Published online: 15 November 2023

The Dangers of Drawing Cohort Profiles From Period Data: A Research Note

Alyson A. van Raalte, Ugofilippo Basellini, Carlo Giovanni Camarda, Marília R. Nepomuceno, and Mikko Myrskylä

ABSTRACT Drawing cohort profiles and cohort forecasts from grids of age–period data is common practice in demography. In this research note, we (1) show how demographic measures artificially fluctuate when calculated from the diagonals of age–period rates because of timing and cohort-size bias, (2) estimate the magnitude of these biases, and (3) illustrate how prediction intervals for cohort indicators of mortality may become implausible when drawn from Lee–Carter methods and age–period grids. These biases are surprisingly large, even when the cohort profiles are created from single-age, single-year period data. The danger is that we overinterpret deviations from expected trends that were induced by our own data manipulation.

KEYWORDS Lexis diagram • Cohort fertility • Cohort life expectancy • Lee–Carter method

Introduction

As demographers, we are all aware that demographic events and person-years at risk can be split and aggregated into different formats on the Lexis diagram: age–period squares, cohort parallelograms, or, for even more flexibility, Lexis triangles (Carstensen 2007; Keiding 1990; Wilmoth et al. 2021). But who among us has never created a cohort profile by "taking the diagonals" of an age–period grid? Perhaps this is because demographic data are usually preformatted for period analysis. Possibly we wanted to analyze cohort trends from a methodology that was designed specifically to be used on an age–period grid. Whatever the reason, taking the diagonals is common practice (e.g., Kermack et al. 1934; Myrskylä et al. 2013; Preston and Wang 2006; Shkolnikov et al. 2011; Vogt et al. 2017), even when we suspect that it is not the best practice.

In this research note, we illustrate the dangers of drawing cohort profiles from period data using examples that we have recently encountered. The first two examples describe how demographic measures artificially fluctuate when calculated from the diagonals of age–period rates because of (1) timing and (2) cohort-size biases. The third example illustrates how drawing forecast uncertainty from the upper and lower bounds of a period Lee–Carter time series parameter results in implausible prediction intervals when completing cohorts from period diagonals.
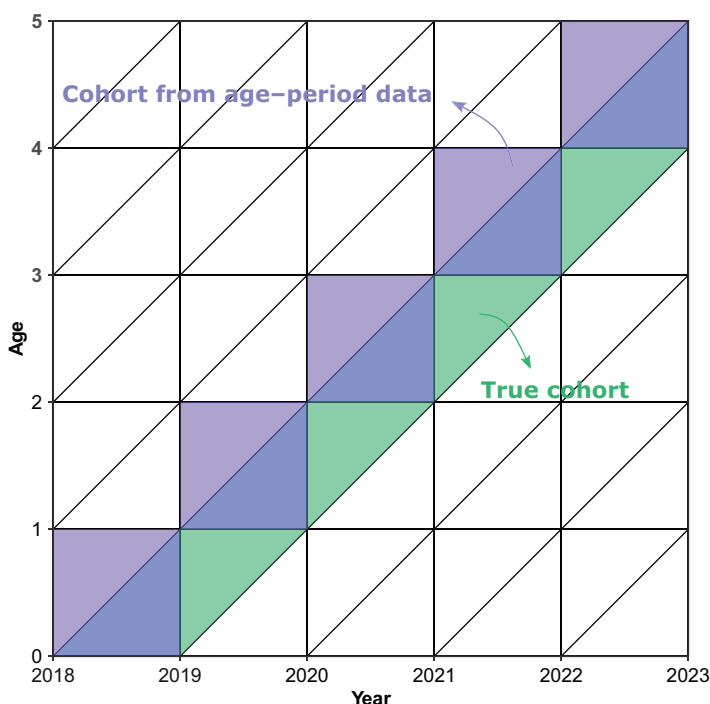
Fig. 1 Lexis representation of the difference between the two data structures. Properly structured age–cohort data are represented in green, and approximating cohorts by "taking the diagonals" of age–period data are represented in purple.

