Supplementary Information

A Complete Estimates and Rankings

Table E1: Estimated life expectancy at birth for 402 German districts over the 2015–2017 period, in alphabetical order. Estimates are listed separately for women and men. We list the district's rank (1=highest, 402=lowest), the posterior median estimate, and the limits of 80% prediction intervals (i.e. 10th & 90th percentiles).

district	Women				Men			
	Rank	Percentiles			Rank	Percentiles		
		10%	50% (Median)	90%		10%	50% (Median)	90%
Please see now MS-Excel-File: `districtEstimatesLongTable.xlsx`								

B Overview: TOPALS Modeling

The basic building block of our estimation process is de Beer's TOPALS approach [1]. Here we briefly illustrate with an example the modified TOPALS model that we applied to estimate district mortality rates by age and sex. More details about this model are available in [2, 3].

Our goal is to estimate a smooth age pattern for mortality that is consistent with known regularities in vital rates, for small populations that may have zero deaths at young ages and considerable random fluctuations in deaths reaching into middle ages. The TOPALS approach models the local age schedule of mortality rates via a set of parametric deviations from a fixed, smooth schedule called the *standard*. As pointed out by de Beer and others [1, 2], TOPALS estimates for age-specific mortality are quite insensitive to the choice of the fixed standard, unlike the classic Brass logit relational model.

Log mortality at any age $x \in 0, 1, \ldots 89$ is modeled as a standard value, $\log \mu^*_x$, plus a linear spline function:

$$\log \hat{\mu}_x = \log \mu^*_x + B'_x \hat{\alpha} \tag{E1}$$

where B is a matrix whose columns are linear B-spline basis functions [4]. In our application B is 90×7; rows correspond to ages $0 \ldots 89$ and B'_x is the 1×7 row of B that corresponds to for age x.

Under the canonical assumption that death counts by age group g are independent Poisson variables [5], we sampled α from a log posterior density that includes a Poisson likelihood:

$$\log P(\alpha|D, N) = c + \sum_g \left(D_g \log \hat{\mu}_g - N_g \hat{\mu}_g \right) - Penalty(\alpha) \tag{E2}$$

where $\log \hat{\mu}_g$ values are age-group averages of the rates defined as in Equation E1, c is a constant that is identical for all values of α, and $Penalty(\alpha)$ is a penalty term based on the smoothing priors discussed in the main text of the paper.

We illustrate the model for males in Vorpommern-Rügen, a district in the northeastern corner of Germany. Observed age-group death rates for 2015–2017, $\log \frac{D_g}{N_g}$, are represented by horizontal bars

in Figure E1, Panel C. The numbers above each bar are the total male deaths in that age group in Vorpommern-Rügen over the three years of data collection.

Larger numbers of deaths in an age group generally mean more reliably estimated rates, in the sense that random fluctuations are less likely to have large effects on observed death/exposure ratios. Vorpommern-Rügen has slightly more inhabitants than an average German district, but nonetheless there were fewer than 10 male deaths in six of eighteen age groups during the three years of observation, and there were no deaths of 5–9 year-old boys.

For our standard schedule (the baseline $\log \mu_x^*$ against which local deviations are expressed) we used 2010–2013 mortality for all German males. We calculated these standard rates from the Human Mortality Database [6]. The dotted line in Figure E1, Panel C shows this reference schedule.

Panel A of Figure E1 illustrates the linear spline basis functions (i.e., the columns of B). Although cubic splines are more common, the TOPALS model uses linear spline offsets from the standard because they are sufficient to produce a smooth age schedule. Spline knots are located at ages 0, 1, 10, 20, 40, 70, and 90. (The number and location of knots is somewhat arbitrary. These ages work well because they are close to well-known turning points in the mortality schedule – infant mortality, lowest mortality, accident mortality, and so forth.)

