The use of population and death count data to inform time-trends in the rates of death and years of life lost to death from tobacco and/or alcohol related diseases

Version 1.0.1

March 2023

Contact:
Duncan Gillespie
Email: duncan.gillespie@sheffield.ac.uk

Contents

1 Introduction

Investigating the past trends in deaths associated with tobacco and alcohol consumption is fundamental to evaluating the effect of past policy and appraising the potential effects of new policy.

The on-going STAPM programme of work at The University of Sheffield is developing public health economic models for policymaker decision support in the fields of alcohol and tobacco control. This work builds on the methods used in the Sheffield Alcohol Policy Model (A. Brennan et al. 2009; Alan Brennan et al. 2015) and is being conducted at a range of geographic scales, including national and local authority level.

The two objectives for the use of mortality data for the STAPM modelling are:

- To investigate past trends in alcohol and/or smoking related deaths and the consequent years of life lost, and the impact of past policy changes on these outcomes.
- To appraise the future effects of potential interventions or changes to policy on deaths related to alcohol and/or tobacco and the consequent years of life lost.

This report is focused on explaining the choice of method used in the STAPM modelling for calculating rates of death and life years lost to death.

2 Methods

2.1 Population and death count data

Data on annual counts of deaths and mid-year population sizes - separately for England, Wales and Scotland. England and Wales data provided by the Office for National Statistics; Scotland data provided by National Records Scotland. Time series of data are used from 2001 for England and Wales, and from 2008 for Scotland. Data for national-level STAPM modelling are stratified by Index of Multiple Deprivation (IMD) quintiles, sex (m/f), single years of age up to the open age interval of 90+ years.

2.2 Preliminary data processing

Deaths are grouped by the ICD-10 definitions of diseases attributable to tobacco and alcohol developed for the STAPM modelling (see the information on the relative risks of disease).

Oesophageal cancer is defined by the ICD-10 code "C15", but has two subtypes "SCC" and "AC" that have different risk relationships to tobacco and alcohol. These subtypes are not identifiable based on the ICD-10 coding system. External data is therefore used to estimate how total oesophageal cancers are split between subtypes (see the disease list reports (Angus et al. 2018; Webster et al. 2018)).

consistency of disease coding

- From 2011 onwards, no acute intoxication deaths recorded (F10.0)
- Before 2011, no acute pancreatitis (alcohol induced) deaths recorded (K85.2)

For some other conditions, e.g. vascular dementia, the time trends indicate changes in the practices of coding deaths that I'm still trying to understand.

2.3 Calculating death rates

Cause-specific central death rates are calculated for each population subgroup by dividing the annual number of deaths from each cause by the mid-year population size.

2.4 Smoothing death rates

Death rates are smoothed by applying a 2d moving average over age and period, by condition, sex and IMD quintile.

2.5 Age-standardising death rates

Age-standardisation uses the 2013 European Standard Population, which provides a standard set of weights for calculating average mortality across all ages.

Age-standardised rates are also calculated based on the population structure in the data for a chosen index year.

2.6 Forecasting death rates

Mortality forecasting methods for STAPM are exploratory

The aim is to project past trends in the age-patterns of cause-specific mortality into the future separately for each sex and Index of Multiple deprivation population subgroup.

It is important to consider how mortality might change into the future due to factors that are not related to tobacco or alcohol. For example, the effect of a future policy on cardiovascular disease mortality might be over-estimated if the ongoing declines in cardiovascular mortality rates are not modelled (Castillo et al. 2014). This presents the challenge of producing forecasts of the age-patterns of cause-specific mortality for each sex and Index of Multiple Deprivation population subgroup. Intervention effects would then be modelled as adjustments to these trends.

The advantages of conducting a cause-specific forecast is that it would enable us to incorporate different trends for different diseases and different subgroups e.g. that that there are fewer deaths from Ischemic Heart disease but more from dementia, and that lung cancer rates are declining in males but rising in females. Cause-specific forecasts stratified by IMD quintiles would also allow us to incorporate the different socio-economic trends in past mortality and to project an expectation of the future socio-economic trends in mortality.