It is well known that the larger the age–period square, the weaker the approximation of a real cohort. As the age–period square gets smaller, cohort patterns drawn from such squares start to approximate real (continuous-time) cohort lines. All of our empirical examples use single-year, single-age data and illustrate that the dangers do not vanish with this fine grid. The examples are fully replicable (https://osf.io/xn5w7/) using data from the Human Mortality Database ([HMD] 2023) and the Human Fertility Database ([HFD] 2023).

Example 1: Cohort Summary Measures From Period Age-Specific Rates—Temporal Bias

Two key biases come into play when we calculate cohort summary measures from the diagonals of age–period data. The first is a bias from the timing of events and exposure, while the second relates to changing cohort sizes.

To illustrate the first of these biases, we turn to the Lexis diagram. In Figure 1, the true cohort data are represented by green parallelograms and "period diagonals" by purple squares. The data from the upper Lexis triangles will always be one calendar year earlier for the age–period compared with age–cohort data, with single-year, single-age data. Lengthening the time/age data format increases discrepancies in timing.
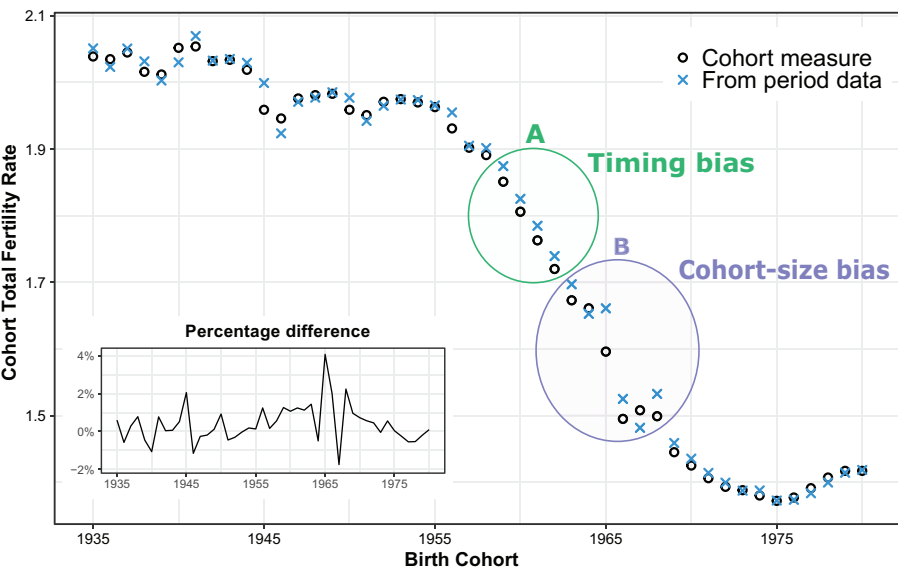
Fig. 2 Completed cohort fertility by age 40 in Japan derived from cohort data (circles) and from diagonals of period data (crosses) for cohorts 1935–1980. The inset panel shows the percentage differences between the two measures (cohort from period diagonal minus the true cohort).

In practice, the extent to which this temporal difference biases the cohort summary measure depends on how the measure is trending. If the processes happen to be stationary, there is no bias. In Figure 2, we contrast cohort completed fertility by age 40, also known as the cohort total fertility rate (TFR), for Japanese females calculated from the two data formats: age–cohort parallelograms and the diagonals of age–period squares. When fertility trends were flat over cohorts (e.g., cohorts 1945–1955), the artificial and true cohort TFRs line up well. When fertility declined sharply (e.g., cohorts 1955–1970), there was a general overestimation of the cohort TFR from the period squares, marked as "A" in Figure 2. The cohort TFR was underestimated from cohort 1975 onward as fertility recovered.