The fifth linear spline function in Panel A is represented by a thicker line. Because this spline is non-zero only between ages 20 and 70, the fifth α coefficient affects the TOPALS estimates of log mortality rates only in that age range. At the knot location (age 40), all other splines are zero. This facilitates interpretation of the fitted spline function that we add to the standard, depicted in the middle Panel B. For example, estimated log mortality at age 40 (a knot location) is simply $\log \hat{\mu}_{40} = \log \mu_{40}^* + \hat{\alpha}_5$. $\hat{\alpha}_5 = .34$ indicates that mortality at age 40 is estimated to be approximately 34% higher in this district, compared to the standard rate. This coefficient also affects estimated mortality over the ages 20–70 at which the associated spline function in Panel A is non-zero, but those effects diminish to zero as one moves farther from age 40. The other estimated coefficients (represented by vertical lines in Panel B) have analogous effects on estimated mortality rates.

The pattern of positive and negative offsets from the standard, illustrated in Panel B, means that local data suggests that male mortality in Vorpommern-Rügen is *lower* than the standard at the lowest ages (below about 7) and at the highest ages (above about 78), but *higher* over most ages (particularly among middle-aged males). We selected Vorpommern-Rügen in part to illustrate the flexibility of the TOPALS approach, which can easily fit such crossovers from a reference mortality schedule.

