Final Report

Submitted to the SCOR Foundation for Science

Coherent mortality forecasts by cause of death and disability level

Submitted by:

Marie-Pier Bergeron-Boucher
Principal investigator, Assistant Professor
Interdisciplinary Centre on Population Dynamics
University of Southern Denmark
Odense, Denmark

July 11, 2022

Contents

1 Background

Economic, social and public health institutions rely on mortality forecasts to estimate, among other things, health care and disability cost, plan social security policies or estimate the pension cost in an aging population. There is then a demand for valid and coherent forecasts for general mortality, but also for some components of mortality, such as causes of death and disability levels.

The number of forecasting models has increased greatly the last three decades. The many options include models based on different indicators [1]; models including cohort effects and smoking effects [2, 3, 4]; models accounting for coherence between populations and components [5, 6]; and so on [7]. A common procedure to forecast mortality is to extrapolate the trends in age-specific death rates. One commonly used model is the Lee-Carter (LC) model [8], which forecasts age-specific death rates in a log-linear way using principal component analysis. Many statistical offices use the LC model, or an extension of this model, to produce national projections. The advantages of this model include its simplicity, little subjective judgments are required and it can factor in uncertainty. However, this model has one major drawback: it tends to underpredict life expectancy. Many extensions and variants of the models have been suggested over the years [9, 7].

There is a reluctance to forecast mortality by other components than ages at death. The reasons being that (1) data access and quality is lesser for mortality components other than ages, such as causes of death, health states and smoking effects; (2) models including such information can be subject to many methodological limitations; and (3) fewer models are available, especially forecast models by health states.

Nevertheless, mortality forecasts for different components of mortality are needed. Oeppen (2008) [6] lists two main advantages to disaggregating mortality forecasts: 1) the risks are known for diverse factors, as age, sex, cause of death, etc. and 2) spending (on research, capital investment, preventive measures or palliative care) could be more efficient if forecasts were known for diverse mortality components. Also, such forecasts can allow for a better specification of the morbidity process.

Given the need for coherent forecasts of mortality by different components, the SCOR Foundation for Science have been funding the current project. The project had three aims:

1. Evaluate forecast models by cause of death and suggest an approach which is not subject to known limitations.
2. Develop models to forecast mortality by health state.
3. Disseminate the knowledge via peer-reviewed publications and presentations.

This report presents the main findings and general conclusions of aims 1 and 2. More details on the methods, data, results and discussion can be found in the papers written as part of aim 3, listed in section 6.

2 Schedule

The project was initially planned to start in January 2020 and to end in December 2020. However, the principal investigator had a career break due to maternity leave from November 2020 to October 2021. The SCOR Foundation for Science thus agreed to extend the project duration until July 2022.

Most of the work on aim 1 was done in the first half of 2020, while most of the work on aim 2 was done in the first half of 2022.

3 The people

The project was carry out by the principal investigator (PI), Marie-Pier Bergeron-Boucher, Assistant Professor at the Interdisciplinary Center on Population Dynamics, University of Southern Denmark. She was the lead on all papers and presentations produced during the project. The PI collaborated with a number of colleagues over the course of the project, including:

1. Søren Kjærgaard
   Postdoc, Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark; and CREATES, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark.

2. Jim Oeppen
   Associate professor, Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark.

3. Violetta Simonacci
   Research Fellow in Statistics, University of Naples Federico II, Naples, Italy.

4. Cosmo Strozza
   Postdoc, Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark.

4 Aim 1: Forecasts by cause of death

Mortality forecasts by age and cause of death are important for more efficient spending on, for example, healthcare and medical technology. Different approaches have been suggested to forecast mortality by cause of death. A common approach is to forecast age-specific death rates for each cause independently and then sum them to obtain an all-causes mortality forecast [10, 11]. However, there is a reluctance in including the cause of death dimension to the forecast, as forecasts by cause are confronted with many methodological problems. Five main problems have been highlighted, with some of them having been resolved in recent literature:

1. Causes-of-death forecasts are often dominated by an increase or slower decrease in certain

causes of death, creating an inherent pessimism, so that life expectancy forecasts by cause are lower than in non-disaggregated forecasts [10]. However, this pessimism have been shown to only emerge under certain particular conditions, such as in linear extrapolative models of death rates [6, 12]. This issue is then model-specific and not a general problem with forecast by cause of death.

2. Extrapolative models by cause can lead to unrealistic trends [11]: If mortality of a given cause has been increasing, it will keep doing so in the forecast, which can lead to a small cause of death becoming the dominant cause in the forecasts.

