Modeling age-space mortality dynamics in small areas

Jacob Martin[1,2] and Carlo Giovanni Camarda[1]

[1]Institut national d'études démographiques
[2]Universidad del País Vasco/Euskal Herriko Unibertsitatea

SCOR chair on mortality research meeting, Paris

8 November 2024

Motivation

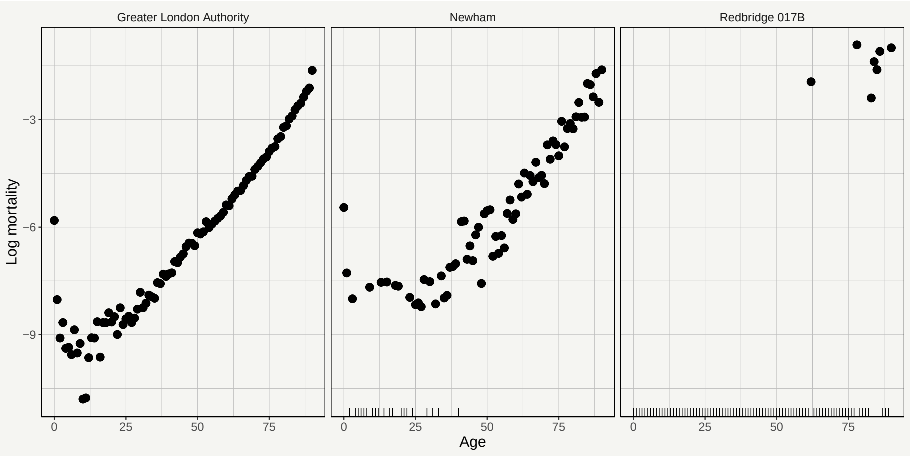


Figure: Observed mortality. Females, 2015. **Left:** London Metropolitan Area. **Center:** single London borough. **Right:** single Lower layer Super Output Area. *Note:* vertical bars at the bottom of the graphs indicate zero observed deaths.

Motivation

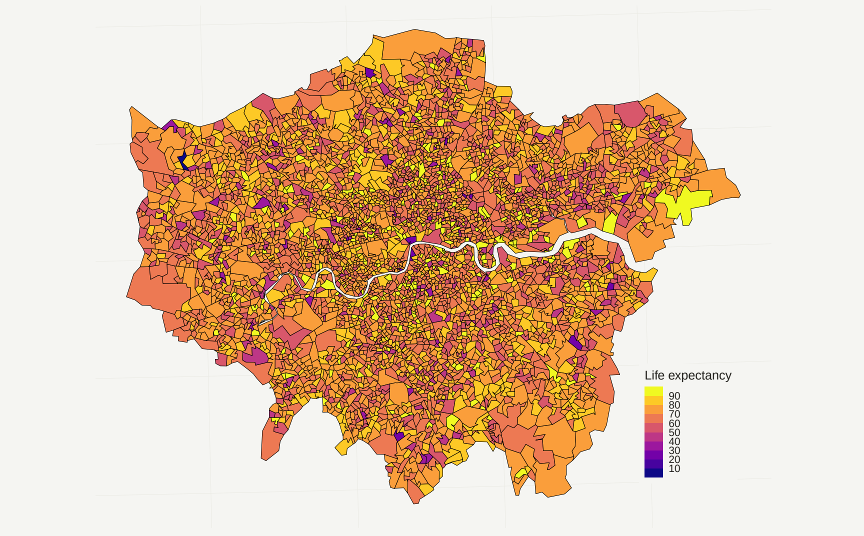


Figure: Life expectancy at birth by Lower layer Single Output Area calculated from observed mortality. Females, 2015.

The problem

- Model mortality in a very large number of very small areas.
- For example, 4835 sub-municipal small areas (Lower layer Super Output Areas (LS0As)) in a metropolitan area (London) and 91 age categories $\rightarrow$ 439.985 observations.