There are several methods that might be used (Li et al. 2019; Arnold and Sherris 2013; Bergeron-Boucher et al. 2017). One issue with cause-specific mortality forecasting is that the forecasts of cause-specific mortality rates might not add up to / correspond with the forecasts of the aggregate all-cause mortality rates. Wilmoth (1995) observed that mortality projections are more pessimistic when disaggregated by cause of death because the causes of death that are decreasing the most slowly or that are increasing come to dominate (i.e. the aggregate trends are most influenced by trends in the causes with the most deaths).

2.6.1 Lee-Carter methods

Our current approach in STAPM is to conduct a Lee-Carter forecast (Lee and Carter 1992) separately for mortality from different causes within ten subgroups defined by combinations of sex and IMD quintile. The Lee-carter method is a "proportional rate of change model". It summarises the age-period surface of log-mortality rates ($\log m_{xt}$) in terms of vectors **a** and **b** along the age dimension and **k** along the time dimension such that

$$\log m_{xt} = a_x + b_x k_t + e_{xt} \tag{1}$$

with restrictions such that the b's are normalized to sum to one and the k's sum to zero, so the a's are average log rates.

The vector **a** can be interpreted as an average age profile of mortality, the vector **k** tracks mortality changes over time, and the vector **b** determines how much each age group changes when k_t changes. When k_t is linear on time each age group changes at its own exponential rate, but this is not a requirement of the model. The error term reflects age-period effects not captured by the model.

The vector **a** is estimated by averaging log rates over time and **b** and **k** via a singular value decomposition (SVD) of the residuals (this is essentially a method for approximating a matrix as the product of two vectors).

Having reduced the time dimension of mortality to a single vector **k**, this is then projected forward by an ARIMA time series model. In our forecasting for STAPM we don't currently incorporate forecast error but if we were to then it would be based on the error in this time series model.

Future mortality is then estimated by inputting the extrapolated values of k_t back into (1).

2.6.2 Experimental application

Before we conduct our cause and subgroup specific mortality forecast, we have experimented with reducing the noise in the matrix of observed rates of mortality by smoothing each matrix over age and period. To do so, we estimated a moving average based on the application of a two-dimensional sliding window (adjusting the degree of smoothing by adjusting the size of the window). We found that smoothing is particularly important for the rarer causes, as the projected future trends for these causes can be strongly influenced by noise in the observed rates.

We conducted the forecast using the R package `demography` (Rob J Hyndman with contributions from Heather Booth and Leonie Tickle and John Maindonald. 2019). We conducted a set of cause-specific forecasts and also a forecast of the remaining other-cause mortality. In the `mort.tools` R package, the experimental forecast method is implemented by the function `mort.tools::CombinedForecast()`, which is a wrapper that calls the function `mort.tools::MortalityForecast()` to run the forecast for each cause and groups the outputs together. The forecast is run for ten population subgroups defined by combinations of sex and IMD quintile.

The parameters that control the forecast for each cause are contained in a separate spreadsheet, which is input into `mort.tools::CombinedForecast()` as `forecast_params` (see Figure). The parameters in this spreadsheet allow the control of the period and and age range that are used to inform the forecast for each cause, and the degree of smoothing to apply to the past trends in mortality.

Type	Condition	year_start	year_end	age_start	age_end	smooth_age	smooth_year	sex_focus	do_forecast	Notes
Cancer	Oral_cavity	2001	2016	45	89	11	11	both	yes	May need to check noise at younger ages
Cancer	Pharynx	2001	2016	45	89	11	11	both	yes	
Cancer	Lung	2009	2016	45	89	13	5	both	yes	Females were reaching a peak in 2010. This was making the male and female trends opposite, therefore the age has been limited to 2010:2016 only
Cancer	Nasopharynx_sinonasal								no	Not a clear trend in observed data.
Cancer	Larynx	2001	2016	45	89	11	11	both	yes	Forecasted age trend a bit weird.
Cancer	Oesophageal_AC	2001	2016	45	89	11	11	both	yes	
Cancer	Oesophageal_SCC	2001	2016	45	89	7	11	both	yes	
Cancer	Stomach	2001	2016	45	89	11	11	both	yes	
Cancer	Pancreas	2001	2016	45	89	11	11	both	yes	
Cancer	Liver	2001	2016	45	89	11	11	both	yes	Can run from 18, but noisey at younger ages.
Cancer	Colorectal	2001	2016	45	89	17	11	both	yes	Smoothing issue: in order to reduce the tail kick, reduces the fit to the observed data
Cancer	Kidney	2001	2016	45	89	13	13	both	yes	Not a clear trend in observed data.
Cancer	Lower_urinary_tract	2001	2016	65	89	13	11	both	yes	Gets noisy if include younger ages than 60, but this is where deaths drop less than 30 anyway
Cancer	Bladder	2001	2016	45	89	11	11	both	yes	
Cancer	Cervical	2001	2016	25	89	13	13	Female	yes	Female only. Screening programme starts at age 25