Example 2: Cohort Summary Measures From Period Age-Specific Rates—Cohort-Size Bias

The second bias results from changing cohort size. The trend in cohort TFR derived from the diagonals of age–period rates shows especially large fluctuations around the 1966 birth cohort (marked as "B" in Figure 2). This is only partially real. The true cohort trend is more stable. Changing cohort size explains the difference between the two trends. In 1965, there were 1.8 million births; births dropped to 1.4 million in 1966 (owing to beliefs related to being born in the year of the "Fire-Horse" (Kaku 1975)), only to rebound to 1.9 million in 1967. Deriving summary cohort fertility measures from the age–period squares uses rates drawn from events and exposures of the smaller 1966 cohort mixed with events and exposures from the larger neighboring

Table 1 Summary statistics of the joint impact of temporal and cohort-size biases on cohort total fertility rate and cohort life expectancy by taking the diagonals of an age–period grid

Measure	Average Deviation (% of absolute values)	Minimum Deviation (%)	Maximum Deviation (%)	Underestimated Values (%)	Overestimated Values (%)
Total Fertility Rate	0.6	−2.1	4.1	36	64
Life Expectancy	0.7	−4.8	2.7	80	20

Notes: "Underestimated" means that the cohort-from-period estimation was lower than the true cohort value. The fertility data come from 10 HFD countries with a long fertility series (Japan, France, Italy, the United States, Spain, Finland, Denmark, Sweden, the Netherlands, and England and Wales), while the mortality data are from seven HMD countries with a long mortality data series (the same countries listed above, excluding Japan, the United States, and Spain, since these countries do not have mortality data for full cohorts; moreover, for all countries we start the analysis from 1865—or later, depending on data availability—to use more reliable historical data).

cohorts. These interactions cause an artificial fluctuation in the corresponding estimated cohort trends.

How Large Is the Bias on Demographic Measures? Empirical Estimates From the HFD and HMD

The timing and cohort-size biases demonstrated with Japanese fertility data affect *any* cohort demographic measure estimated from the diagonals of age–period squares. In practice, these two biases are challenging to separate.

We estimated the joint impact of these biases on empirical trends of fertility (cohort TFR), as well as mortality (cohort life expectancy), from countries with long time series of available data from the Human Fertility and Mortality Databases (Table 1). Overall, on average there was a 0.6% (in absolute values) difference in cohort TFR between true cohort data and cohort-from-period diagonals data. Life expectancy differences amounted to 0.7% on average (in absolute values). Cohort-from-period fertility was underestimated about one third of the time and overestimated about two thirds of the time, while cohort-from-period life expectancy was usually underestimated. This is because although cohort fertility generally declined during the time frame investigated, there were also periods of recovery, while cohort life expectancy more consistently increased.

Example 3: Cohort Lifespan Variation Drawn From Period Lee–Carter Forecasts

Cohort mortality forecasts are often done by "taking the cohort diagonals" of a forecast age–period surface (Alburez-Gutierrez et al. 2021; Andreev and Vaupel 2006; Shkolnikov et al. 2011). Our third example shows that we need to be careful in attributing forecast uncertainty with such an approach.

The Lee–Carter (1992) model[1] and its various extensions have become the dominant mortality forecasting methodology, used worldwide by statistical agencies and international organizations (Basellini et al. 2023). The prediction intervals are derived from the mortality time series parameter k_t, usually either by:

1. simulating trajectories of k_t, constructing the life table measure of interest for each trajectory, and calculating the uncertainty from this simulated distribution; or
2. using the upper and lower bounds of the simulated k_t trajectories, and constructing the life table measure of interest that corresponds to these bounds.

The latter method is computationally simpler,[2] in particular since sometimes it is feasible to avoid the simulation altogether and calculate the bounds of k_t analytically. This approach is widely used, for instance, in the popular "demography" R package (Hyndman et al. 2023), and was suggested by Lee (2000:83) himself for period—not cohort—life table functions, since the variations in the log death rates will be "perfectly correlated with one another, because all are linear function of the same time-varying parameter *k*." For *cohort* life table measures, however, mortality schedules are completed from k_t, k_{t+1}, k_{t+2}, and so on, and the associated complex relationship between their uncertainty bounds (Lee 2000).