Understanding fine-grain inequalities in mortality

- Small populations lead to substantial fluctuations in observed death counts, creating very noisy mortality signal.
- At the same time, underlying mortality risk can differ substantially in space.
- Difficult to distinguish between real differences and random variation in risk of death.

Existing approaches

- Bayesian "**principal components**" or "**SVD** (Singular Value Decomposition)" model (Alexander, Zagheni, and Barbieri 2017):
  - Model area-specific mortality schedule as a linear combination of several principal components of a matrix of standard mortality schedules (e.g. mortality rates over age and time for the whole country).
  - Pool information with hierarchical structure (counties within a state in US) or across space (autoregressive prior).

Existing approaches

- Bayesian "**principal components**" or "**SVD** (Singular Value Decomposition)" model (Alexander, Zagheni, and Barbieri 2017):
    - Model area-specific mortality schedule as a linear combination of several principal components of a matrix of standard mortality schedules (e.g. mortality rates over age and time for the whole country).
    - Pool information with hierarchical structure (counties within a state in US) or across space (autoregressive prior).
- **TOPALS** model (Gonzaga and Schmertmann 2016; Schmertmann and Gonzaga 2018):
    - Model area-specific mortality schedule as piecewise-linear deviations from a standard schedule.
    - Difference penalty on coefficients of deviations to ensure stability of estimates.

Existing approaches

- Bayesian "**principal components**" or "**SVD** (Singular Value Decomposition)" model (Alexander, Zagheni, and Barbieri 2017):
  - Model area-specific mortality schedule as a linear combination of several principal components of a matrix of standard mortality schedules (e.g. mortality rates over age and time for the whole country).
  - Pool information with hierarchical structure (counties within a state in US) or across space (autoregressive prior).
- **TOPALS** model (Gonzaga and Schmertmann 2016; Schmertmann and Gonzaga 2018):
  - Model area-specific mortality schedule as piecewise-linear deviations from a standard schedule.
  - Difference penalty on coefficients of deviations to ensure stability of estimates.
- **D-Splines** (Schmertmann 2021):
  - Model area-specific mortality schedule as a linear combination of cubic splines, penalized with a custom-made penalty derived from a standard schedule of mortality.

Drawbacks of existing approaches

- Presence of a standard schedule that must be estimated from real data, but uncertainty surrounding the estimation and incorporation of the standard is not taken into account (Alexander, Zagheni, and Barbieri 2017; Schmertmann and Gonzaga 2018; Gonzaga and Schmertmann 2016).
- Spatial structure of data is not exploited (Schmertmann and Gonzaga 2018; Gonzaga and Schmertmann 2016; Schmertmann 2021).

The age-space model

Data requirements

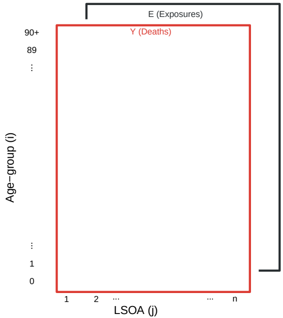
- Deaths and exposures for a single year for multiple spatial units.
- Enough deaths in the sum of spatial units to distinguish a mortality signal.

Model assumptions

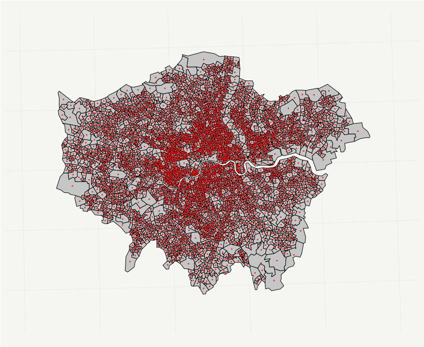
- Mortality in a single area:
  - resembles the mortality schedule of the total of all area being studied,
  - deviates from the standard schedule smoothly in age and space,
  - can show breaks from the overall smooth pattern in space.
- Standard schedule and deviations from standard are estimated simultaneously $\rightarrow$ uncertainty in standard is accounted for.
- Spatial structure exploited to borrow strength across areas.