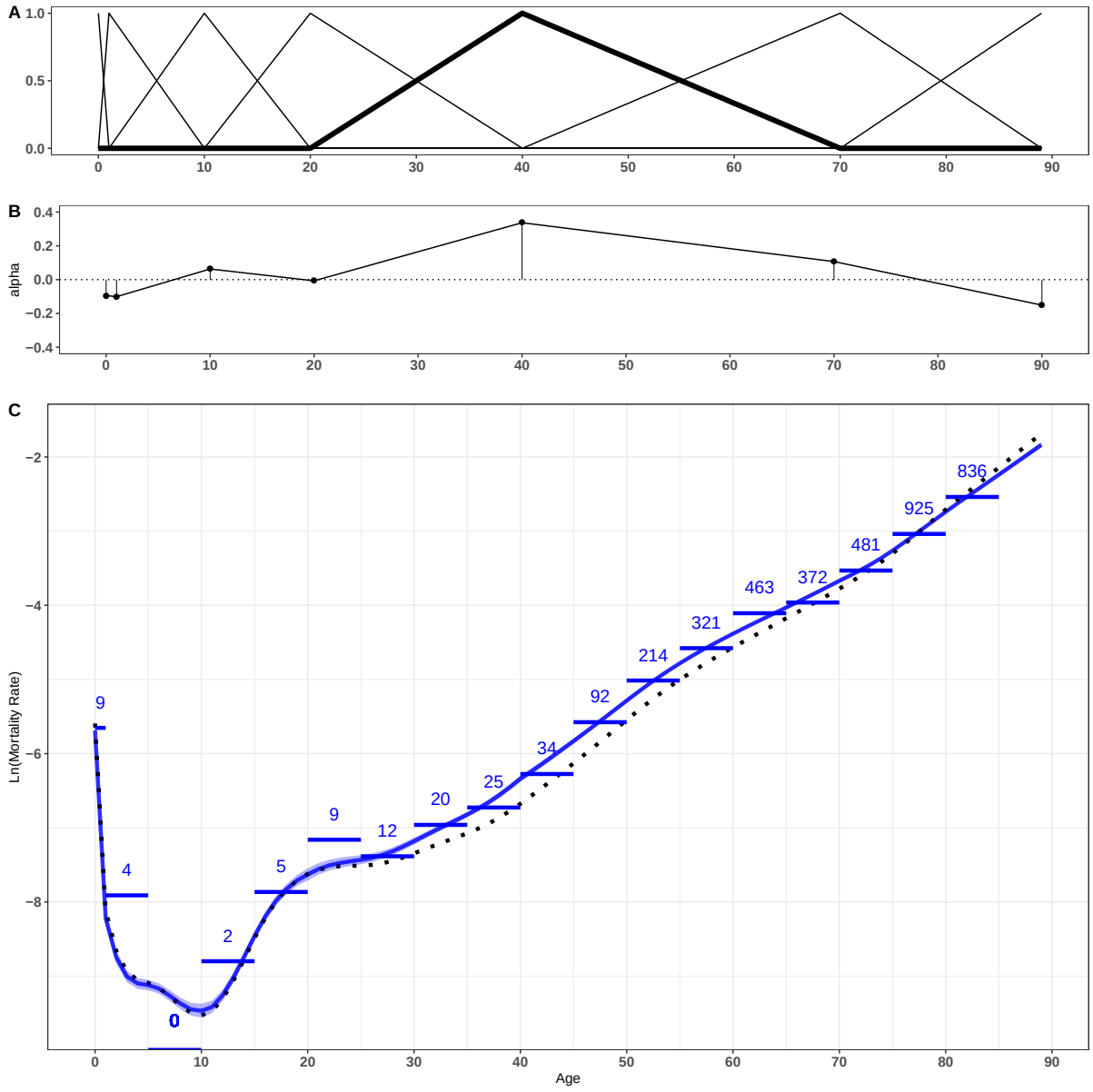


Figure E1: Panel A: B-spline basis functions (columns of matrix B);
Panel B: estimated linear spline function $B\,\hat{\alpha}$ added to the standard schedule;
Panel C: Standard mortality schedule (dotted curve), observed $\log(D_g/N_g)$ for age groups from 2015–2017 data (horizontal bars, with numbers of deaths in the age group recorded above each bar) and TOPALS-estimated mortality rates for Vorpommern-Rügen males at ages 0–89 (solid curve).

C Smoothing Priors for TOPALS Parameters

C.1 TOPALS priors

We simultaneously estimate mortality schedules for all districts and both sexes via Bayesian models that link local TOPALS parameters across small populations. Generally speaking, Bayesian models allow us to "borrow strength" by exploiting similarities between populations that are expressed via prior distributions [7]. Our model employs three distinct priors.

1) We use a hierarchical spatial prior for TOPALS parameters. Each of the seven local coefficients that determine a district's age- and sex-specific mortality rates shares a common national mean (NUTS-0), shares a common effect with all other districts in the same state (NUTS-1), and (only for states that have regional subdivisions, called *Regierungsbezirke*) shares a common effect with all other districts in the same substate region (*Regierungsbezirk*, NUTS-2). There are also district-specific (NUTS-3) effects for each parameter. This hierarchical spatial specification emphasizes likely similarities between districts by political geography. We consider this to be a sensible approach because policy (mental health facilities, hospital planning, etc.) at different administrative levels may influence health and mortality. We also experimented with an alternative spatial approach that emphasized spatial adjacency of districts rather than political boundaries. Both spatial approaches improve the precision of estimates, with slightly better results for the political hierarchy model. Here we present only results from the hierarchy model.

2) We smooth age schedules by using a normal prior distribution for the differences between TOPALS coefficients at consecutive age groups for each district and sex. This stabilizes estimated schedules in regions with extremely small populations [2].

3) We use another prior to regularize male-female differences in log mortality rates within districts. Sex differences by age are fairly consistent across similar populations. Adding this prior rewards estimates that conform more closely to expected sex differences, which permits us to "borrow strength" between male and female mortalilty information within each district. Because male-female mortality differences tend to be larger in Eastern Germany, we based our prior on the East German pattern of sex differences in life tables for the years 2010–2015 from the Human Mortality Database (HMD) for districts in the former GDR, and on the West German HMD pattern for districts located in the former West Germany. For Berlin we used the overall German pattern.

In the full Bayesian model we simultaneously estimate 5,628 parameters (2 sexes, 402 districts, 7 parameters per mortality schedule) by Markov Chain Monte Carlo (MCMC) sampling from the joint posterior distribution. Each draw from the posterior represents a possible map of mortality rates by sex, district, and age group, with combinations of parameters sampled in proportion to their probability. Thus more likely maps are sampled more often, and vice versa.

For each simulated map we use standard methods [8] to calculate male and female life expectancies (e_0) in each district, after extrapolating mortality for the highest ages (85–120) with the two-parameter "Kannisto-Model" approach that enforces a logistic shape, which is also employed for official estimates [9, 10, 11, 12].