3. Modifications to the International Classification of Diseases (ICD) create discontinuities over time, making the use of long time series difficult. This issue is still a current problem and will remain so until efforts are made to ensure continuity between ICD revisions.

4. Trajectories of causes of death are generally considered to be independent. But in reality, they are interconnected [6]. This correlation between the components is often ignored when forecasting. To remedy to this problem, Oeppen [6] suggested abandoning the conventional way to forecast death rates and to forecast death distributions using Compositional Data Analysis (CoDA). Compositional data are vectors of relative information, constrained to sum to a constant, such as proportions. Oeppen used the distribution of death by age and cause from multiple-decrement life tables to forecast mortality by age and cause simultaneously. By treating life table deaths as compositional data and using a CoDA framework, the deaths are constrained to vary between 0 and the life table radix (e.g. 1 or 100,000), conditioning the relationship between components. Deaths are thus directly dependent on each other on the aggregate level, such that the decrease in deaths from one cause will lead to an increase in deaths from at least one other cause.

5. Forecasts by cause are often inconsistent with all-causes forecasts. This issue relates to how changes in mortality by age and cause interact. Deaths have been simultaneously shifted towards older ages and more diverse causes of death over time [13]. Hence, the question is how the different methods model this shift, which varies with the level of disaggregation and the number of categories.

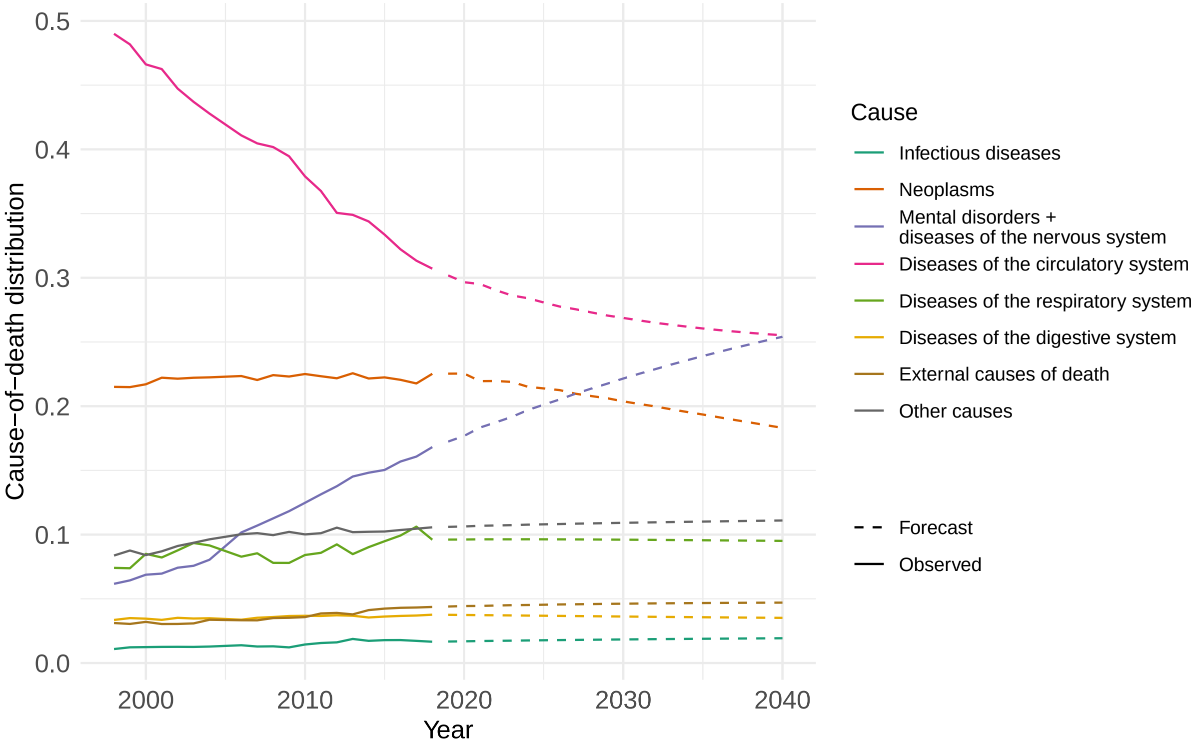
We found in the literature that some of these limitations are still current issues. How the age and cause dimensions interact when forecasting has been overlooked. Should they be forecast simultaneously or separately? How do different approaches affect the forecast?