Data structure

Figure: Data inputs



(a) Data structure.



(b) Centroids of GLA LSOAs.

- Data: two $m \times n$ matrices, m ages, n spatial units: deaths $\mathbf{Y}$, exposures $\mathbf{E}$. Centroids of territorial units serve as spatial information.

Methodology

- We model deaths in a Poisson setting:

$$\ln E[\mathbf{y}] = \ln(\mathbf{e}) + \ln(\boldsymbol{\mu}) = \ln(\mathbf{e}) + \boldsymbol{\eta}. \tag{1}$$

- Goal: model

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}^0 \\ \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_j \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix} = \mathbf{X}\boldsymbol{\theta} \text{ where } \begin{cases} \boldsymbol{\eta}^0 & \text{a common age schedule,} \\ \boldsymbol{\eta}_j & = \boldsymbol{\eta}^0 + \boldsymbol{\delta}_j + \gamma_j \\ & = \text{area } j \text{ schedule} \end{cases} \tag{2}$$

  - $\boldsymbol{\delta}_j$ deviations from standard that vary smoothly in age and space
  - γ_j area-specific intercepts that allow for unsmooth variation.

Building the model matrix

- Smooth standard over age $\rightarrow \mathbf{B}_a$, $m \times k_{\mathrm{a}}$, a rich cubic B-spline basis over age ($\lfloor m/5 \rfloor$ internal knots) adjusted for infant mortality (Camarda 2019).
- Smooth age-space deviations:
  - $\mathbf{B}_s$, $n \times k_{\mathrm{s}}$, basis over space:

$$\mathbf{B}_s = \mathbf{B}_{\mathrm{lat}} \square \mathbf{B}_{\mathrm{lon}} = (\mathbf{B}_{\mathrm{lat}} \otimes \mathbf{1}'_{k_{\mathrm{lon}}}) \odot (\mathbf{1}'_{k_{\mathrm{lat}}} \otimes \mathbf{B}_{\mathrm{lon}}), \tag{3}$$

    $\mathbf{B}_{\mathrm{lon}}$, $n \times k_{\mathrm{lon}}$ basis over longitude, $\mathbf{B}_{\mathrm{lat}}$, $n \times k_{\mathrm{lat}}$ basis over latitude.
  - $\breve{\mathbf{B}}_a$, $m \times \breve{k}_{\mathrm{a}}$ reduced basis over age ($\lfloor m/10 \rfloor$ internal knots).
  - $\mathbf{B}_s \otimes \breve{\mathbf{B}}_a$
- Unsmooth area specific intercepts $\rightarrow \mathbf{I}_n$ identity matrix.
- Model matrix: $\mathbf{1}_n \otimes \mathbf{B}_a : \mathbf{B}_s \otimes \breve{\mathbf{B}}_a : \mathbf{I}_n \otimes \mathbf{1}_m$

Model specification: ensuring identifiablility

- Add sum of areas to ensure convergence $\rightarrow n+1$ spatial units.
- Model matrix:

$$\mathbf{X}\,\theta = \left[\begin{array}{c|c|c} \mathbf{1}_{n+1} \otimes \mathbf{B}_a & \multicolumn{2}{c}{\mathbf{0}_{m,k_\mathrm{a}k_\mathrm{s}+n}} \\ \cline{2-3} & \mathbf{B}_s \otimes \breve{\mathbf{B}}_a & \mathbf{I}_n \otimes \mathbf{1}_m \end{array}\right] \boldsymbol{\theta}\,. \tag{4}$$

Penalization

- Second difference penalty on coefficients for $\boldsymbol{\eta}^0$ and $\boldsymbol{\delta}_j$s. Ridge penalty for γ_j.
- Four smoothing parameters (λ_a, λ_{lon}, λ_{lat}, $\breve{\lambda}_a$) and one ridge penalty (κ).

