

# Estimating age- and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010\*

Marcos Roberto Gonzaga\*\*  
Carl Paul Schmertmann\*\*\*

**High variability** in recorded vital events creates serious problems for small-area mortality estimation by age and sex. Many existing approaches to fitting local mortality schedules, including those most often used in Brazil, estimate rates by making rigid mathematical assumptions about local age patterns. Such methods assume that all areas within a larger area (for example, microregions within a mesoregion) have identically-shaped log mortality schedules by age. We propose a more flexible statistical estimation method that combines Poisson regression with the TOPALS relational model (DE BEER, 2012). We use the new method to estimate age-specific mortality rates in Brazilian small areas (states, mesoregions, microregions, and municipalities) in 2010. Results for Minas Gerais show notable differences in the age patterns of mortality between adjacent small areas, demonstrating the advantages of using a flexible functional form in regression models.

**Keywords:** Mortality. Small areas. TOPALS method. Poisson regression.

---

\* This research was supported by Capes Foundation, Ministry of Education of Brazil. Support from the Research Projects 470866/2014-4 (Estimativas de mortalidade e construção de tabelas de vida para pequenas áreas no Brasil, 1980 a 2010 – MCTI/CNPQ/MEC/Capes/Ciências Sociais Aplicadas) and 454223/2014-5 (Estimativas de mortalidade e construção de tabelas de vida para pequenas áreas no Brasil, 1980 a 2010/MCTI/CNPQ/Universal 14/2014) are gratefully acknowledged by Marcos R. Gonzaga.

\*\* Universidade Federal do Rio Grande do Norte (UFRN), Natal-RN, Brazil (mracs.roberto@gmail.com).

\*\*\* Florida State University, Tallahassee, FL, United States of America (schmert@fsu.edu).

## Introduction

### *Approaches to small-area mortality estimation*

Reliable estimates of levels, age patterns, and sex differences in small-area mortality are important for evaluating and targeting public investment. However, the combination of low risk and small populations can make small-area estimation a difficult task. Under those circumstances, observed event/exposure rates are often unstable across ages and times, and estimation of underlying mortality patterns is difficult (RIGGAN et al., 1991; BERNADINELLI; MONTOMOLI, 1992; PLETCHER, 1999; THATCHER et al., 2002; ASSUNÇÃO et al., 2005; DIVINO; EGIDI; SALVATORE, 2009).

For countries with good vital registration systems and annual population updates, researchers have recently made important advances in statistical modeling and estimation of complete mortality schedules in small areas (TOSON; BAKER, 2003; EAYRES; WILLIAMS, 2004; BRAVO; MALTA, 2010). The most sophisticated new approaches use Bayesian models to estimate small-area mortality rates and life expectancies (and their uncertainty) by “borrowing strength” over ages, sexes, times, and/or places. Demographers have played a role in developing Bayesian models (MCKINNON, 2010; TSIMBOS et al., 2014; ALEXANDER et al., 2016), but much of the progress in estimating small-area mortality schedules has been made by statistical epidemiologists (CONGDON, 2009; OCAÑA-RIOLA; MAYORAL-CORTÉS, 2010; JONKER et al., 2012; STEPHENS et al., 2013).

Studies of complete age- and sex-specific mortality schedules at subnational levels are rare in developing countries. There is a voluminous and well-known demographic literature on estimating *partial* mortality schedules (especially infant and child mortality indicators) in developing countries. Many classic methods rely on indirect information from surveys or censuses, rather than direct information from vital registration (e.g., BRASS; COALE, 1968; SULLIVAN, 1972; FEENEY, 1980; UNITED NATIONS, 1983; HILL, 1991; MOULTRIE et al., 2013). When applied to census data, classic indirect methods and their modern variants can produce useful child mortality indicators for small subnational areas e.g., Rajaratnam et al. (2010) for Mexico; UNDP (2013) for Brazil.

Researchers have also explored new methods for estimating small-area mortality and life expectancy for Brazil. McKinnon (2010) used Bayesian spatial smoothing to produce municipal-level estimates of child mortality ( ${}_5q_0$ ) from the 2000 census. Freire et al. (2015) combined formal demographic methods for estimating vital registration coverage with Empirical Bayes spatial smoothing. They estimated complete life tables by sex at the municipal level for 2010 from vital registration data, using indirect standardization of municipal rates based on mesoregional schedules. Lima et al. (2016) recently experimented with a variety of Poisson regression models, some with intra-state spatial smoothing of mortality rates, for estimating municipal-level life expectancy from vital registration data in the states of São Paulo and Maranhão.

Despite these advances, accurate estimation of age-specific rates for subnational populations in developing countries remains a challenge for demographers and public health researchers (HORTA et al., 1998; JUSTINO; FREIRE; LUCIO, 2012; UNDP, 2013; LIMA; QUEIROZ; SAWYER, 2014; FREIRE et al., 2014; LIMA; QUEIROZ, 2014). In particular, there is still value in developing robust estimation methods for complete mortality schedules, which could then serve as building blocks for more complex models.

### *Indirect standardization for sparse data*

Indirect standardization (IS) is a common method for dealing with sparse data for small populations. IS assumes a specific pattern of relative mortality by age, and estimates only the *level* of local mortality from local, age-specific death and exposure data. The assumed age pattern for a small area is usually the empirical pattern observed in a larger geographic aggregate, such as the state to which the small area belongs.

Researchers in Brazil and elsewhere have used IS in two different ways, depending on the quality of vital records. For small areas in regions with reliable death registration, researchers have applied IS smoothing to estimate mortality rates directly (CURTIN; KLEIN, 1995; UNDP, 2013; RAM et al., 2015). Death registration is incomplete in many regions of Brazil, however (VASCONCELOS, 1998; PAES; ALBUQUERQUE, 1999; SZWARCOWALD et al., 2002; MELO-JORGE; LAURENTI; GOTLIEB, 2007; LIMA; QUEIROZ, 2011; JUSTINO; FREIRE; LUCIO, 2012; FRIAS et al., 2013; LIMA; QUEIROZ; SAWYER, 2014; LIMA; QUEIROZ, 2014). In regions with problematic vital records, IS smoothing produces estimates of (uncorrected) mortality schedules, which then serve as inputs to procedures that estimate the degree and age pattern of under-registration (e.g., HORTA et al., 1998; UNDP, 2013; QUEIROZ et al., 2013; LIMA QUEIROZ; SAWYER, 2014; FREIRE et al., 2014; FREIRE et al., 2015). In both cases, the quality of mortality estimates depends on the fundamental IS assumption – namely, that a specific pattern of relative mortality rates by age is correct. In most applications of IS to small areas, this requires that mortality age patterns must be identical in a larger region (such as a state) and in each of its component subregions (such as microregions or municipalities).

In this paper we propose an alternative to IS that allows estimation of small-area schedules without imposing strong assumptions about the age pattern of mortality rates. Like IS, the new method can be used two ways: (1) to estimate complete schedule of log mortality rates in areas where the vital registration coverage is complete, or (2) to smooth reported death rates in areas with defective vital registration before applying undercount adjustments.

