirao L *et al. Coronavirus Pandemic (COVID-19)*. OurWorldInData.org; 2020. https://ourworldindata.org/coronavirus (9 March 2023, date last accessed).
8. NSW Government. Health Protection NSW. *COVID-19 (Coronavirus)*. https://www.health.nsw.gov.au/infectious/covid-19/pages/default.aspx (8 October 2024, date last accessed).
9. Karlinsky A, Kobak D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife* 2021;**10**:e69336.
10. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports* 1963;**78**:494.
11. Centers for Disease Control and Prevention. *Excess Deaths Associated with COVID-19* [Internet]. National Center for Health Statistics, 2023. https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm#references (9 March 2023, date last accessed).
12. New South Wales Government. *Births, Deaths and Marriages Registration Act 1995 No 62* [Legislation]. Sydney: New South Wales Government, NSW Legislation, 2010. https://legislation.nsw.gov.au/view/html/inforce/current/act-1995-062 (3 June 2024, date last accessed).
13. EuroMOMO. *Methods*. 2022. https://www.euromomo.eu/how-it-works/methods/ (24 March 2024, date last accessed).
14. Virtanen MJ, Espenhain L, Nielsen J *et al. thl-mjv/euromomo: EuroMoMo algorithm*. 2019. https://rdrr.io/github/thl-mjv/euromomo/f/README.md (24 March 2024, date last accessed).
15. Department of Planning and Environment. *Projections*. 2022. https://www.planning.nsw.gov.au/research-and-demography/population-projections/projections (10 October 2024, date last accessed).
16. Muscatello DJ, Nazareno AL, Turner RM, Newall AT. Influenza-associated mortality in Australia, 2010 through 2019: High modelled estimates in 2017. *Vaccine* 2021;**39**:7578–83.
17. AIHW 2021. *The First Year of COVID-19 in Australia: Direct and Indirect Health Effects*. Cat. no. PHE 287. Canberra: AIHW. https://www.aihw.gov.au/getmedia/a69ee08a-857f-412b-b617-a29acb66a475/aihw-phe-287.pdf
18. Gregory G, Zhu L, Hayen A, Bell KJ. Learning from the pandemic: mortality trends and seasonality of deaths in Australia in 2020. *Int J Epidemiol*. 2022;**51**:718–26.
19. Lin L, Demirhan H, P. Johnstone-Robertson S, Lal R, M. Trauer J, Stone L. Assessing the mass vaccination campaigns in australia and their impact over the delta and omicron outbreaks. *PLoS ONE* 2024;**19**:e0299844.
20. Australian Government. Department of Health and Aged Care. *AHPPC Statement on the Omicron Public Health Implications and Response Options*. 2021. https://www.health.gov.au/news/ahppc-statement-on-the-omicron-public-health-implications-and-response-options (11 December 2023, date last accessed).
21. OECD. Mortality, by week: Excess deaths by week, 2020–2023. https://stats.oecd.org/index.aspx?queryid=104676 (1 December 2023, date last accessed)
22. SAS Institute Inc. *SAS® Enterprise Guide® 8.3: User's Guide*. Cary, NC: SAS Institute Inc, 2020.
23. The ROBUSTREG procedure. *SAS/STAT(R) 9.2 User's Guide*. Cary: SAS Institute, 2008. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.6381&rep=rep1&type=pdf
24. New South Wales Government. *Public Health Act 2010*. Sydney. New South Wales Government, NSW Legislation, 2010. https://legislation.nsw.gov.au/view/html/inforce/current/act-2010-127 (19 July 2021, date last accessed).

International Journal of Epidemiology, 2024, 53, 1–5
https://doi.org/10.1093/ije/dyae145
Original article