Model specification: penalty

- Let $\mathbf{D}_a$, $\mathbf{D}_{\text{lon}}$, $\mathbf{D}_{\text{lat}}$ $\breve{\mathbf{D}}_a$ be the difference penalties associated with the age standard, longitude, latitude, and the second age basis.
- Penalty matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_3 \end{bmatrix},$$

where

$$\mathbf{P}_1 = \lambda_a \mathbf{D}_a' \mathbf{D}_a,$$

$$\mathbf{P}_3 = \lambda_{\text{lon}} \mathbf{I}_{k_{\text{lat}}} \otimes \mathbf{D}_{\text{lon}}' \mathbf{D}_{\text{lon}} \otimes \mathbf{I}_{\breve{k}_{\text{a}}} + \lambda_{\text{lat}} \mathbf{D}_{\text{lat}}' \mathbf{D}_{\text{lat}} \otimes \mathbf{I}_{k_{\text{lon}}} \otimes \mathbf{I}_{\breve{k}_{\text{a}}} + \breve{\lambda}_a \mathbf{I}_{k_{\text{lat}}} \otimes \mathbf{I}_{k_{\text{lon}}} \otimes \breve{\mathbf{D}}_a' \breve{\mathbf{D}}_a,$$

$$\mathbf{P}_3 = \kappa \mathbf{I}_n.$$

Model specification and fitting

- Model fitting: Iteratively re-Weighted Least Squares (IWLS) using Generalized Linear Array Model (GLAM) (Currie, Durbán, and Eilers 2006) arithmetic:
$$(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{P})\hat{\boldsymbol{\theta}} = \mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}, \tag{5}$$
  - $\hat{\boldsymbol{\theta}}$ next iteration of coefficients to be estimated
  - $\mathbf{W}$ diagonal matrix of weights
  - $\mathbf{z}$ working dependent variable
  - Tilde (˜) current approximation.

Greater London Authority Data

Data

- $n = 4835$ Lower layer Super Output Areas (LS0As) in Greater London Authority.
- Deaths and mid-year population by $m = 91$ $(0, 1, ..., 90+)$ ages.
- Total female population between 298 and 4413.
- One year: 2015

Model fitting

- Smoothing parameters equal to 1, ridge penalty equal to 100.
- 7782 coefficients to estimate.
- Effective dimension of 408.

Estimated values of components

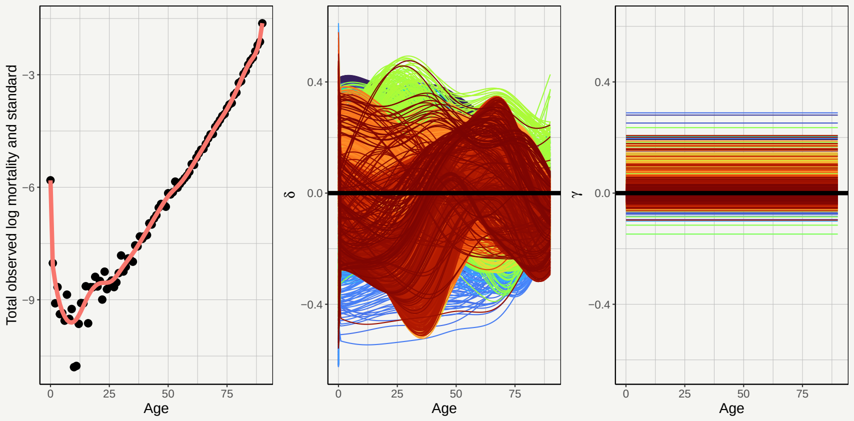


Figure: **Left:** estimated standard schedule and observed mortality of total GLA. **Center:** estimated values of δ_j. **Right:** estimated values of γ_j.