### *Poisson regression with a TOPALS relational model*

We propose a Poisson regression method based on TOPALS, a relational model developed by De Beer (2012) for smoothing and projecting age-specific probabilities of death. Like all relational models (e.g. BRASS, 1971; MURRAY et al., 2003; WILMOTH

et al., 2012), TOPALS builds complete schedules of age-specific rates via mathematical adjustments to a specified standard schedule. Our version of TOPALS constructs a fitted schedule of log mortality rates at ages 0...99 by adding a linear spline function with seven parameters ( $\alpha_0 \dots \alpha_1$ ) to a pre-specified standard schedule. We estimate parameters by maximizing a penalized Poisson likelihood function for age-specific deaths, conditional on age-specific exposure.

To our knowledge, demographers have not previously used TOPALS in regression models. Although relational models can be sensitive to the choice of the baseline schedule, we demonstrate that TOPALS regression results are extremely insensitive to the choice of the standard.

As a relational model with multiple, age-specific parameters, TOPALS is flexible enough to “bend” a standard schedule into a variety of alternative shapes. This flexibility can be important when age-specific mortality patterns vary across small areas within a larger region.

Mortality patterns differ significantly across large regions in Brazil (AGOSTINHO, 2009), but it is less clear what happens in small areas. Even adjacent regions can vary in the stage of epidemiological transition, and in levels of urbanization, development, and public investment. Consequently, they can also differ in age-specific mortality by cause of death, and in age patterns of all-cause mortality (PRATA, 1992; BARRETO et al., 1993; SCHRAMM et al., 2004; BARRETO; CARMO, 2007; ARAÚJO, 2012). In these circumstances, applying IS to estimate mortality rates before applying undercount adjustments could hide important differences between areas’ underlying mortality patterns – especially for males and for some specific causes of death.

The method that we present in this paper, TOPALS regression, offers at least three advantages over IS. First, rather than using multi-year age groups and a maximum age of 60 or 70, the new approach uses data for single years of age 0, 1, 2, ..., 99, even when the corresponding risk populations are very small. Second, Poisson regression allows appropriate use of the “zero” cells for specific ages and age groups in which there are no recorded deaths. Finally, and most importantly, our method does not assume a fixed age pattern of mortality rates in advance. Based on the empirical tests that we report in the paper and in the supplementary material on our project website, we conclude that the TOPALS method works well, even for areas with small populations and zero deaths at many ages.

In the text of this paper, we present only regression results for small areas of Brazil in which vital registration is nearly complete. In these cases TOPALS estimates may be interpreted directly as estimates of local mortality rates. On a supplementary website<sup>1</sup> we report regression results for all small areas of Brazil (all states, mesoregions, microregions, and municipalities). We strongly caution readers that many regression estimates on the website correspond to areas with incomplete vital registration. In such areas, TOPALS estimates, like IS estimates, still require adjustment for under-reporting of deaths.

<sup>1</sup> Available at <<http://topals-mortality.schmert.net>>.

## Data and methods

### Data

Population and deaths by sex and single-year ages, from 0 to 99, come from the Demographic Census (2010) and from the Ministry of Health's Mortality Information System (SIM/Datasus), respectively. Both types of data were collected at municipal level and then aggregated to microregions, mesoregions, and state levels, as defined by the Brazilian National Statistics Office (IBGE).

We used IBGE's automated retrieval system (SIDRA)<sup>2</sup> and *TabWin* software available on the Datasus website<sup>3</sup> to collect data on population and deaths, respectively, for 5565 municipalities, 100 single-year ages, and 2 sexes. We recorded the 2010 census population, the number of deaths over calendar years 2009-2011, and geographic identifiers for each of the 1,113,000 combinations of (municipality, age, sex). We used the 2010 census populations to estimate age- and sex-specific exposure over 2009-2011.<sup>4</sup> Despite using three years of exposure, almost half (49.2%) of the (municipality, age, sex) cells have no recorded deaths. The complete dataset is available for other researchers in the supplementary material on our project website.

### Poisson regression with TOPALS

Several statistical models smooth mortality rates or probabilities using spline functions (MACNAB; DEAN, 2001; CURRIE; DURBAN; EILERS, 2004; DE BEER, 2012; CAMARDA, 2012). De Beer (2012)'s TOPALS model uses a linear spline to model the pattern of ratios between age-specific probabilities of death in a population and the corresponding probabilities in a standard schedule.

Our variant of TOPALS also uses a standard schedule and a linear spline offset. However, we use the spline to represent additive offsets on the log mortality rate scale, rather than multiplicative offsets on the probability scale. Specifically, we assume that the schedule of log mortality rates over ages 0...99 is the sum of a standard schedule  $\lambda^* \in \mathbb{R}^{100}$  and a linear spline function:

$$\lambda(\alpha) = \lambda^* + B \alpha \quad (1)$$

$100 \times 1$     $100 \times 1$     $100 \times 7$     $7 \times 1$

where  $\lambda$  is a vector of log rates in a small area;  $\lambda^*$  is a standard schedule,<sup>5</sup>  $B$  is a matrix of constants in which each column is a linear B-spline basis function (DE BOOR, 1978;

<sup>2</sup> <<http://www.sidra.ibge.gov.br>>.

<sup>3</sup> <<http://www.datasus.gov.br>>.

<sup>4</sup> Details of this method are included on our project website. For all but the smallest areas, estimated exposure for a given age and sex is very close to three times the 2010 population at that age and sex.

<sup>5</sup> The vector of mean sex- and age-specific log rates in the Human Mortality Database (HMD, 2015) over all schedules observed in any country after 1969.

EILERS; MARX, 1996),<sup>6</sup> and  $\alpha$  is a vector of parameters representing offsets to the standard schedule. We define knots at ages  $t_0, \dots, t_6 = (0, 1, 10, 20, 40, 70, 100)$ . For ages  $x$  in  $\{0, 1, 2, \dots, 99\}$  and columns  $k$  in  $\{0, \dots, 6\}$  the basis functions in  $B$  are:

$$B_{x,k} = \begin{cases} \frac{x - t_{k-1}}{t_k - t_{k-1}} & \text{if } x \in [t_{k-1}, t_k] \\ \frac{t_{k+1} - x}{t_{k+1} - t_k} & \text{if } x \in [t_k, t_{k+1}] \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

With this parameterization,  $\alpha$  values represent additive offsets ( $\lambda_x - \lambda_x^*$ ) to the log rate schedule at exact ages (0, 1, 10, 20, 40, 70, 100) and offsets change linearly with age between those knots.

For any set of observed age-specific deaths and populations  $\{D_x, N_x\}_{x=0, \dots, 99}$ , we assume that deaths are distributed as independent Poisson variables,  $D_x \sim \text{Pois}[N_x \exp(\lambda_x)]$ , so that the log likelihood is:

$$\log L(\alpha) = \text{constant} + \sum_x [D_x \lambda_x(\alpha) - N_x \exp(\lambda_x(\alpha))] \quad (3)$$

In order to avoid implausible fitted schedules for very small populations with very low numbers of deaths, we add a penalty term to the log likelihood that increases as the linear spline offsets become less smooth (EILERS; MARX, 1996). After omitting the constant (which does not affect the fit), the penalized log likelihood is:<sup>7</sup>

$$Q(\alpha) = \sum_x [D_x \lambda_x(\alpha) - N_x \exp(\lambda_x(\alpha))] - \sum_{k=0}^5 (\alpha_{k+1} - \alpha_k)^2 \quad (4)$$

The penalty term has virtually no effect on the fit for areas with moderate to large numbers of deaths  $D_x$  and exposure  $N_x$ , but it stabilizes estimates in the smallest municipalities.