The resulting distribution yields point estimates of district life expectancies (median e_0 values), and also produces estimates of uncertainty (expressed here as an 80% probability interval between the 10th and 90th percentiles of simulated e_0 values within each district).

C.2 TOPALS parameters

Parameters are α_{krs} for knot positions $k = 1 \dots K$, regions $r = 1 \dots R$, and sexes $s = \{M, F\}$. The TOPALS schedule of log mortality rates at ages $x \in \{0 \dots (A-1)\}$ in region r for sex s is

$$\lambda_{rs} = \lambda_s^* + \boldsymbol{B}\alpha_{rs}$$

where $\boldsymbol{B}$ is an $A \times K$ matrix of linear spline constants and $\alpha_{rs} \in \mathbb{R}^K$ is the vector of α elements $(\alpha_{1rs} \dots \alpha_{Krs})'$. In the analysis of German districts, $K = 7$, $A = 90$, and $R = 402$.

Arranging the α parameters into two $K \times R$ matrices,

$$\boldsymbol{\alpha}_s = \begin{bmatrix} \alpha_{11s} & \dots & \alpha_{1Rs} \\ \vdots & \ddots & \vdots \\ \alpha_{K1s} & \dots & \alpha_{KRs} \end{bmatrix} \quad s = M, F$$

the complete $A \times R$ matrix of (age,region)-specific log mortality rates for sex s is

$$\boldsymbol{\Lambda}_s = (1_R' \otimes \lambda_s^*) + \boldsymbol{B}\boldsymbol{\alpha}_s, \quad s = M, F$$

C.3 Spatial Model

The fundamental idea behind the spatial model is that there are likely to be regional patterns among the R TOPALS parameters for a given knot k and sex s. These sets of α parameters correspond to rows in the $K \times R$ $\boldsymbol{\alpha}_s$ matrix above:

$$\boldsymbol{\alpha}_s = \begin{bmatrix} \alpha_{1s}' \\ \vdots \\ \alpha_{Ks}' \end{bmatrix}$$

In the spatial hierarchical we assume that there are $Q \geq R$ *effects* that determine the α values for each (k, s) combination. The number of effects depends on the hierarchical map. For Germany there are $n_1 = 16$ first-level effects (one per state), $n_2 = 115$ second-level effects (86 for districts in states without *Regierungsbezirke*, and 29 for substate *Regierungsbezirke*), and $n_3 = 314$ third-level effects (for districts within *Regierungsbezirke*), so $Q = 16 + 115 + 314 = 445$ effects determine $R = 402$ district parameter values.

The parameters for knot k and sex s are

$$\alpha_{ks} = \mu_{ks} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \omega_1 \boldsymbol{H}_1 u_{(1)ks} + \omega_2 \boldsymbol{H}_2 u_{(2)ks} + \omega_3 \boldsymbol{H}_3 u_{(3)ks}$$

where μ_{ks} is the mean of α_{ks} over the entire map,$u_{(j)} \in \mathbb{R}^{n_j}$ are the hierarchical effects at level j, and $\boldsymbol{H_j}$ are $R \times n_j$ matrices of ones and zeroes that determine which effects correspond to each district at each hierarchical level.

C.3.1 Sum-to-Zero Constraints on Hierarchical Effects

Within each level of the hierarchy we enforce sum-to-zero constraints on subsets of u that have the same "parent". For example, in the German case we assume that the first-level effects (for the 16 states) must all sum to zero, that second-level effects (for *Regierungsbezirke* or districts within states) must sum to zero within each state, and that third-level effects (for districts within *Regierungsbezirke*) must sum to zero within each *Regierungsbezirk*.

Table E2: German Geographic Hierarchy

Level (j)	**Effects for...**	**Parent(s)**	**Dimensions of G_j $(q_j \times n_j)$**
1	State (16)	Germany (1)	1×16
2	District (86) *Regierungsbezirk* (29)	State without *Regierungsbezirke* (7) States with *Regierungsbezirke* (7)	14×115
3	District (314)	*Regierungsbezirk* (29)	29×314

We can enforce constraints by considering $q_j \times n_j$ grouping matrices $\boldsymbol{G}_j$ containing ones and zeroes such that the q_j constraints on sums within level j are

$$\boldsymbol{G}_j \, u_{(j)} = 0$$

where we temporarily drop the (k, s) subscripts for clarity, remembering that there are different hierarchical effects for each (k, s) combination.