We illustrate this with a Lee–Carter forecast[3] of Swedish female mortality data, where cohort age-specific mortality was completed by employing the diagonals of a forecast age–period grid, using the 95% prediction intervals from the upper and lower bounds of the simulated k_t trajectories (Figure 3, panel c). While the life expectancy prediction intervals (panel a) appear reasonable, the interquartile range intervals (panel b) do not. The interval first grows, then shrinks, goes to zero, and then grows again. The decline in uncertainty over cohorts is clearly implausible. These dynamics occur regardless of the chosen index of variation, the input data (age–period or age–cohort), or the population (see Figures A1–A4 in the online appendix).
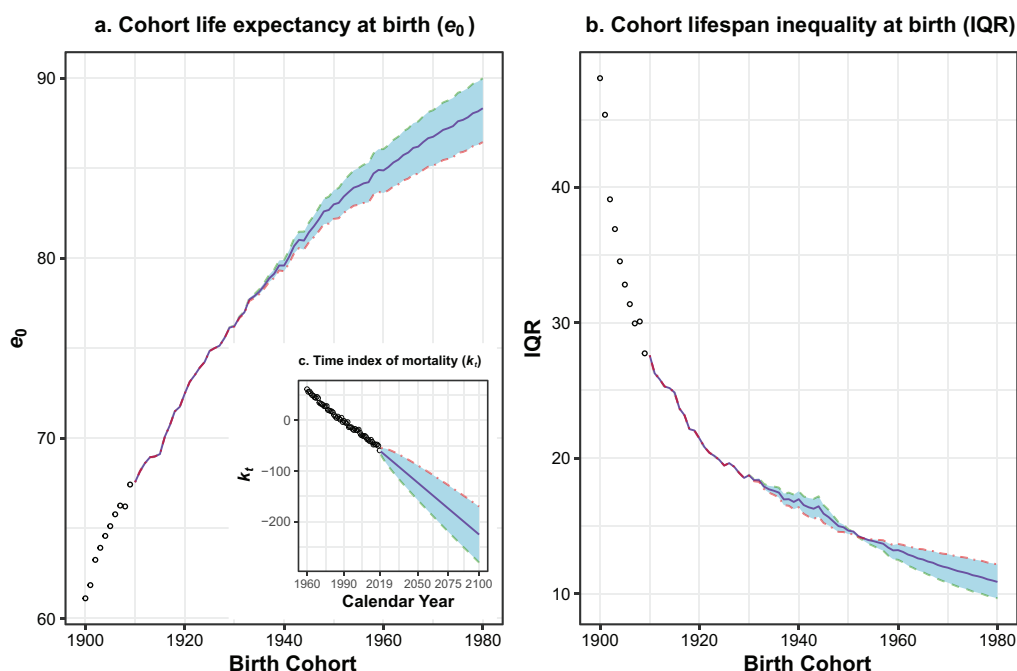
Discussion

Summary

Drawing cohort profiles from age–period data or models can lead to patently false conclusions. In the first instance, using age–period data to calculate cohort TFR and

---

[1] The Lee–Carter model expresses variation in death rates across age and time as follows: $\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}$, where a_x is the pattern of log mortality at age x, k_t is an index of the level of mortality at time t, b_x describes how much mortality at a given age changes with the overall mortality level changes, and $\varepsilon_{x,t}$ is the residual.

[2] This was far more of a concern in the early 1990s than it is today. With modern computing, the difference in the time needed to calculate prediction intervals from the two methods is trivial. We argue that the first approach should now be used even in period calculations.

[3] In our results, following the standard Lee–Carter estimation procedure, the model was fit to data from 1960 to 2019, a_x was set to equal the means over time of the $\ln m_{x,t}$, and the parameters b_x and k_t were estimated with singular value decomposition. An adjusted k_t (to match the observed number of deaths in each year) was forecast to decline linearly to 2100 using a random walk with drift.

a. Cohort life expectancy at birth (e_0)

b. Cohort lifespan inequality at birth (IQR)

c. Time index of mortality (k_t)

Fig. 3 Panel a: Observed (circles) and forecast (lines) cohort life expectancy at birth, e_0, for birth cohorts 1900–1980. Panel b: Observed (circles) and forecast (lines) cohort interquartile range, for birth cohorts 1900–1980. Panel c: Estimated k_t for the years 1960–2019 (circles) and forecast k_t with 95% prediction intervals for the years 2020–2100 (lines) for Swedish females, ages 0–110+. The colors highlight the median, upper, and lower percentile forecasts of k_t. IQR = interquartile range.

cohort e_0 resulted in estimates that were on average 0.6% and 0.7% different from those calculated from age–cohort data, respectively. The largest deviations were a 4.1% difference in cohort TFR and a –4.8% difference for cohort life expectancy at birth (e_0). In the second example, drawing cohort profiles from period Lee–Carter forecasts resulted in implausible prediction intervals for lifespan variability measures, when based on taking the upper and lower bounds of the k_t time series parameter.