The estimated mortality schedule for an area is  $\lambda^* + B\hat{\alpha}$ , where  $\hat{\alpha}$  is the value that maximizes the penalized likelihood  $Q$ . This objective function is non-linear in  $\alpha$ , but standard software (such as R's *optim* function or the *solver* add-in in Microsoft Excel) can easily find the offsets  $\alpha$  that maximize the penalized log likelihood.<sup>8</sup>

Graph 1 provides an example regression fit, for females in the Pará de Minas microregion (IBGE code 31029), which is part of the Belo Horizonte metropolitan mesoregion (code 3107). This microregion had a 2010 female population of 62,248, with 891 female deaths recorded over 2009-2011. The graph shows the logarithms of deaths/(estimated exposure) at each single year of age as circles, and illustrates the fit to the  $(D_x, N_x)$  data with offsets

<sup>6</sup> B-splines are only one of many possible methods for constructing linear spline basis functions. De Beer (2012) uses a different parameterization for linear splines, for example, but the columns spaces of his  $B$  matrix and ours are identical. B-splines have an advantage in regression problems, because the columns of the basis function are less collinear and parameter estimators have lower covariances. B-spline coefficients are also especially easy to interpret, because they represent the values of the spline at the knots.

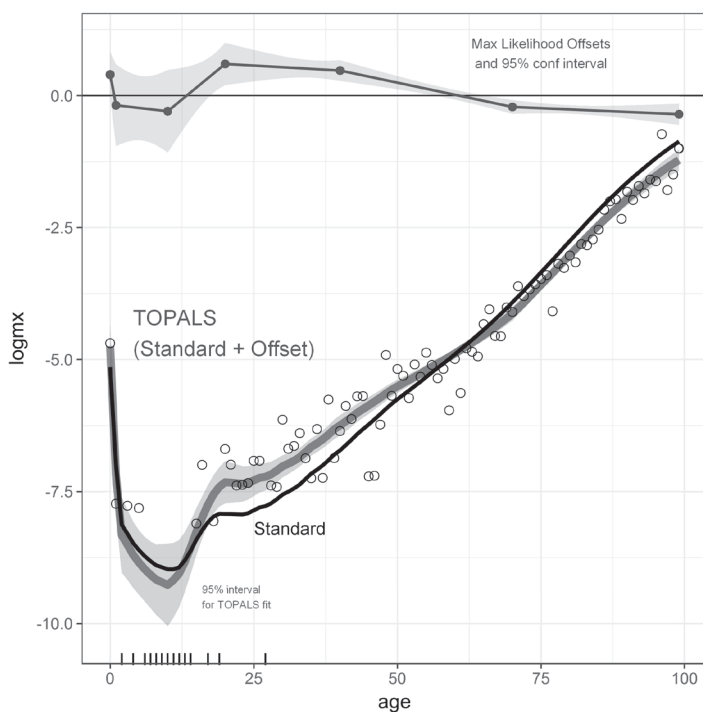
<sup>7</sup> The penalty term has a Bayesian interpretation:  $Q(\alpha)$  is the posterior log likelihood that arises when using an improper prior distribution  $(\alpha_{k+1} - \alpha_k) \sim N(0, 1/2)$  with the Poisson likelihood. The penalty term is also related to the IS estimator  $\lambda_x = \lambda_x^* + y \nabla x$ . In this second sense, IS is a special case of TOPALS, in which the penalty for differences in consecutive  $\alpha$  terms is infinite and all age-specific offsets  $\alpha_0 = \alpha_1 = \dots = \alpha_6 (= \gamma)$  must be identical.

<sup>8</sup> In the supplementary material for this paper (<http://topals-mortality.schmert.net>) we include an example spreadsheet, Pará de Minas.xlsx, that illustrates how to use Excel's *solver*. The R programs on the site use the *optim* function.

$\alpha$  that maximize the likelihood function in Equation (4). Several features of the plot merit attention:

- Regression estimates produce the smooth schedule ( $\lambda^* + B\alpha$ ) represented by the thick line. This fitted schedule has a plausible age pattern because it is based on an observed standard, but it is appropriately adjusted to local levels and age patterns of mortality via the choice of offsets  $\alpha$ .
- A non-linear Poisson regression approach allows estimates of the entire schedule of single-year rates, even when there are no deaths at some ages. For Pará de Minas females 2009-2011, fourteen single-year ages had  $D_x=0$  (see figure caption), and there was one five-year group (10-14 year olds) with no recorded deaths.
- Because the regression method borrows information from both the standard and the local rates at younger ages, it is possible to extend rate estimates for small areas to single-year ages above 80.

GRAPH 1  
Maximum likelihood TOPALS fit of log mortality by age, females  
Pará de Minas microregion – 2010



Source: Human Mortality Database (2015) and IBGE (2010) and Ministry of Health (<http://www.datasus.gov.br>).

Note: Open circles represent  $\ln(D_x/N_x)$  for each single year of age. Tick marks on the horizontal axis represent 14 ages with no reported female deaths (2, 4, 6-14, 17, 19 and 27). The thin solid curve is the standard schedule  $\lambda_x^*$ . Segmented line at the top represents age-specific offsets from the standard ( $\lambda_x - \lambda_x^*$ ); heights of the solid points are equal to maximum likelihood estimates for  $\alpha$ . Final fitted schedule  $\lambda_x$  equals standard schedule plus estimated offset. Shaded bands represent 95% pointwise confidence intervals for linear spline offsets and for log mortality schedule  $\lambda_x$ .

- A Poisson regression approach allows calculation of standard errors for the estimated spline offsets and the corresponding standard errors in estimated log mortality rates. Pointwise 95% confidence intervals (discussed in more detail in section 2.4) appear as the light bands in Graph 1.
- As demonstrated in the next subsection, the shape of the final estimated schedule is extremely insensitive to the choice of standard. Because the approach allows different offsets at different ages, almost any choice of standard will produce a fitted schedule very similar to the thick line labeled TOPALS when applied to this data set.

### *Insensitivity to choice of standard*

Selection of an appropriate standard can be a big challenge when using a relational model. One major advantage of our regression approach is that the estimated mortality rates change very little with different choices for the standard  $\lambda^*$ , so that demographers can use almost any schedule as a starting point.

To illustrate the (in)sensitivity of our proposed method, we estimated mortality rates for Brazilian areas using seven extremely different standards, based on average log mortality rates over alternative subsets of the HMD (2015): lifetables from (1) all countries, (2) Chile, (3) Sweden, (4) France, (5) Eastern European countries, (6) Anglophone countries, and (7) Asian countries.

The top left panel of Graph 2 shows the seven different standards. The other three panels of Graph 2 present Poisson regression estimates of male mortality rates using each of the seven standards for three municipalities with very different population sizes in 2010: Ribeirão Preto - SP (6,186 deaths over 2009-2011, with a 2010 census population of 290,165 males), Itajubá - MG (1,008 deaths, 44,489 males), and Fernando de Noronha (12 deaths, 1,292 males).

In all areas, the mortality pattern estimated by the regression model is very similar for all seven choices of standard schedule, and goodness of fit is almost identical. In order to compare fits, we calculate the deviance-based  $R^2$  measure for Poisson regression models proposed by Cameron and Windmeijer (1996, p. 211).<sup>9</sup> Each standard  $s$  produces a different fitted model and therefore a different goodness of fit  $R_s^2$  to the observed deaths and exposure.