To answer these questions we used a model framework which models ages and causes simultaneously; causes-of-death distribution (CDD) within each age group; and age-at-death distribution (ADD) within each cause. All approaches are performed in a CoDA framework to avoid issue 4. We specified multiple models within each of the three approaches to obtain a better understanding of the importance of the age and cause interactions for forecasting.

To evaluate the accuracy of the various models, we used an out-of-sample approach. The out-of-sample method consists in forecasting observed trends, based on a fitting period outside the forecast horizon. We did out-of-sample tests for multiple fitting periods and forecast horizons.

We found that forecasting the CDD within each age group generally provides the most accurate forecasts, while avoiding limitations 1,4 and 5 listed above. To avoid issue 2, we suggested forecasting the CDD with a Holt-linear damped trend, rather than an ARIMA model with linear trend (the more traditional approach). Instead of forecasting a continuous increase into the future, the Holt-linear damped trend "dampens" the trend at some point in the future. By using this model, we limit the transfer of deaths towards the cause which showed the fastest increase in the fitting period (issue 2). Figure 1 shows an example of the forecast of the cause-of-death distribution with the best performing approach in Australia.

Figure 1: Cause-of-death distribution observed and forecast over time for Australian females, 1998 to 2030



Data and calculation: Data from [14] and [15] and calculation by Marie-Pier Bergeron-Boucher (PI) and Søren Kjærgaard. Reference: Bergeron-Boucher, Marie-Pier, and Søren Kjærgaard. 2022. "Mortality Forecasts by Age and Cause of Death: How to Forecast Both Dimensions?" SocArXiv. June 28. doi:10.31235/osf.io/d7hbp

We concluded that recent methodological developments allowed to resolve some of the issues with forecasts by cause of death. With this project, we contributed to the discussion around forecasting by cause of death by assessing how to forecast both the age and cause dimensions within a CoDA framework. Nowadays, forecasting by cause of death is possible without carrying too much methodological problems.

5 Aim 2: Forecasts by health state

We live longer than ever. But are we also living healthier? Whether the extra years of life are being lived in good or poor health is of importance to society. The quality of these extra years of life has implications for individuals and society, including an increased burden of caregiving for surviving family members, increased pressure on healthcare systems, as well as changes in the dependency ratio [16]. As forecasts support social, economic and medical decisions, as well as individuals' choices, there is a clear rationale for forecasting healthy life expectancy (HLE). However, forecast of HLE remains uncommon.

HLE estimates the number of years expected to live in good health, informing on the overall population health. There are two main approaches to estimate HLE: the Sullivan's method and the multistate life table (MSLT) method. The Sullivan method estimates the number of years lived in good or poor health using cross-sectional information on mortality and prevalence in each health-state. The MSLT model estimates health-prevalence as the results of transitions across states (e.g. healthy, unhealthy and death).

Forecasts of HLE are often based on scenarios. However, scenarios are more of a "what if" situation and it is arduous to assess the likelihood of each scenario [17]. Other authors used microsimulation to forecast health and mortality for individuals. This approach has the disadvantage to be very data demanding, making it hard to apply to diverse populations and contexts [18]. Other authors forecast HLE assuming that the transition probabilities across health states in the MSLT remain constant in the forecast for different components or variables, e.g. age, sex and education [19, 20].

Stochastic models to forecast HLE are rare. Majer and colleagues [21] were among the first to introduce such a model. They suggested using the LC model [8] to forecast the age-specific transition rates estimated with a MSLT, using a separate forecasts for the mortality rates of the nondisabled, mortality rates of the disabled and the incidence rates. The model assumes independence between mortality rates for non-disabled and disabled and incidence rates. There is, however, a dependence between the different states. If individuals become less and less disabled over time, more and more individuals will remain nondisabled and die in that state.

During the duration of the project, we developed models which can (1) forecast simultaneously mortality and health prevalence; (2) consider the dependence between age-groups and between health-states; and (3) account for changes in transition probabilities and health prevalence over time. We developed two models: one based on the Sullivan method and the other on the MSLT method. Both models make use of CoDA, which has been shown to account for correlation between mortality and health components [22].

There are differences in the estimation and forecast of HLE whether we used the Sullivan or the MSLT models. But, we found that the differences between models was not significant in most cases. Both Sullivan and MSLT models showed similar trends over time. The MSLT model is generally seen as the best approach in estimating the number of years lived in good or poor health, but it is more data demanding. If the data to estimate MSLT are not available, the Sullivan method can be an acceptable alternative for monitoring such trends in health expectancies, as previously shown [23, 24]. Figure 2 shows the forecast of life expectancy, disability-free life expectancy and severe disability-free life expectancy for Spanish females. We see that more and more years are being lived without limitations. However, the increase in disability-free life expectancy (in blue) increased slower than the total life expectancy (in red), suggesting also more years lived with limitations.