Should the artificial fluctuations in cohort fertility and mortality caused by the input data be considered large or small? To put this difference into perspective, the average data-driven artifact estimated for any single cohort TFR amounts to about one third of the recent average yearly drop in U.S. period TFR over the period 2007–2019 (HFD 2023), which is noteworthy for its size and has generated considerable discussion (Guzzo and Hayford 2020; Hartnett and Gemmill 2020). In terms of mortality, the drop in period e_0 in either the United States or the United Kingdom from 2014 to 2017 was less than 0.1% (HMD 2023). In the pre-COVID-19 world, these life expectancy declines received extensive attention in the media, as well as in demographic and public health circles (Hiam et al. 2018; Ho and Hendi 2018). The declines were seven times less than the average data-induced cohort life expectancy artifact.

While the examples used related to fertility and mortality, these issues are equally pertinent in migration research. Migration flows fluctuate by year and season. In

Europe, 1.3 million refugees, primarily from Syria, Iraq, and Afghanistan, claimed asylum in 2015 (Eurostat 2016). This was more than double the level seen in any year over the preceding decade. Two thirds of these 2015 asylum seekers arrived in July or later that year. In a Lexis diagram with the year of arrival on the *x*-axis ("year") and years since arrival on the *y*-axis ("age"), the late-2015 migration cohort would have some weight in "year 2015–age 0," heavy weight in "year 2016–age 0," some weight in "year 2016–age 1," heavy weight in "year 2017–age 1," and so on with the fluctuation continuing as this migration cohort is followed throughout its life course. An analysis of decisions on asylum seekers by year of arrival, based on square-based cohort inference, could be very different from decisions based on properly structured age–cohort data.

As a field, demography is slowly moving away from the analysis of aggregate population-level data toward the usage of rich individual-level data (Lee 2001), which we can aggregate ourselves by year of birth. These changes started with the implementation of repeated cross-sectional and panel surveys a few decades ago (Crimmins 1993) and have progressed to include administrative and other big data sources (Kashyap 2021). The longer some of these studies (or registers) run, the more we can expect research designs that examine temporal change over the cohort dimension, as well as methodological development for proper cohort inference (see, e.g., Yang and Land (2013) for a discussion of how individual-level data are shaping methodological developments in the field of age–period–cohort analysis). Nevertheless, analysis of aggregate data remains a core task of demographers, and it is likely we will face aggregated age–period data from statistical offices for some time to come.

Potential Solutions to Mitigate the Problems

Calculating Cohort Summary Measures From Period Data

National statistical agencies often provide data in aggregated age–period formats that are particularly suited to period analysis. When conducting cohort analyses, we should first split the data into Lexis triangles and then reshape it into cohort parallelograms (Carstensen 2007). These methods have their assumptions, in particular, that births (or the event that defines the cohort) are evenly distributed at the beginning and end of the calendar year. This assumption is sometimes violated, for instance, after exceptional events such as wars, famines, or epidemics (Aassve et al. 2020; Agadjanian and Prata 2002; Chandra et al. 2018; Lindstrom and Berhanu 1999); in anticipation of loss or gain in social benefits such as baby bonuses (Brunner and Kuhn 2011; Thévenon and Gauthier 2011); depending on cultural beliefs, for example, the period boost in fertility among Chinese accompanying dragon years (Goodkind 1991, 1995); or due to seasonal fluctuations such as the 2015 European migration wave. In these cases, exposures can be distributed by month of birth or other event defining the cohort, as is now done for the exposure populations in the HMD and HFD (Wilmoth et al. 2021).

If the data we have on hand come in the form of age–period rates and not the actual event and exposure counts, we envisage an adjustment of the rates that depends on the changes in cohort sizes, if known.