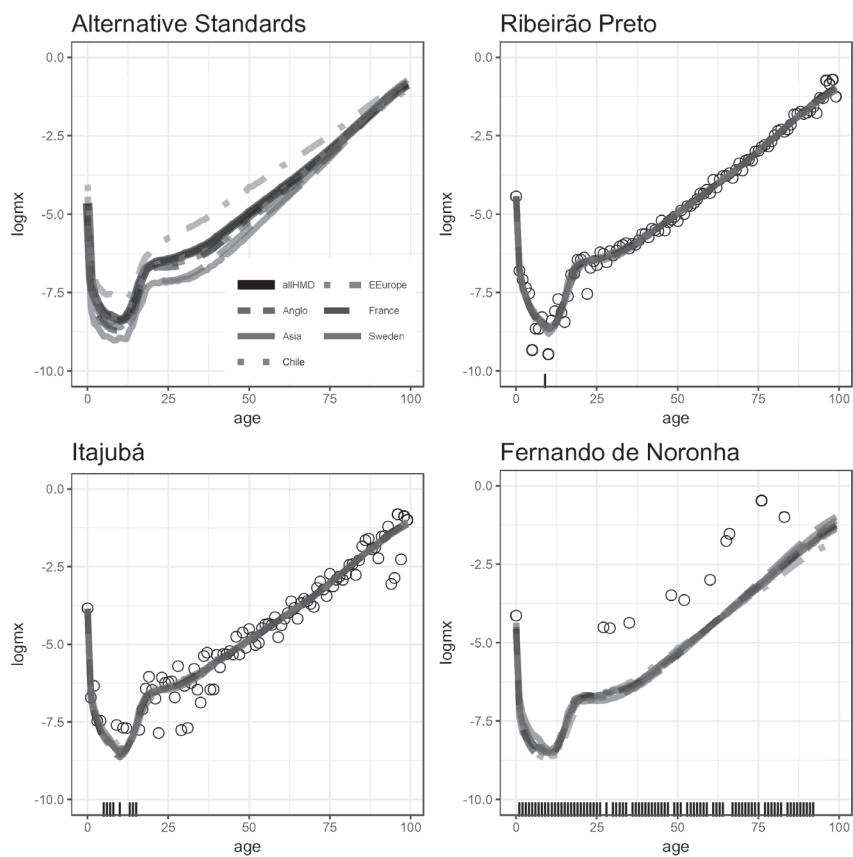
Ribeirão Preto (top right panel) had at least one male death over 2009-2011 at every single-year age except 9, and three years of exposure is sufficient to generate a regular pattern in age-specific rates even before smoothing (open circles). Regardless of the choice of standard, a TOPALS regression model reproduces the observed pattern well. All standards yield similar goodness of fit. Rounded to three decimal places, the seven Cameron-Windmeijer  $R_s^2$  values for  $s \in \{All\_HMD \dots Asian\}$  all equal 0.999, indicating very good and extremely similar fits.

<sup>9</sup> In our model this measure is  $R_{DEV,P}^2 = 1 - \frac{\sum_x D_x \ln(D_x / \hat{D}_x) - (\sum_x D_x - \hat{D}_x)}{\sum_x D_x \ln(D_x / \bar{D}_x)}$ , where  $\hat{D}_x = N_x \exp(\hat{\lambda}_x)$ ,  $\bar{D} = (\sum_x D_x) / (\sum_x N_x)$ , and by convention  $0 \cdot \ln(0) = 0$ .



In a municipality with a moderate-sized population, such as Itajubá in the bottom left panel, directly observed rates (open circles) are noisier and there are more ages at which no deaths are recorded. In this case regression smoothing is more necessary, but again the choice of standard has only minor effects on the fitted rates: estimated schedules are very similar for all eight standards.  $R_s^2$  for  $s \in \{All\_HMD \dots Asian\}$  all round to 0.993, again indicating very similar fits.

**GRAPH 2**  
**Seven alternative standard schedules and corresponding estimated male mortality rates ( $\log m_x$ ) by age**  
**for three municipalities with different population sizes**  
**Ribeirão Preto, Itajubá and Fernando de Noronha – 2010**



Source: IBGE (2010) and Ministry of Health (<http://www.datasus.gov.br>).

Finally, the bottom right panel shows regression results for an extreme case with a very small municipal population. Fernando de Noronha had only twelve recorded deaths for males over 2009-2011 (one infant death, one death each at ages 27, 29, 35, 52, 60, 65, 66, 76, and 83, and two deaths at age 48). The male resident population was also zero at many ages above 75 and at all ages 91-100. Remarkably, TOPALS regression produces reasonable (although highly uncertain) estimated mortality schedules, even in this case where mortality

and exposure data is very sparse. As in the other panels, the fitted schedules are similar for all choices of standard, although with only twelve observed deaths fits are of course slightly less stable.  $R_s^2$  values for  $s \in \{All\_HMD \dots Asian\}$  are (.689, .686, .676, .673, .681, .678, .670).

### Calculation of standard errors

A nonlinear regression approach to estimating the linear spline offsets  $\alpha$  allows approximation of standard errors for  $\alpha$  and  $\lambda$  terms. From standard nonlinear regression analysis (e.g. RUUD 2000, p. 327), the 7x7 covariance matrix of offset estimators  $\hat{\alpha}$  is approximately equal to the negative of the inverse Hessian matrix of second derivatives of  $Q(\alpha)$  from equation (4):  $V(\hat{\alpha}) \approx -\left[\frac{\partial^2 Q(\hat{\alpha})}{\partial \alpha \partial \alpha'}\right]^{-1}$ . In the TOPALS model the negative of the Hessian is

$$-\left(\frac{\partial^2 Q(\hat{\alpha})}{\partial \alpha \partial \alpha'}\right) = \sum_x \left[ \hat{D}_x(\hat{\alpha}) b_x b_x' \right] + 2\Delta \quad (5)$$

where  $\hat{D}_x = N_x e^{\lambda_x(\hat{\alpha})}$  is the number of deaths at age  $x$  predicted by the fitted model,  $b_x$  is a 7x1 column vector containing the (transposed) row of the spline matrix  $B$  that corresponds to age  $x$ , and  $\Delta$  is a 7x7 matrix of differencing constants related to the penalty term in equation (4).<sup>10, 11</sup> Inverting this matrix produces an estimate of the 7x7 covariance matrix  $V(\hat{\alpha})$ , which in turn produces a 100x100 covariance matrix for estimated log mortality rates:  $V(\hat{\lambda}) = V(\lambda^* + B\hat{\alpha}) = BV(\hat{\alpha})B'$  (6)

The square roots of the diagonal of  $V(\hat{\lambda})$  are the estimated pointwise standard errors for log mortality rates at ages 0...99.

Graph 1 includes example uncertainty estimates for the Pará de Minas female schedule. At each age, bands at (estimates  $\pm 1.96$  standard errors) illustrate 95% confidence intervals for both the spline offsets and the log rates. Analogous calculations are possible for any TOPALS regression estimates. Equation (5) shows that uncertainty about log mortality rates will be high when there are few expected local deaths around the age of interest. Standard errors are therefore large when age-specific populations are small (as in a small municipality) and when mortality rates are very low (as at ages 5-15).

### Estimation errors for TOPALS and IS with simulated data

The evaluations in the previous subsections use real data in cases for which true mortality rates are unknown. It is also important to evaluate TOPALS regression in simulations with known rates in order to understand the method's statistical performance – particularly in comparison with IS.

