Indirect Estimation of Cause-of-Death Mortality from Life Expectancy.

Andrea Nigri [1], Susanna Levantesi [2], Gabriella Piscopo [3]

August 14, 2021

[1]University of Foggia
[2]Sapienza, University of Rome
[3]University of Naples

Outline

Outline

- Background & Motivation
- Indirect Estimation - State of the Art
- Causes of Death indirect estimation - Methodological framework
- Application: Spain and Russia - Case studies
- Conclusion & Limitations

Background and Motivation

Motivation

- Life expectancy at birth has attracted interest in various fields. Its appeal relies on the ability to enclose and summarize all the factors affecting longevity.
- Estimating age-specific vital rates is needed to analyze demographic patterns at different ages.
- More granular information is provided by cause-of-death (CoD) rates...unfortunately, their availability across countries and over the years is not always guaranteed.
- We propose an indirect model to produce estimates of death rates due to specific causes using the summary indicator of life expectancy at birth

Background

- The rise in human longevity during the 20th century and the recent decline in life expectancy
- Increase in mortality for some classes of causes of death
- Overall mortality as a composition of causes of death... implausible estimation modeling each cause of death independently
- **Our approach:** we exploit the relationship between cause-of-death mortality rates and life expectancy at birth in log scale to reconstruct, for a given age, the cause-specific mortality pattern. **Our contribution:** (i) ensuring the consistency of CoD estimations against the total mortality. (ii) indirect estimations allow us to overcome the problem of incomplete data.

Indirect Estimation

Indirect Estimation

- Demography has a long-standing tradition in indirect estimation approach (United Nations, 1955, 1967; Murray et al., 2003, Brass et al., 1968;1971,Wilmoth et al., 2012, Ševčíková et al. 2016 and Pascariu et al. 2020)
- **"Indirect"** qualifies the demographic estimation technique that origins in the fact that such technique produces estimates of certain parameters on the basis of information that is only indirectly related to its value.

Indirect Estimation

- **Ševčíková et al. 2016**
  - Estimate the Lee-Carter parameters α_a, k_t and β_x using the observed death rates $m_{a,t}$
  - For a given value of projected life expectancy at birth $e_0(\tau)$, the method solves for future k_τ
  - The age specific log-death rates are derived as follows:

$$\log\left(m_{a,\tau}\right) = \hat{\alpha}_a + \hat{\beta}_a \hat{k}_\tau$$

- **Linear-Link model**
  - by Pascariu et al. (2020) derives specific death rates $a, m_{a,t}$, as a linear function of the logarithm of life expectancy at birth ($e_{0,t}$) and at time t given by:

$$\log\left(m_{a,t}\right) = \beta_a \log\left(e_{0,t}\right) + \nu_a k + \varepsilon_{a,t}$$

Causes of Death indirect estimation - Methodological framework

Model

$D_x(c,t) \sim Poisson\{E_x(t) \cdot m_x(c,t)\}$ and $\forall x \in X$, the aim is to convert a given value of $e_0(t)$, into a list of cause-specific death rates $m_x(c,t)$. Formally:

$$\log m_x(c,t) = \beta_x(c) \log e_0(t) + \delta_x(c)\gamma_x(t) \tag{1}$$

- $\beta_x(c)$: sensitivity of cause-specific mortality to the variation of $\log e_0(t)$,
- $\delta_x(c)$: provides the cause-of-death mortality structure,
- $\gamma_x(t)$ is the parameter which satisfies the condition:

$$\sum_{c \in C} \hat{m}_x(c,t) = m_x(t) \tag{2}$$

Parameters' estimation

$\forall x \in X$ given $m_x(c,t)$ and $m_x(t)$ as inputs, our model is fitted by maximizing the log-likelihood function, $\log L\left[\beta_x(c), \delta_x(c), \gamma_x(t)\right]$, which is given by:

$$\log L\left[\beta_x(c), \delta_x(c), \gamma_x(t)\right] = \sum_{x,t}\left\{D_x(c,t)\left[\beta_x(c)\log e_0(t) + \delta_x(c)\gamma_x(t)\right] - E_x(t)\exp\left[\beta_x(c)\log e_0(t) + \delta_x(c)\gamma_x(t)\right]\right\} + C \tag{3}$$