Completing Cohort Profiles From Period Forecasts

The problem of the crossing prediction intervals is not a general feature of completing cohort mortality from period forecasts, but relates to an often-used approach of attributing uncertainty within the Lee–Carter forecasting methodology that is clearly inappropriate. Lee himself argued against using such an approach for cohort measures (Lee 2000) but never empirically elaborated on the point. Crossovers do not occur when measures are calculated on the implied death rates for each simulated time series trajectory (see Figure A5 in the online appendix). However, this approach has not been empirically tested to determine whether the obtained uncertainty intervals are well calibrated.

Conclusion

Drawing cohort profiles from period data and forecasts can lead to misleading conclusions, including erroneous fluctuation in cohort summary measures and implausible uncertainty in cohort forecasts. No one will be surprised that we find bias. But we expect many to be surprised by the magnitude of bias, even when cohort profiles and forecasts are drawn from a single-age, single-period grid. The danger is that we go down rabbit holes explaining deviations from expected trends when the deviations themselves are an artifact of our own data manipulation.

We argue that cohort summary measures of fertility, mortality, and migration should always be calculated on the basis of age–cohort data. When unavailable, age–period data should be split into Lexis triangles and reformatted to cohort data using the most plausible assumptions. In the case of cohort forecasting, crossing prediction intervals can be avoided by period-based forecasting methodologies that are not based on the Lee–Carter framework, or by simulation approaches within the Lee–Carter framework. However, we are of the opinion that uncertainty can only be properly accounted for with a cohort-based forecasting methodology. We end with a plea for researchers to further test and develop new cohort-based methods and models. ∎

References

Aassve, A., Cavalli, N., Mencarini, L., Plach, S., & Livi Bacci, M. (2020). The COVID-19 pandemic and human fertility. *Science, 369,* 370–371.

Agadjanian, V., & Prata, N. (2002). War, peace, and fertility in Angola. *Demography, 39,* 215–231.

Alburez-Gutierrez, D., Kolk, M., & Zagheni, E. (2021). Women's experience of child death over the life course: A global demographic perspective. *Demography, 58,* 1715–1735. https://doi.org/10.1215/00703370-9420770

Andreev, K. F., & Vaupel, J. W. (2006). *Forecasts of cohort mortality after age 50* (MPIDR Working Paper, No. WP 2006-012). Rostock, Germany: Max Planck Institute for Demographic Research. Retrieved from https://www.demogr.mpg.de/papers/working/wp-2006-012.pdf

Basellini, U., Camarda, C. G., & Booth, H. (2023). Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting, 39,* 1033–1049.

Brunner, B., & Kuhn, A. (2011). *Financial incentives, the timing of births, birth complications, and newborns' health: Evidence from the abolition of Austria's baby bonus* (Working Paper No. 48). Zurich, Switzerland: University of Zurich, Department of Economics. https://doi.org/10.2139/ssrn.1963809

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine, 26,* 3018–3045.

Chandra, S., Christensen, J., Mamelund, S.-E., & Paneth, N. (2018). Short-term birth sequelae of the 1918–1920 influenza pandemic in the United States: State-level analysis. *American Journal of Epidemiology, 187,* 2585–2595.

Crimmins, E. M. (1993). Demography: The past 30 years, the present, and the future. *Demography, 30,* 579–591.

Eurostat. (2016, March 4). *Asylum in the EU member states: Record number of over 1.2 million first time asylum seekers registered in 2015* (News release). Retrieved from https://ec.europa.eu/eurostat/documents/2995521/7203832/3-04032016-AP-EN.pdf/790eba01-381c-4163-bcd2-a54959b99ed6

Goodkind, D. M. (1991). Creating new traditions in modern Chinese populations: Aiming for birth in the Year of the Dragon. *Population and Development Review, 17,* 663–686.

Goodkind, D. M. (1995). The significance of demographic triviality: Minority status and zodiacal fertility timing among Chinese Malaysians. *Population Studies, 49,* 45–55.

Guzzo, K. B., & Hayford, S. R. (2020). Pathways to parenthood in social and family contexts: Decade in review, 2020. *Journal of Marriage and Family, 82,* 117–144.