<sup>10</sup>  $\Delta$  is 7x7 with (1,2,2,2,2,2,1) on the main diagonal, -1 in every element of the first subdiagonal and first superdiagonal, and zero everywhere else.

<sup>11</sup>  $R$  calculates the Hessian automatically if the *optim* function includes the argument *hessian=TRUE*.

For this purpose, we compared TOPALS regression and IS by generating thousands of Monte Carlo samples of different sizes from a known mortality schedule (the national  $\mu_x$  schedule for Brazilian males in 2010). Every sample in our simulations had a predetermined number of individuals  $N_x$  at each single year of age, and a random number of deaths  $D_x \sim \text{Poisson}(N_x \mu_x)$ .

We calculated two TOPALS regression fits from each sample – one using a standard schedule with an *incorrect* shape (the all-HMD average used earlier), and another using a standard with the *correct* shape (the  $\mu_x$  values that actually generated the data). We also calculated two IS estimates for each sample, using both the HMD standard (incorrect shape) and the true Brazilian schedule (correct shape).<sup>12</sup>

The main questions for the Monte Carlo exercise are (1) Do typical errors approach zero as sample sizes increase? and (2) Does using an incorrect standard cause significant bias? The second question is particularly important for IS estimates in small areas, because the mortality schedule for a subregion may not have the same shape as the schedule for the larger region in which it resides. In that case, estimating the subregional schedule using the (locally incorrect) shape of the regional schedule could produce large errors.

Complete Monte Carlo results are available in the supplemental files on the project website. Here we briefly summarize by reporting mean absolute errors (MAE) in estimated log mortality rates and in estimated life expectancy over samples with different estimators and small-area sample sizes (Table 1).

TABLE 1 Mean absolute errors of log mortality rates and estimated life expectancy, with alternative methods, standards, and sample sizes				
$N_x$	With correct standard		With incorrect standard	
	IS	TOPALS	IS	TOPALS
$ \ln \hat{\mu}_x - \ln \mu_x $ over all ages and samples				
1	.415	.417	.733	.692
100	.040	.142	.539	.187
10000	.004	.028	.538	.057
$ \hat{e}_0 - e_0 $ over all samples				
1	4.79	4.94	6.31	6.10
100	0.52	1.29	5.54	1.43
10000	0.05	0.15	5.54	0.15

Source: Simulated data as described in text.  
Note: IS=indirect standardization,  $N_x$ =number of individuals at each single-year age

TOPALS regression, which is able to “bend” any standard mortality schedule via age-specific offsets, performs similarly, regardless of the chosen standard. At extremely small sample sizes (such as  $N_x=1$ , which would correspond to total local population of only 100 people), TOPALS errors are fairly large,<sup>13</sup> but they quickly become small as sample sizes

<sup>12</sup> In our notation, IS may be expressed as  $\lambda_x = \lambda_x^* + \hat{\gamma}$ , where  $\hat{\gamma} = \ln \left[ \sum_x D_x / \sum_x (N_x \exp(\lambda_x^*)) \right]$  is the logarithm of (observed deaths)/(expected deaths). IS shifts a standard schedule of log rates up or down, by the same amount at each age.

<sup>13</sup> For all methods, in very small samples the reported errors are almost entirely caused by underestimates of mortality rates and overestimates of  $e_0$ . Scherbov and Ediev (2011) show that this bias is a mathematical feature of small-sample estimates.

increase (Table 1). This occurs for almost any chosen standard, as comparison of the last two columns in Table 1 suggests.

In contrast, the performance of IS depends critically on the chosen shape of the standard mortality schedule. If the standard shape is correct (first columns of table 1), then errors are smaller than TOPALS errors and decrease similarly with sample size. However, if the chosen standard schedule does *not* match the true schedule's shape (last columns of table 1), then IS can exhibit strong bias even as sample sizes become very large.

We conclude from the simulations that IS is slightly better than TOPALS when the shape of the local mortality schedule matches the assumed standard. (This makes statistical sense: when a strong mathematical assumption about the exact shape of the mortality schedule is correct, then adding extra parametric flexibility via TOPALS can only increase the variance of mortality estimates). However, TOPALS decisively outperforms IS when the shape of the local mortality schedule differs from the standard.

### The underlying mortality pattern in Brazilian small areas

A flexible regression model like TOPALS is valuable when regional and subregional age patterns of mortality differ. We now return to contemporary Brazilian data to investigate whether that situation is common.

We illustrate intra-state differentials in underlying mortality patterns with results from the state of Minas Gerais and some of its subregions. We chose Minas Gerais as an example because its vital registration is very complete (LIMA; QUEIROZ, 2011; LIMA; QUEIROZ; SAWYER, 2014), and because it is a state with heterogeneous subregions in terms of social demographic variables (UNDP, 2013).

#### *TOPALS regression versus IS*

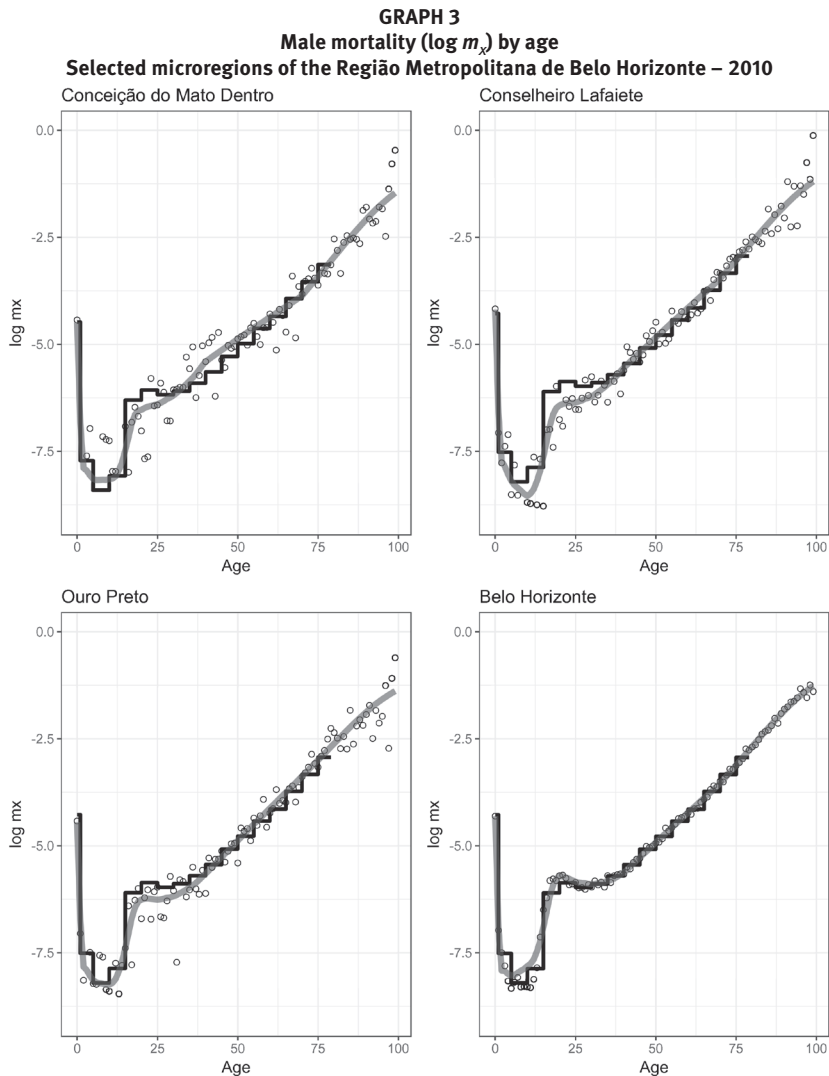
IS methods use the mortality pattern in a large region as a standard shape for the patterns in component subregions (HORTA et al., 1998; LIMA; QUEIROZ; SAWYER, 2014), which imposes a strong homogeneity assumption. In contrast, TOPALS regression allows estimation and smoothing of age-specific mortality rates for small areas without assuming homogeneity.