Table E2 shows the structure of the administrative hierarchy for German districts, with types, numbers, and constraints on the effects at three levels. For example, at level $j = 1$ in the German hierachical model, all states have the same "parent", namely Germany. So there is only $q_1 = 1$ constraint on the $n_1 = 16$ state effects: their sum must be zero. Thus $\boldsymbol{G}_1$ is a 1×16 vector of 1s.

At level $j = 2$ in Germany the hierarchical constraints are more complex. The second level in the geographic hierarchy is either (1) *a district* (for seven states that do not have legal *Regierungsbezirke*; there are 86 such districts); (2) *a substate Regierungsbezirk* (for seven states that are legally subdivided into *Regierungsbezirke*; there are 29 such *Regierungsbezirke* in those seven states); or (3) *nothing* (for Hamburg and Berlin, which are both states and districts and have no second level). Thus for each of $q_2 = 14$ states we require that the sum of second-level effects within the state must sum to zero. For example, 18 second-level effects are for districts in (undistricted) Brandenburg state. So one row of $\boldsymbol{G}_2$ is a 1×115 row vector containing 18 ones (in columns corresponding to districts within Brandenburg) and 97 zeroes. Four second-level effects are for substate *Regierungsbezirke* in Niedersachsen, so another row of $\boldsymbol{G}_2$ has four ones (in columns corresponding to Niedersachsen *Regierungsbezirke*) and 111 zeroes. And so forth. The end result is a 14×115 matrix $\boldsymbol{G}_2$ for which $\boldsymbol{G}_2 u_{(2)} = 0$.

The third level in the German hierarchy consists entirely of districts within *Regierungsbezirke* in the seven districted states. There are $n_3 = 314$ such districts belonging to $q_3 = 29$ *Regierungsbezirke*, so the $\boldsymbol{G}_3$ matrix for Germany is 29×314.

C.3.2 Basis Functions for Constrained Effects

We can enforce the constraints on the hierarchical effects by building $n_j \times (n_j - q_j)$ basis functions Z_j such that

$$\underbrace{u_{(j)}}_{n_j \times 1} = \underbrace{\boldsymbol{Z}_j}_{n_j \times (n_j - q_j)} \underbrace{\varepsilon_j}_{(n_j - q_j) \times 1}$$

guarantees that $\boldsymbol{G}_j \, u_{(j)} = 0$ for any $(n_j - q_j) \times 1$ vector ε_j.

Matrices $\boldsymbol{Z}_j$ are defined as follows:

1. calculate $\boldsymbol{M}_j = \boldsymbol{G}_j' \boldsymbol{G}_j$; $\boldsymbol{M}_j$ is $n_j \times n_j$, with rank q_j.

2. perform an eigendecomposition $\boldsymbol{M}_j = \boldsymbol{X}_j[\boldsymbol{D}_j]\boldsymbol{X}_j' + \boldsymbol{Z}_j[\boldsymbol{0}]\boldsymbol{Z}_j'$, where $\boldsymbol{X}_j$ is the $n_j \times q_j$ matrix of eigenvectors corresponding to non-zero eigenvalues, and $\boldsymbol{Z}_j$ is the $(n_j \times (n_j - q_j)$ matrix of eigenvectors corresponding to zero eigenvalues.