Hartnett, C. S., & Gemmill, A. (2020). Recent trends in U.S. childbearing intentions. *Demography, 57,* 2035–2045.

Hiam, L., Harrison, D., McKee, M., & Dorling, D. (2018). Why is life expectancy in England and Wales 'stalling'? *Journal of Epidemiology and Community Health, 72,* 404–408.

Ho, J. Y., & Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: Retrospective observational study. *BMJ, 362,* k2562. https://doi.org/10.1136/bmj.k2562

Human Fertility Database. (2023). Rostock, Germany: Max Planck Institute for Demographic Research; Vienna, Austria: Vienna Institute of Demography. Available from www.humanfertility.org.

Human Mortality Database. (2023). Berkeley, CA (USA): University of California, Berkeley; Rostock, Germany: Max Planck Institute for Demographic Research. Available from www.mortality.org

Hyndman, R., Booth, H., Tickle, L., Maindonald, J., & Wood, S. (2023). *Demography: Forecasting mortality, fertility, migration and population data* [R package version 2.0]. Retrieved from http://CRAN.R-project.org/package=demography

Kaku, K. (1975). Increased induced abortion rate in 1966, an aspect of a Japanese folk superstition. *Annals of Human Biology, 2,* 111–115.

Kashyap, R. (2021). Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Population Studies, 75*(Suppl. 1), 47–75.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions: Physical Sciences and Engineering, 332,* 487–509.

Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death-rates in Great Britain and Sweden: Some general regularities and their significance. *Lancet, 223,* 698–703.

Lee, R. (2000). The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal, 4*(1), 80–91.

Lee, R. (2001, March). *Demography abandons its core*. Panel discussion remarks presented at the annual meeting of the Population Association of America, Washington, DC. Retrieved from https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/rlee/FormalDemog.pdf

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association, 87,* 659–671.

Lindstrom, D. P., & Berhanu, B. (1999). The impact of war, famine, and economic decline on marital fertility in Ethiopia. *Demography, 36,* 247–261.

Myrskylä, M., Goldstein, J. R., & Cheng, Y. A. (2013). New cohort fertility forecasts for the developed world: Rises, falls, and reversals. *Population and Development Review, 39,* 31–56.

Preston, S. H., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography, 43,* 631–646.

Shkolnikov, V. M., Jdanov, D. A., Andreev, E. M., & Vaupel, J. W. (2011). Steep increase in best-practice cohort life expectancy. *Population and Development Review, 37,* 419–434.

Thévenon, O., & Gauthier, A. H. (2011). Family policies in developed countries: A 'fertility-booster' with side-effects. *Community, Work & Family, 14,* 197–216.

Vogt, T., van Raalte, A., Grigoriev, P., & Myrskylä, M. (2017). The German East–West mortality difference: Two crossovers driven by smoking. *Demography, 54,* 1051–1071.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D., & Riffe, T. (2021, January 26). *Methods protocol for the Human Mortality Database* (Version 6). Available at https://mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf

Yang, Y., & Land, K. C. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications*. Boca Raton, FL: CRC Press.

Alyson A. van Raalte (corresponding author)
vanraalte@demogr.mpg.de

van Raalte • Max Planck Institute for Demographic Research, Rostock, Germany; Max Planck–University of Helsinki Center for Social Inequalities in Population Health, Rostock, Germany, and Helsinki, Finland; https://orcid.org/0000-0002-0676-8921

Basellini • Max Planck Institute for Demographic Research, Rostock, Germany; https://orcid.org/0000-0003-0292-1404

Camarda • French Institute for Demographic Studies (INED), Aubervilliers, France; https://orcid.org/0000-0003-0191-4093

Nepomuceno • Max Planck Institute for Demographic Research, Rostock, Germany; https://orcid.org/0000-0002-5687-5776

Myrskylä • Max Planck Institute for Demographic Research, Rostock, Germany; Max Planck–University of Helsinki Center for Social Inequalities in Population Health, Rostock, Germany, and Helsinki, Finland; Center for Social Data Science and Population Research Unit, University of Helsinki, Helsinki, Finland; https://orcid.org/0000-0003-4995-027X