We first show results for several microregions in the Belo Horizonte metropolitan mesoregion, *Região Metropolitana de Belo Horizonte* (RMBH). We applied IS by (1) calculating mesoregion-level male mortality rates for standard age groups (0,1-4,5-9,...,75-79)<sup>14</sup> as death/exposure ratios for the entire RMBH mesoregion; (2) calculating the expected number of deaths in each microregion at those rates, given local age-specific exposure; (3) calculating a multiplier for each microregion as (observed deaths)/(expected deaths at mesoregional rates); (4) assuming that the rate schedule in each microregion is its local multiplier times the mesoregional schedule. We also estimated TOPALS regressions for each microregion using the HMD standard.

<sup>14</sup> Standard practice (e.g. FREIRE et al., 2015) uses an upper age limit, often 80, for IS. We repeat that practice for the calculations in Graph 3.

Because the Belo Horizonte microregion contains 76% of the male population of the RMBH mesoregion, it dominates these calculations: the shape of the RMBH schedule calculated in step (1) of IS is essentially determined by the age pattern of mortality in Belo Horizonte. This means, in turn, that the subregional schedules calculated in step (4) will also have the Belo Horizonte age pattern.

Graph 3 shows the results for four of the eight microregions in the RMBH mesoregion, and illustrates how TOPALS regression adds flexibility that can capture meaningful differences in mortality schedules across adjacent small areas.

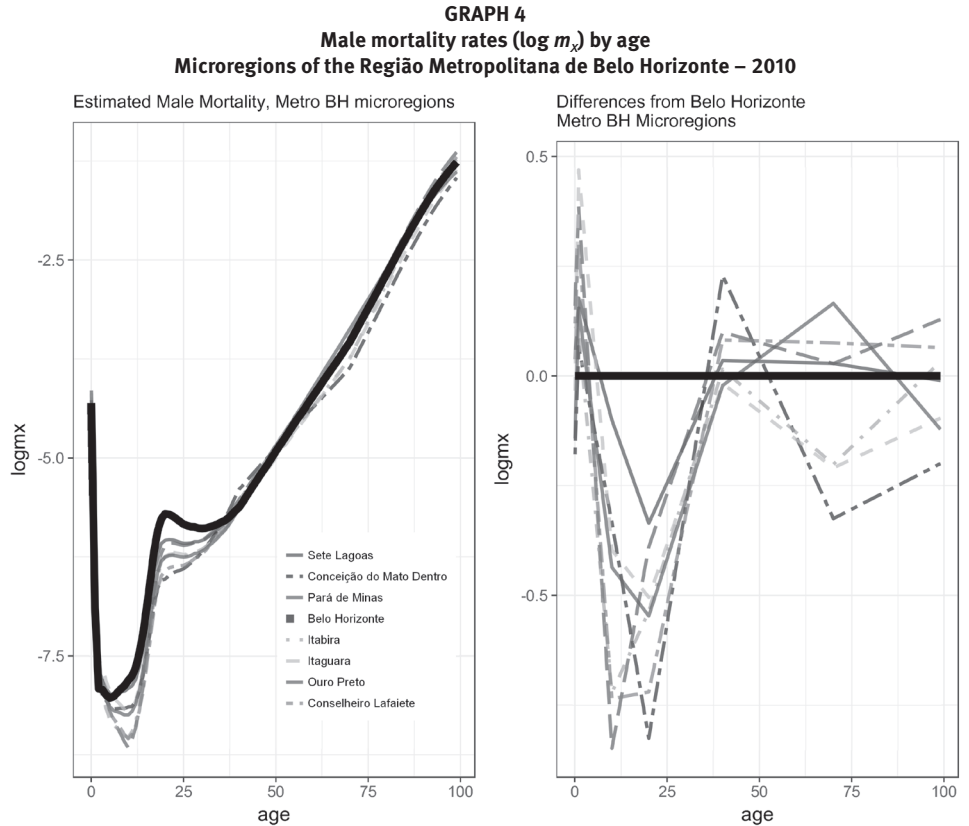


Source: IBGE (2010) and Ministry of Health (<http://www.datasus.gov.br>).  
Note: Step functions over ages 0-79 are estimates from indirect standardization based on the mesoregion's aggregate rates. Open circles are  $\ln(D_x/N_x)$  from recorded deaths. Smooth lines are local TOPALS regression estimates. There are eight microregions in the Belo Horizonte metropolitan mesoregion; these four were selected in order to illustrate the potential variety in local age patterns.

*Differences within a mesoregion: microregions in RMBH*

TOPALS regression allows each subregion to have a different age pattern of mortality. Graph 4 presents TOPALS estimates of mortality schedules for all microregions within the RMBH mesoregion. In the left panel of Graph 4 one can see that at the youngest ages, at young adult ages, and at older ages 70+, there are notable differences in mortality rates between microregions. These differences become clearer in the right panel of Graph 4, which shows the differences between each microregion’s log rates and those of the Belo Horizonte microregion.<sup>15</sup>

Brazilian data quality studies show that RMBH had nearly complete coverage of deaths by sex and age in 2010 (LIMA; QUEIROZ, 2011; LIMA; QUEIROZ; SAWYER, 2014). Clear differences between the schedules in Graph 4 show that homogeneity in the underlying mortality patterns between adjacent small areas (in this case, adjacent Brazilian microregions) may be a poor assumption.



Source: IBGE (2010) and Ministry of Health (<http://www.datasus.gov.br>).

<sup>15</sup> Because all fitted schedules are the sum of a common standard schedule plus a location-specific linear spline, the differences between local schedules are also linear splines.

### *Differences between mesoregions: Minas Gerais*

In Graph 5 we shift the level of analysis to larger geographic scales. In contrast to the previous analyses, the RMBH mesoregion is now one of several subregions, with the entire state of Minas Gerais as the larger area. The graph shows TOPALS regression fits for male mortality rates by mesoregion (left panel). Differences between each mesoregion's fitted schedule and the RMBH fit appear in the right panel. The left panel shows different underlying male mortality patterns between mesoregions, mainly for infants, young children, young adults, and at the oldest ages. Also, one can see that there is one mesoregion, Vale do Mucuri, with the highest mortality rates at almost every age. For females (results not shown in Graph 5),<sup>16</sup> mesoregions other than Vale do Mucuri had more homogeneous mortality patterns than males, except at the oldest ages.

The notable differences between the Vale do Mucuri male mortality pattern and all other mesoregions in Minas Gerais need to be more carefully investigated. There is evidence that mortality due to violence and transit accidents in Brazil is increasing among young adult males (SOUZA; LIMA, 2007). There is also emerging evidence that mortality rates due to homicides are increasing in Brazil outside of metropolitan areas (SOUZA; LIMA, 2007; WAISELFISZ, 2013). However, these trends do not explain why mortality rates for children and adults are higher only in the Vale do Mucuri mesoregion, as we see in Graph 5.