Figure 2: Life expectancy, disability-free life expectancy and severe disability-free life expectancy at age 50 observed and forecast with the Sullivan and MSLT approaches, Spanish females, 2004–2030

Data: Data from [25] and [15]. The results are based on data from Eurostat, EU-SILC 2004-2020.
The responsibility for all conclusions drawn from the data lies entirely with the authors.
Calculation: Calculations by Marie-Pier Bergeron-Boucher (PI), Cosmo Strozza, Violetta Simonacci and Jim Oeppen.
Reference: Bergeron-Boucher, Marie-Pier, Cosmo Strozza, Violetta Simonacci, and Jim Oeppen. 2022. “Modeling and Forecasting Healthy Life Expectancy with Compositional Data Analysis.” SocArXiv. July 9. doi:10.31235/osf.io/ksrbj.

As only a few models were developed to forecast HLE, the introduced models are a welcome addition to the literature. These models solved some of the methodological issues with forecast of HLE, by simultaneously forecasting health and mortality in a coherent manners, while accounting for changes in prevalence and transition probabilities over time. These models are a first step towards simple and coherent forecasts of the number of years lived in good and poor health and can be further developed by future research.

6 Aim 3: Knowledge dissemination

6.1 Peer-reviewed articles

During the duration of the project, two articles were produced. At the time the final report was submitted, both articles were under review in peer-reviewed journal. Nevertheless, both articles can be found on a preprint server:

- Bergeron-Boucher, Marie-Pier, and Søren Kjærgaard. 2022. "Mortality Forecasts by Age and Cause of Death: How to Forecast Both Dimensions?" SocArXiv. June 28. doi:10.31235/osf.io/d7hbp. `https://osf.io/preprints/socarxiv/d7hbp/`
- Bergeron-Boucher, Marie-Pier, Cosmo Strozza, Violetta Simonacci, and Jim Oeppen. 2022. "Modeling and Forecasting Healthy Life Expectancy with Compositional Data Analysis." SocArXiv. July 9. doi:10.31235/osf.io/ksrbj. `https://osf.io/preprints/socarxiv/ksrbj/`

6.2 Presentations

The methods developed and results produced were presented during three seminars/conferences:

- Bergeron-Boucher, Marie-Pier. 2020. "Mortality Forecasts by Age and Cause of Death: How to Forecast Both Dimensions?" Presentation, SCOR seminar, Online (October 28).
- Bergeron-Boucher, Marie-Pier, Cosmo Strozza, Violetta Simonacci and Jim Oeppen. 2022. "Modelling and forecasting healthy life expectancy. A Compositional Data Analysis approach?" Presentation, SCOR seminar, Paris (April 6).
- Bergeron-Boucher, Marie-Pier, Cosmo Strozza, Violetta Simonacci and Jim Oeppen. 2022. "Modelling and forecasting healthy life expectancy. A Compositional Data Analysis approach?" Presentation, European Population Conference (EPC), Groningen, the Netherlands (July 2).

7 Conclusions

Mortality forecasts by age, cause of death and health state are important for more efficient planning. However, there was a reluctance to include information on causes of death and health states to the forecasts, due to data and methodological problems. With the support from the SCOR Foundation for Science, we showed that many of the methodological problems with such forecasts can be solved or avoided. We hope that future research will build on the developed methods to provide forecasts for multiple components of mortality.

References

[1] M.-P. Bergeron-Boucher, S. Kjærgaard, J. Oeppen, and J. W. Vaupel, "The impact of the choice of life table statistics when forecasting mortality," Demographic Research, vol. 41, pp. 1235–1268, 2019.

[2] A. E. Renshaw and S. Haberman, "A cohort-based extension to the Lee–Carter model for mortality reduction factors," Insurance: Mathematics and Economics, vol. 38, no. 3, pp. 556–570, 2006.

[3] Y. Li and A. E. Raftery, "Accounting for smoking in forecasting mortality and life expectancy," The Annals of Applied Statistics, vol. 15, no. 1, p. 437, 2021.

[4] F. Janssen, A. Bardoutsos, S. El Gewily, and J. De Beer, "Future life expectancy in Europe taking into account the impact of smoking, obesity, and alcohol," Elife, vol. 10, p. e66590, 2021.