3. Use $\boldsymbol{Z}_j$ as the basis for $u_{(j)}$.

Substituting the "deep" parameters ε into the hierarchical specification produces

$$\alpha_{ks} = \mu_{ks}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix} + \omega_1 \boldsymbol{H}_1\boldsymbol{Z}_1\varepsilon_{(1)ks} + \omega_2\boldsymbol{H}_2\boldsymbol{Z}_2\varepsilon_{(2)ks} + \omega_3\boldsymbol{H}_3\boldsymbol{Z}_3\varepsilon_{(3)ks}$$

Defining $\boldsymbol{S}_j = (\boldsymbol{H}_j\boldsymbol{Z}_j)'$, transposing, and stacking over parameters k then produces

$$\boldsymbol{\alpha}_s = (1_R' \otimes \mu_s) + \omega_1\boldsymbol{E}_{(1)s}\boldsymbol{S}_1 + \omega_2\boldsymbol{E}_{(2)s}\boldsymbol{S}_2 + \omega_3\boldsymbol{E}_{(3)s}\boldsymbol{S}_3$$

where $\boldsymbol{E}_{(j)s}$ is a $K \times (n_j - q_j)$ matrix of iid N(0,1) variables that represent the "deep effects" for level j.

C.3.3 Small Example

As a simple example, suppose that we had the hierarchy depicted in Figure E2, in which $R = 7$ districts are represented by solid dots. This hierarchy has $n_1 = 3$ first-level effects, $n_2 = 4$ second-level effects, and $n_3 = 4$ third-level effects, with

$$\boldsymbol{H}_1 = \begin{pmatrix}1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\\0&0&1\\0&0&1\end{pmatrix}, \quad \boldsymbol{H}_2 = \begin{pmatrix}0&0&0&0\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\end{pmatrix}, \quad \boldsymbol{H}_3 = \begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$\boldsymbol{G}_1 = \begin{pmatrix}1&1&1\end{pmatrix}, \quad \boldsymbol{G}_2 = \begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix}, \quad \boldsymbol{G}_3 = \begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix}$$

$$\boldsymbol{Z}_1 = \begin{pmatrix}.82&0\\-.41&-.71\\-.41&.71\end{pmatrix}, \quad \boldsymbol{Z}_2 = \begin{pmatrix}0&.71\\0&-.71\\.71&0\\-.71&0\end{pmatrix}, \quad \boldsymbol{Z}_3 = \begin{pmatrix}0&.71\\0&-.71\\.71&0\\-.71&0\end{pmatrix}$$

After constructing $\boldsymbol{S}_j$ matrices the hierarchical model for this simple example is

$$\begin{aligned}\boldsymbol{\alpha}_s = 1_7' \otimes \begin{pmatrix}\mu_{1s}\\ \vdots \\ \mu_{Ks}\end{pmatrix} &+ \omega_1\boldsymbol{E}_{1s}\begin{pmatrix}.82&-.41&-.41&-.41&-.41&-.41&-.41\\0&-.71&-.71&.71&.71&.71&.71\end{pmatrix}\\ &+ \omega_2\boldsymbol{E}_{2s}\begin{pmatrix}0&0&0&0&.71&.71&-.71&-.71\\0&.71&-.71&0&0&0&0&0\end{pmatrix}\\ &+ \omega_3\boldsymbol{E}_{3s}\begin{pmatrix}0&0&0&0&0&0&.71&-.71\\0&0&0&.71&-.71&0&0&0\end{pmatrix}\end{aligned}$$

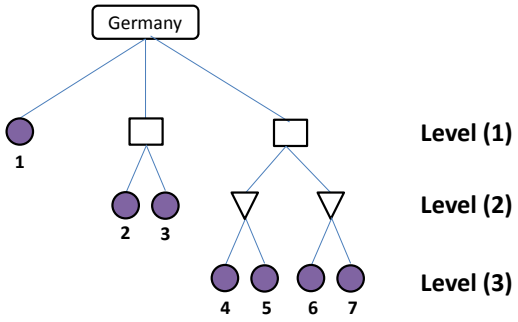
C.4 Smoothing Priors

The smoothing prior says that within each (r, s) combination the differences between offsets at consecutive knots should be small. In a "big matrix" approach we can construct two large $(K-1) \times R$ matrices for each sex

$$\boldsymbol{U}_s = \begin{bmatrix}\alpha_{2,1,s} & \cdots & \alpha_{2,R,s}\\ \vdots & \ddots & \vdots \\ \alpha_{K,1,s} & \cdots & \alpha_{K,R,s}\end{bmatrix} \qquad \boldsymbol{L}_s = \begin{bmatrix}\alpha_{1,1,s} & \cdots & \alpha_{1,R,s}\\ \vdots & \ddots & \vdots \\ \alpha_{K-1,1,s} & \cdots & \alpha_{K-1,R,s}\end{bmatrix}$$

Our smoothing prior is that, for both sexes $s = M, F$, each element of $\boldsymbol{U}_s - \boldsymbol{L}_s$ should be independently distributed $N(0, \text{sd} = \sqrt{\frac{1}{2}})$.