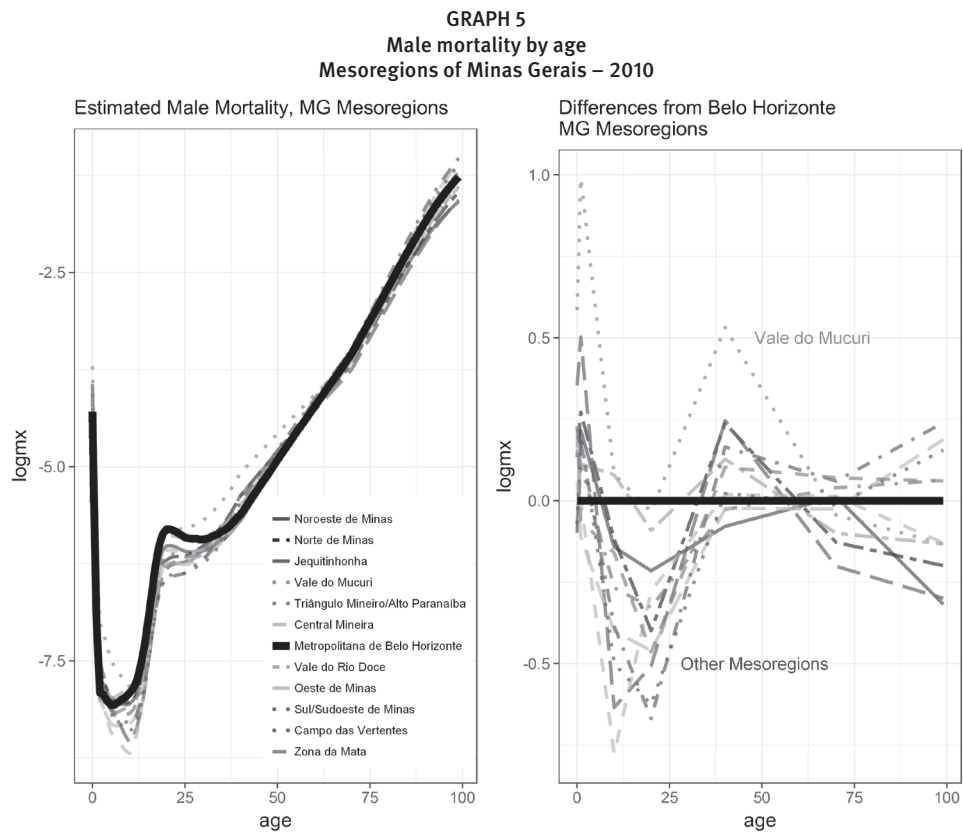
In the right panel of Graph 5 we show differences between the estimated log mortality rates of each mesoregion and the RMBH mesoregion.<sup>17</sup> Setting Vale do Mucuri aside temporarily, there are notable differences between RMBH and other Minas Gerais mesoregions.<sup>18</sup> In particular, child and young adult mortality for males is much lower outside of RMBH. Mesoregions located in the North and Northeast of the state are characterized by higher infant mortality rates and lower young adult mortality rates (except for Vale do Mucuri). The high infant mortality rates outside of RMBH are consistent with the findings of Almeida and Szwarcwald (2014) for Minas Gerais. Municipalities in northern Minas Gerais, especially in the Jequitinhonha mesoregion, probably experienced the highest rates (CASTRO; SIMÕES, 2009; ALMEIDA; SZWARC WALD, 2014). The geographical distribution of child mortality risk seems similar to the infant mortality pattern (MCKINNON, 2010).

Differences between mesoregions in violence and transit accidents could explain the regional differences in the mortality rates of young adult males. Although mortality due to violence is increasing among young adults around many areas in Brazil (SOARES FILHO et al., 2007; SOUZA; LIMA, 2007; WAISELFISZ, 2013), mortality due to transit accidents is also responsible for higher rates in state capitals and metropolitan areas (MELLO-JORGE; LATORRE, 1994; MELLO-JORGE; GAWRYSZEWSKI; LATORRE, 1997; GAWRYSZEWSKI; KOIZUMI; MELLO-JORGE, 2004).

<sup>16</sup> Results for females are available on the project website [URL redacted for review]

<sup>17</sup> The shape of the mortality schedule for the RMBH mesoregion is very similar to the shape for the entire state of Minas Gerais.

<sup>18</sup> Our project website has color versions of all graphics.



Source: IBGE (2010) and Ministry of Health (<http://www.datasus.gov.br>).  
Note: Right panel shows  $\log(\mu_x) - \log(\mu_{x, \text{RMBH}})$  for each mesoregion in Minas Gerais.

In contrast to the situation for infants and young adults, at the oldest ages we have little evidence with which to evaluate the mesoregional differences. Due to the poor quality of both death and population data, there is no evidence about the underlying mortality pattern at oldest ages in Brazil, even in larger areas. A study about centenarians showed that age-specific mortality rates in Brazil are likely to be underestimated due to misreported ages (GOMES; TURRA, 2009). Brazilian centenarians are probably exaggerating their ages in the census (GOMES; TURRA, 2009). The results in Graph 5, in which mesoregions located in the North and Northeast of the state have the lowest mortality rates at the oldest ages, suggest that the data quality problems could lead to underestimation of mortality rates in those parts of Minas Gerais.

*Consistency of small- and large-area TOPALS estimates*

Consistency between disaggregated and aggregated schedules is a desirable property for a small-area estimation procedure. Ideally, a region's estimated mortality schedule should be identical regardless of whether one (1) aggregates death and exposure data for the entire region and then estimates a single mortality schedule, or (2) estimates separate mortality schedules for subregions and then aggregates those schedules.



TOPALS estimates of age-specific mortality rates do not have this mathematical property exactly, but in practice they are extremely consistent when aggregated. For purposes of analysis, note that empirical consistency implies that the estimated rate from aggregated regional data ( $\tilde{\mu}_x$ ) is related to subregional estimates ( $\hat{\mu}_{x1}, \hat{\mu}_{x2}, \dots$ ) at each age as

$$\tilde{\mu}_x = \sum_i \left( \frac{N_{xi}}{N_x} \right) \hat{\mu}_{xi} \quad (7)$$

where  $N_{xi}$  is the exposure at age  $x$  in subregion  $i$  and  $N_x = \sum_i N_{xi}$ . Multiplying both sides of equation (7) by  $N_x$  produces  $N_x \tilde{\mu}_x = \sum_i N_{xi} \hat{\mu}_{xi}$ , or even more simply in terms of predicted deaths  $\tilde{D}_x = \sum_i \hat{D}_{xi}$ . Thus, consistency between estimated regional and subregional schedules implies that, at every age, the sum of predicted deaths in each small area equals the number of predicted deaths for the larger area.

To assess the empirical consistency of TOPALS estimates, we aggregated predicted deaths by sex and age from smaller to larger areas at three geographic levels – from 137 Brazilian mesoregions to 27 state schedules, from 558 microregions to 137 mesoregional schedules, and from 5565 municipalities to 558 microregional schedules. In all cases we will call the higher-level aggregate the “regional” schedule. There are thus  $27+137+558 = 722$  aggregate regions and  $2 \cdot 722 = 1444$  sex-specific regional schedules. For each regional schedule we calculated the TOPALS-predicted number of deaths at each age  $x=0 \dots 99$  from aggregated regional data ( $\tilde{D}_x$ ) and the sum of predicted deaths from the corresponding subregional TOPALS estimates ( $\hat{D}_x = \sum_i \hat{D}_{xi}$ ).