[5] N. Li and R. Lee, "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method," Demography, vol. 42, no. 3, pp. 575–594, 2005.

[6] J. Oeppen, "Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data," in Population Association of America (PAA) 2008 Annual Meeting, 2008.

[7] H. Booth and L. Tickle, "Mortality modelling and forecasting: A review of methods," Annals of Actuarial Science, vol. 3, no. 1-2, pp. 3–43, 2008.

[8] R. D. Lee and L. R. Carter, "Modeling and forecasting US mortality," Journal of the American Statistical Association, vol. 87, no. 419, pp. 659–671, 1992.

[9] H. Booth, R. J. Hyndman, L. Tickle, and P. De Jong, "Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions," Demographic Research, vol. 15, pp. 289–310, 2006.

[10] J. R. Wilmoth, "Are mortality projections always more pessimistic when disaggregated by cause of death?," Mathematical Population Studies, vol. 5, no. 4, pp. 293–319, 1995.

[11] G. Caselli and J. Vallin, Est-il utile de prendre en compte les causes de décès pour extrapoler les tendances de la mortalité? Università degli Studi di Roma-La Sapienza, Dipartimento di Scienze Demografiche, 1996. Rome, 22 p.

[12] E. Tabeau, P. Ekamper, C. Huisman, and A. Bosch, "Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends," European Journal of Population/Revue européenne de Démographie, vol. 15, no. 2, pp. 153–183, 1999.

[13] M.-P. Bergeron-Boucher, M. D. Pascariu, J. M. Aburto, J.-A. Alvarez, U. Basellini, S. Rizzi, and J. W. Vaupel, "Alternative forecasts of Danish life expectancy," in Developments in Demographic Forecasting, pp. 131–151, Springer, Cham, 2020.

[14] WHO, "WHO Mortality Database, url: https://www.who.int/data/data-collection-tools/who-mortality-database," 2021.

[15] HMD, "Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), url: http://www.mortality.org," October 2021.

[16] H. Beltrán-Sánchez, S. Soneji, and E. M. Crimmins, "Past, present, and future of healthy life expectancy," Cold Spring Harbor Perspectives in Medicine, vol. 5, no. 11, p. a025957, 2015.

[17] H. Booth, "Demographic forecasting: 1980 to 2005 in review," International Journal of Forecasting, vol. 22, no. 3, pp. 547–581, 2006.

[18] C. Jagger, E. M. Crimmins, Y. Saito, R. T. De Carvalho Yokota, H. Van Oyen, and J.-M. Robine, eds., International Handbook of Health Expectancies, vol. 9 of International Handbooks of Population. Cham: Springer International Publishing, 2020.

[19] J. P. Ansah, R. Malhotra, N. Lew, C.-T. Chiu, A. Chan, S. Bayer, and D. B. Matchar, "Projection of young-old and old-old with functional disability: does accounting for the changing educational composition of the elderly population make a difference?," PLoS One, vol. 10, no. 5, p. e0126471, 2015.

[20] N. Biddle and H. Crawford, "Projections of the number of Australians with disability aged 65 and over eligible for the National Disability Insurance Scheme: 2017–2026," Australasian Journal on Ageing, vol. 36, no. 4, pp. E43–E49, 2017.

[21] I. M. Majer, R. Stevens, W. J. Nusselder, J. P. Mackenbach, and P. H. van Baal, "Modeling and forecasting health expectancy: theoretical framework and application," Demography, vol. 50, no. 2, pp. 673–697, 2013.

[22] M.-P. Bergeron-Boucher, V. Canudas-Romo, J. E. Oeppen, and J. Vaupel, "Coherent forecasts of mortality with Compositional Data Analysis," Demographic Research, vol. 37, no. 17, pp. 527–566, 2017.

[23] C. D. Mathers and J.-M. Robine, "How good is Sullivan's method for monitoring changes in population health expectancies?," Journal of Epidemiology & Community Health, vol. 51, no. 1, pp. 80–86, 1997.

[24] Y. Murakami, R. Tsukinoki, K. Miura, T. Okamura, A. Kadota, A. Okayama, and H. Ueshima, "Comparison of methods for calculating healthy life expectancy in a Japanese population: an analysis of Nippon Data90," Revue d'Épidémiologie et de Santé Publique, vol. 66, pp. S329–S330, 2018.

[25] Eurostat, "Database: Statistics on Income and Living Conditions (EU-SILC), 2004 to 2020," 2022.