Given all the potential unreliability of a cause-specific forecast, it is important to consider how far in the future the trends in mortality might be extrapolated reliably. Once this final year is determined, if we want the STAPM projections to run beyond it, we assume that the mortality rates in the control arm of our model will remain constant into the future from this final year.

There is lots of room for improvement in this forecasting method, particularly looking into incorporating forecast uncertainty, and developing methods that improve the consistency between the forecasts of cause-specific mortality and those of all-cause mortality (Li et al. 2019; Arnold and Sherris 2013; Bergeron-Boucher et al. 2017).

References

Angus, Colin, M Henney, L Webster, and Duncan Gillespie. 2018. *Alcohol-Attributable Diseases and Dose-Response Curves for the Sheffield Alcohol Policy Model Version 4.0*. https://doi.org/10.15131/shef.data.6819689.v1.

Arnold, Séverine, and Michael Sherris. 2013. "Forecasting Mortality Trends Allowing for Cause-of-Death Mortality Dependence." Journal Article. *North American Actuarial Journal* 17 (4): 273–82. https://doi.org/10.1080/10920277.2013.838141.

Bergeron-Boucher, Marie-Pier, Vladimir Canudas-Romo, Jim Oeppen, and James W. Vaupel. 2017. "Coherent Forecasts of Mortality with Compositional Data Analysis." Journal Article. *Demographic Research* 37: 527–66. https://doi.org/10.4054/DemRes.2017.37.17.

Brennan, Alan, Petra Meier, Robin Purshouse, Rachid Rafia, Yang Meng, Daniel Hill-Macmanus, Colin Angus, and John Holmes. 2015. "The Sheffield Alcohol Policy Model – a Mathematical Description." Journal Article. *Health Economics* 24 (10): 1368–88. https://doi.org/10.1002/hec.3105.

Brennan, A, R Purshouse, K Taylor, R Rafia, A Booth, D O'Reilly, T Stockwell, et al. 2009. "Modelling the Potential Impact of Pricing and Promotion Policies for Alcohol in England: Results from the Sheffield Alcohol Policy Model Version 2008 (1-1). Independent Review of the Effect of Alcohol Pricing and

Promotion: Part b." Report. University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.95621!/file/PartB.pdf.

Castillo, Maria Guzman, Duncan OS Gillespie, Kirk Allen, Piotr Bandosz, Volker Schmid, Simon Capewell, and Martin O'Flaherty. 2014. "Future Declines of Coronary Heart Disease Mortality in England and Wales Could Counter the Burden of Population Ageing." *PLoS One* 9 (6): e99482.

Lee, Ronald D, and Lawrence R Carter. 1992. "Modeling and Forecasting US Mortality." Journal Article. *Journal of the American Statistical Association* 87 (419): 659–71.

Li, Han, Hong Li, Yang Lu, and Anastasios Panagiotelis. 2019. "A Forecast Reconciliation Approach to Cause-of-Death Mortality Modeling." Journal Article. *Insurance: Mathematics and Economics* 86: 122–33. https://doi.org/10.1016/j.insmatheco.2019.02.011.

Rob J Hyndman with contributions from Heather Booth and Leonie Tickle and John Maindonald. 2019. *demography: Forecasting Mortality, Fertility, Migration and Population Data.* %7Bhttps://CRAN.R-project.org/package=demography%7D.

Webster, Laura, Colin Angus, Alan Brennan, and Duncan Gillespie. 2018. "Smoking and the Risks of Adult Diseases." The University of Sheffield. https://doi.org/10.15131/shef.data.7411451.v1.

Wilmoth, J. R. 1995. "Are Mortality Projections Always More Pessimistic When Disaggregated by Cause of Death?" Journal Article. *Math Popul Stud* 5 (4): 293–319, 377. https://doi.org/10.1080/08898489509525409.