- I STEP: We propose to estimate the parameters modifying the scheme proposed by Brouhns et al. 2002 based on the Newton-Raphson algorithm.
- II STEP: Optimizing $\hat{\gamma}_x(t)$, where the optimized value $\hat{\gamma}_x^*(t)$ satisfies the condition in eq. 2. Hence, denoting $\Delta_x(c,t) = \left|\sum_{c\in C}\log \hat{m}_x(c,t) - \sum_{c\in C}\log m_x(c,t)\right|$ as the difference between the sum of estimated and observed cause-specific mortality rates in log scale, given a certain age x, the optimization problem can be formalized as follows:

$$\begin{aligned} &\min_{\gamma_x(t)} \ \Delta_x(c,t) \\ &\text{subject to} \\ &\sum_{c\in C}\hat{m}_x(c,t) = m_x(t) \end{aligned} \tag{4}$$

Application - Spain and Russa case studies

Human Cause-of-Death Database

Three time windows (w) for the model fitting: 1983-2002 (w=1), 1988-2002 (w=2), 1993-2002 (w=3). While, the estimation period is set to 2003-2012 and remains the same.

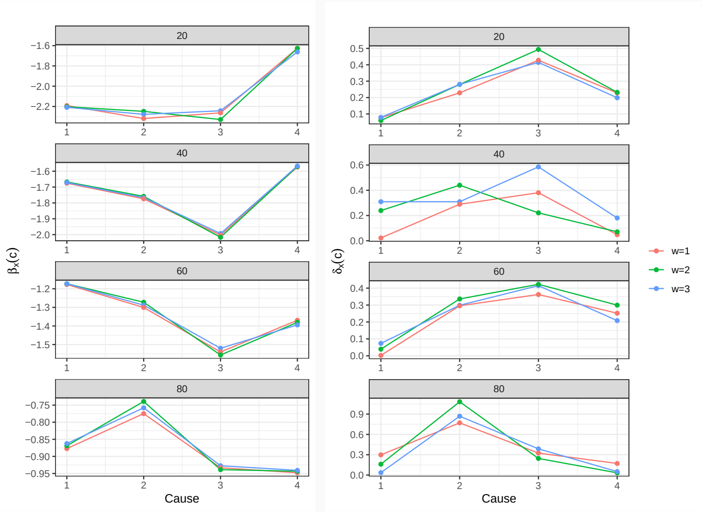
Spain:

- (1) Cancer (C00-D48),
- (2) Circulatory (I00-I52, G45, I60-I69, I70-I99),
- (3) Respiratory (J00-J22, U04, J30-J98) and Infectious (A00-B99),
- (4) Others (residual class).

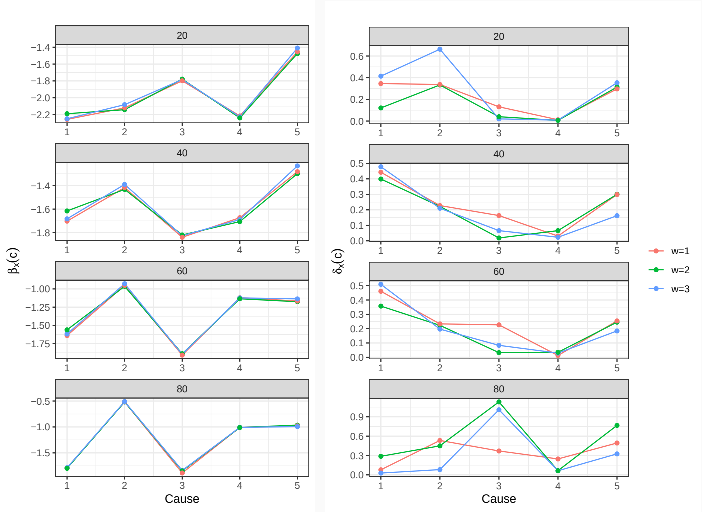
Russia:

- (1) Wholly attributable to alcohol (F10, K70, K74,X45),
- (2) Circulatory (I00-I52, G45, I60-I69, I70-I99),
- (3) Transportation accidents (V01-V99),
- (4) Cancer (C00-D48),
- (5) Others (residual class).

Spain: Parameter Estimation



Russia: Parameter Estimation



Numerical Results

Table 1: RMSE and MAE over years 2003-2012 for different time windows of estimation. Spain.

Cause	1983-2002		1988-2002		1993-2002	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
1: Cancer	0.000684	0.000291	0.000400	0.000200	0.000260	0.000121
2: Circulatory	0.000776	0.000250	0.000442	0.000176	0.000372	0.000152
3: Respir. + Infect.	0.000424	0.000148	0.000351	0.000138	0.000305	0.000103
4: Others	0.000487	0.000219	0.000444	0.000266	0.000329	0.000196

Table 2: RMSE and MAE over years 2003-2012 for different time windows of estimation. Russia.