Mapping components

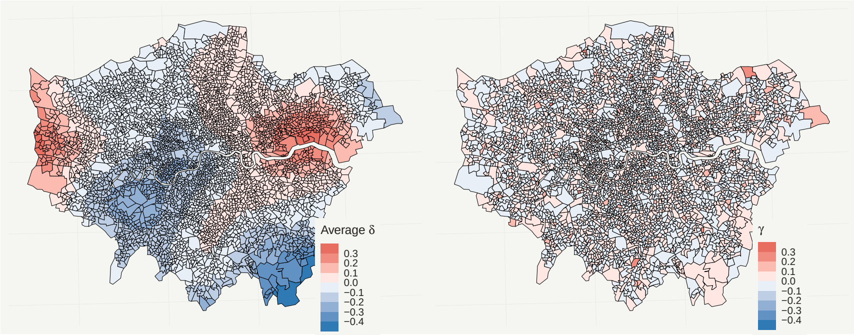


Figure: Average over age of $\boldsymbol{\delta}_j$ and γ_j.

Fitted values

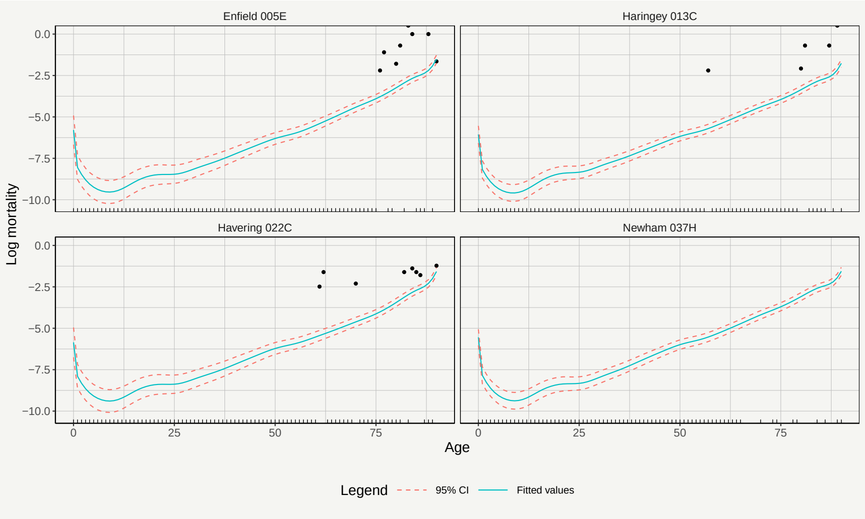


Figure: Fitted log-mortality schedule for four LSOAs. Females, 2015.

Estimating sub-municipal life-expectancy

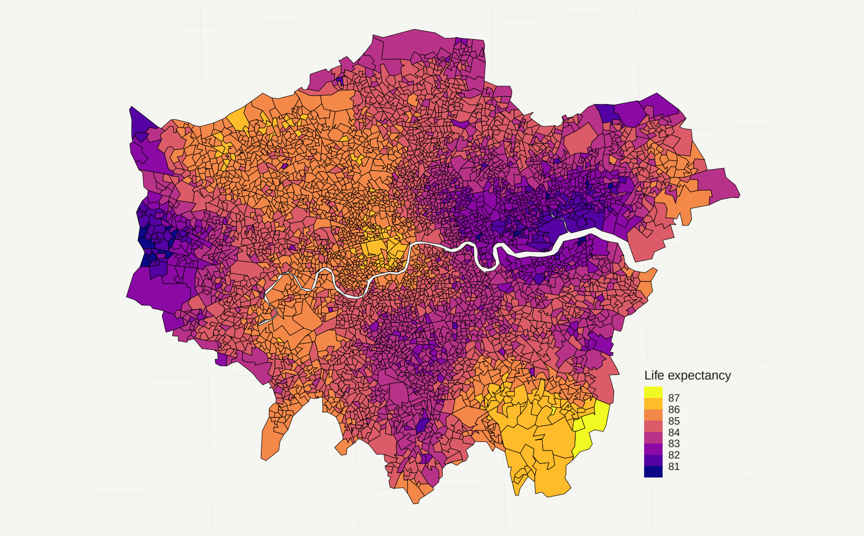


Figure: Estimated life expectancy at birth. Females, 2015.