Figure E2: A simple hierarchy



C.5 Sex-Difference Priors

Sex-difference priors differ from smoothing and spatial priors, because they are about similarities in log mortality rates (λ) rather than similarities in the underlying parameters (α). The fundamental idea of the sex-difference priors for German data is as follows. We first assign each region $r = 1 \ldots 402$ to one of three zones: Western Germany (W), Eastern Germany (E), or the Berlin area (B). Use the notation $z(r) \in \{W, E, B\}$ to denote the zone for region r. Within each zone we used aggregated empirical data to define a typical pattern of age-specific sex differences, with δ^*_{xz} representing the typical male disadvantage (male log rate - female log rate) at age x in zone z. That is, *a priori* we expect $(\lambda_{x,r,M} - \lambda_{x,r,F})$ to be similar to $\delta^*_{x,z(r)}$.

In a "big matrix" approach the estimated sex differences are $\mathbf{\Lambda_M} - \mathbf{\Lambda_F}$, which we expect *a priori* to be

$$\mathbf{\Delta}^* = \begin{bmatrix} \delta^*_{0,z(1)} & \delta^*_{0,z(2)} & \cdots & \delta^*_{0,z(R)} \\ \vdots & \vdots & \ddots & \vdots \\ \delta^*_{89,z(1)} & \delta^*_{89,z(2)} & \cdots & \delta^*_{89,z(R)} \end{bmatrix}$$

The sex-difference prior is that each element of $(\mathbf{\Lambda}_M - \mathbf{\Lambda}_F) - \mathbf{\Delta}^*$ should be independently distributed $N(0, \text{sd} = \sigma_{sex})$.

References

[1] de Beer JAA. Smoothing and projecting age-specific probabilities of death by TOPALS. Demogr Res. 2012;27(20):543–592.

[2] Gonzaga MR, Schmertmann CP. Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. Revista Brasileira de Estudos de População. 2016;33(3):629–652.

[3] Schmertmann CP, Gonzaga MR. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. Demography. 2018;55(4):1363–1388.

[4] de Boor C. A Practical Guide to Splines. Heidelberg, New York: Springer-Verlag; 1978.

[5] Brillinger DR. The Natural Variability of Vital Rates and Associated Statistics. Biometrics. 1986;42:693–734.

[6] University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock, (Germany). Human Mortality Database; 2017. Available at www.mortality.org.

[7] Congdon P. Life expectancies for small areas: a Bayesian random effects methodology. Int Stat Rev. 2009;77(2):222–240.

[8] Preston SH, Heuveline P, Guillot M. Demography. Measuring and Modeling Population Processes. Oxford: Blackwell Publishers; 2001.

[9] zur Nieden F, Rau R, Luy M. Allgemeine Sterbetafel 2010/12 — Neue Ansätze zur Glättung und Extrapolation der Sterbewahrscheinlichkeiten. Wirtschaft und Statistik. 2016;(1):63–74.

[10] DAV-Unterarbeitsgruppe Rentnersterblichkeit. Herleitung der DAV-Sterbetafel 2004 R für Rentenversicherungen. Blätter der DGVFM. 2005;27(2):199–313.

[11] Thatcher RA, Kannisto V, Vaupel JW. The force of mortality at ages 80 to 120. Monographs on Population Aging, 3. Odense: Odense University Press; 1998.

[12] Thatcher AR. The long-term pattern of adult mortality and the highest attained age. J R Stat Soc A Stat. 1999;162(1):5–43.