Cause	1983-2002		1988-2002		1993-2002	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
1: Alcohol	0.000245	0.000162	0.000236	0.000149	0.000197	0.000127
2: Circulatory	0.001405	0.000893	0.001207	0.000752	0.001953	0.000885
3: Transp. acc.	0.000055	0.000043	0.000056	0.000042	0.000044	0.000032
4: Cancer	0.000898	0.000571	0.000754	0.000479	0.000673	0.000380
5: Others	0.001514	0.000777	0.000964	0.000506	0.000811	0.000557

Application - Case Studies

Check model reliability: Lee-Carter comparison on a single CoD

Starting from age 5 up to 80, we focus the analysis on two relevant causes of death: mortality wholly attributable to alcohol in Russia and cancer in Spain. The models' performances are compared using RMSE and MAE.

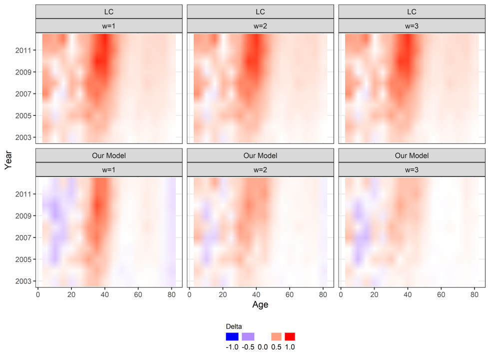
Country	Cause	Model	1983-2002		1988-2002		1993-2002	
			RMSE	MAE	RMSE	MAE	RMSE	MAE
Spain	Cancer	Our model	0.000684	0.000291	0.000400	0.000200	0.000260	0.000121
		Lee-Carter	0.000814	0.000456	0.000772	0.000428	0.000803	0.000445

Country	Cause	Model	1983-2002		1988-2002		1993-2002	
			RMSE	MAE	RMSE	MAE	RMSE	MAE
Russia	Alcohol	Our model	0.000245	0.000162	0.000236	0.000149	0.000197	0.000127
		Lee-Carter	0.000302	0.000170	0.000367	0.000204	0.000209	0.000134

To observe age-specific deviations we use relative differences of the central death rates between our model/Lee-Carter model and the observed values as

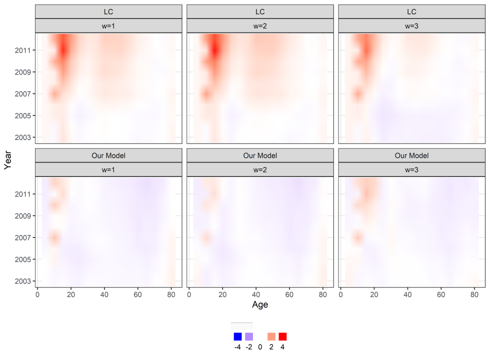
$$Delta_c(x, t) = \frac{\hat{m}_c(x, t) - m_c(x, t)}{m_c(x, t)}$$

Spain: Cancer



- Lung, colorectal, and smoking-related- cancers, also remain important causes of death in Spain in 2018
- We bring evidence of improvements for cancer mortality supported by the OECD report (the anti-tobacco law adopted in 2005 and reinforced in 2010.)
- Our model estimations improve as the time window progresses over time, which does not happen with the LC.

Russia: Alcohol



- We bring evidence of the sensitivity to the anti-alcohol campaign of 1985-1988 that showed a rapid increase in life expectancy in Russia
- After reaching lowest levels, life expectancy at birth increased rapidly between 1994 and 1998, and after 1998, began a new decline albeit not as fast as in the early 1990s.

Conclusion and limitation

Conclusion

- CoD information is very limited compared to overall mortality - ICDs structural breaks issue
- Our model offers the advantage of an indirect and complementary way to approximate death rates specific for age and causes
- Valuable in contexts where population studies are hindered by financial constraints, for national registry systems who do not support the open data system
- Indirect estimation offers a benchmark with which overall mortality can be compared to, thus dealing with very uncertain scenarios.

limitation

- Our model, as other indirect methods (Pascariu et al. (2020), Ševčíková et al. (2016)), leverages on the log-linear assumption, exploiting the regularities of age patterns of mortality that might not hold for some specific causes.
- On the other hand, we are focusing on ages 0-80, we overcome a past contributions that provided estimation to mid-life mortality only.

Thank you for your attention.

Questions / Suggestions ?