Smoothing parameter selection

- Normally we would select the smoothing parameters through a grid search to minimize either Aikaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).
- However, BIC and AIC may be inadequate in the presence of under- or over-dispersion.
- We can correct AIC and BIC with ψ^2:

$$\psi^2 = \frac{\text{Dev}}{\text{df}},$$

where Dev is the deviance of the model and df is the number of observations with nonzero offset minus the effective dimension of the model.

Further extensions

- Accounting for under- or over-dispersion.
- Allowing the non-smooth effects to vary by age.
- Incorporating time to model multiple years of data.
- Incorporating hierarchical structure in the model.
- Adjusting the distributional assumption to model prevalence of health conditions to produce small-area health expectancy estimates.

Extension: incorporating time

- Now m ages, n spatial units, l years.
- Construct $\mathbf{B}_t$, $l \times k_{\mathrm{a}}$, basis over time.
- Possible model matrices (omitting zeros and sum for identifiablility):

$$\mathbf{B}_t \otimes \mathbf{1}_n \otimes \mathbf{B}_a : \mathbf{B}_t \otimes \mathbf{B}_s \otimes \breve{\mathbf{B}}_a : \mathbf{I}_l \otimes \mathbf{I}_n \otimes \mathbf{1}_m$$

or

$$\mathbf{B}_t \otimes \mathbf{1}_n \otimes \mathbf{B}_a : \mathbf{B}_t \otimes \mathbf{B}_s \otimes \breve{\mathbf{B}}_a : \mathbf{1}_l \otimes \mathbf{I}_n \otimes \mathbf{1}_m.$$

- Multiple years of data may give enough strength to be able to abandon the standard:

$$\mathbf{B}_t \otimes \mathbf{B}_s \otimes \mathbf{B}_a : \mathbf{I}_l \otimes \mathbf{I}_n \otimes \mathbf{1}_m.$$

Conclusion

- Our model borrows strength across age and space in order to produce mortality estimates in data-sparse contexts.
- Our modeling approach can be used to identify spatial patterns in mortality, as well as localized breaks from the general spatial pattern.
- Further work needed for selecting smoothing parameters and modeling multiple years of data.
- Contact: `jacob.martin@ined.fr`

References I

- Alexander, Monica, Emilio Zagheni, and Magali Barbieri (2017). "A Flexible Bayesian Model for Estimating Subnational Mortality". In: *Demography* 54.6, pp. 2025–2041. DOI: 10.1007/s13524-017-0618-7.
- Camarda, Carlo Giovanni (2019). "Smoothed constrained mortality forecasting". In: *Demographic Research* 41, pp. 1091–1130. DOI: 10.4054/DemRes.2019.41.38.
- Currie, Ian, María Durbán, and Paul Eilers (2006). "Generalized linear array models with applications to multidimensional smoothing". In: *Journal of the Royal Statistical Society Series B* 68, pp. 259–280.
- Gonzaga, Marcos and Carl Schmertmann (2016). "Estimating age- and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010". In: *Revista Brasileira De Estudos De População* 33.3, pp. 629–652. DOI: 10.20947/S0102-30982016c0009.

References II

- Rizzi, Siliva, Jutta Gampe, and Paul Eilers (2015). "Efficient Estimation of Smooth Distributions From Coarsely Grouped Data". In: *American Journal of Epidemiology* 182, pp. 138–147. DOI: 10.1093/aje/kwv020.
- Schmertmann, Carl (2021). "D-splines: Estimating rate schedules using high-dimensional splines with empirical demographic penalties". In: *Demographic Research* 44, pp. 1085–1114. DOI: 10.4054/DemRes.2021.44.45.
- Schmertmann, Carl and Marcos Gonzaga (2018). "Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas With Defective Vital Records". In: *Demography* 55.4, pp. 1363–1388. DOI: 10.1007/s13524-018-0695-2.