Perfect consistency between regional and subregional schedules occurs if all 100 calculated differences  $\tilde{D}_x - \hat{D}_x$  are zero. Estimates are highly consistent if almost all differences are small. For each of the 1444 aggregated regional schedules, we evaluated consistency by calculating the mean absolute difference of the TOPALS estimates from higher- and lower-level geography,  $MAD = \sum_x |\tilde{D}_x - \hat{D}_x| / 100$ , and the mean absolute percent difference  $MAPD = \sum_x \left| \frac{\tilde{D}_x - \hat{D}_x}{\tilde{D}_x} \right|$ .

With TOPALS estimates, 1408 of the 1444 regional schedules had either  $MAD < 1$  death or  $MAPD < 1$  percent. In the other 36 regional schedules the MAD never exceeded 2.12 deaths. These are very small inconsistencies.

For males in the RMBH mesoregion, for example, the TOPALS mortality estimates illustrated in Graphs 4 and 5 (for the eight component microregions and the larger RMBH mesoregion, respectively) are highly consistent. In the RMBH mesoregion the TOPALS-predicted number of deaths in 2009-2011 from the mesoregional data is 58179, while the sum of deaths predicted in the component microregions is 58183. The fitted age-specific schedules for RMBH are also extremely similar: over 100 ages, the MAD between regional and subregional predictions is 0.84 deaths, and the MAPD is 0.3%.

In sum, TOPALS estimates for higher- and lower-level geographies are not perfectly consistent in theory. But they are highly consistent in practice.

## Conclusion and discussion

In this paper we dealt with the problem of unstable event/exposure estimates in small populations. We proposed a Poisson regression method for estimating age-specific mortality rates that is based on the TOPALS relational model, and we illustrated the new method with small-area data from Brazil in 2010. In contrast to alternative estimation methods, our approach uses disaggregated, single-year age data, it does not impose a specific age pattern of mortality rates, and it allows appropriate use of the “zero” cells in which there are no recorded deaths.

Our application to Brazilian small areas was motivated by the hypothesis that different degrees of urbanization and development in subnational populations could mean that even adjacent areas might have different mortality patterns by age. Under this hypothesis, IS techniques commonly used to estimate age-specific mortality rates could produce biased estimates of small-area schedules.

To demonstrate the plausibility of this hypothesis, we estimated mortality rates between and within mesoregions in Minas Gerais, one of Brazil’s most socially and demographically heterogeneous states. We observed notable differences in relative levels of child, young adult, and old-age mortality between and within mesoregions. TOPALS regression can capture these meaningful differences in mortality schedules across adjacent small areas, which would be obscured by IS.

Many subnational areas in Brazil had almost full coverage of deaths in 2010, especially in the South and Southeast. For those areas, our method could be used to estimate complete life tables and to compare levels of mortality in terms of life expectancy at birth. Our method also allows comparisons of mortality rates at high ages (60-99).

TOPALS regression represents a *potential* advance for mortality estimation in small areas with incomplete death registration that require undercount adjustments. Previous studies have identified many areas in Brazil with poor vital registration coverage (PAES; ALBUQUERQUE, 1999; LIMA; QUEIROZ, 2011; JUSTINO; FREIRE; LUCIO, 2012; LIMA; QUEIROZ; SAWYER, 2014; FREIRE et al., 2015). Current approaches for estimating mortality rates in such areas (e.g., HORTA et al., 1998; LIMA; QUEIROZ; SAWYER, 2014) use a two-step process: (1) IS to smooth small-area rates and estimate expected numbers of deaths by age group, followed by (2) correction for under-registration by applying Death Distribution Methods to the smoothed data.

As we have demonstrated above, IS based on higher-level geographic aggregates can lead to systematic mistakes in the first step of this process. These mistakes could then lead to incorrect conclusions about the completeness of registration in the second step. For example, if one used IS to smooth rates for Conselheiro Lafaiete (the step function in the top right panel of Graph 3), then the cluster of lower-than-expected mortality rates at child and young adult ages in would be misinterpreted as under-registration.

A method like TOPALS regression that does not impose a particular shape on the first-stage estimates would clearly work better in cases like Conselheiro Lafaiete, where the data is good, and a flexible model will more accurately estimate real local patterns. However, it remains unclear whether flexibility in the first stage has net benefits when observed death/exposure ratios are strongly affected by under-registration.<sup>19</sup> TOPALS regression approach for first-stage estimation and smoothing seems promising as a component for more complex correction models, but we have not investigated its advantages and disadvantages in this paper.

## References

- AGOSTINHO, C. S. **Estudo sobre a mortalidade adulta, para Brasil entre 1980 e 2000 e Unidades da Federação em 2000**: uma aplicação dos métodos de distribuição de mortes. Tese (Doutorado em Demografia) – Universidade Federal de Minas Gerais, Belo Horizonte, 2009.
- ALEXANDER, M.; ZAGHENI, E.; BARBIERI, M. A Flexible Bayesian model for estimating subnational mortality. In: ANNUAL MEETING OF THE POPULATION ASSOCIATION OF AMERICA. Washington, 2016. Available at: <<https://paa.confex.com/paa/2016/mediafile/ExtendedAbstract/Paper6721/PAA.pdf>>. Accessed on: 23 Sep. 2016.
- ALMEIDA, W. S.; SZWARCOWALD, C. L. Mortalidade infantil nos municípios brasileiros: uma proposta de método de estimação. **Revista Brasileira de Saúde Materno Infantil**, v. 14, n. 4, p. 331-342, 2014.
- ARAÚJO, J. D. Polarização epidemiológica no Brasil. **Epidemiologia e Serviços de Saúde**, Brasília, v. 21, n. 4, p. 533-538, 2012.
- ASSUNÇÃO, R. M.; SCHMERTMANN, C. P.; POTTER, J. E.; CAVENAGHI, S. M. Empirical Bayes estimation of demographic schedules for small areas. **Demography**, v. 42, n. 3, p. 537-558, 2005.
- BARRETO, M. L.; CARMO, E. H. Padrões de adoecimento e de morte da população brasileira: os renovados desafios para o Sistema Único de Saúde. **Ciência & Saúde Coletiva**, v. 12, supl., p.1779-1790, 2007.
- BARRETO, M. L.; CARMO, E. H.; NORONHA, C. V.; NEVES, R. B. B.; ALVES, P. C.; Mudanças dos padrões de morbi-mortalidade: uma revisão crítica das abordagens epidemiológicas. **PHYSIS – Revista de Saúde Coletiva**, v. 3, n. 1, 1993.
- BERNADINELLI, L.; MONTOMOLI, C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. **Statistics in Medicine**, v. 11, n. 8, p. 983-1007, 1992.
- BRASS, W. On the scale of mortality. In: BRASS, W. (Ed.). **Biological aspects of demography**. London: Taylor and Francis, 1971.
- BRASS, W.; COALE, A. J. Methods of analysis and estimation. In: BRASS, W.; COALE, A.J.; DEMENY, P.; LORIMER, F. et al. (Ed.). **The demography of tropical Africa**. Princeton: Princeton University Press, 1968.
- BRAVO, J. M.; MALTA, J. Estimating life expectancy in small population areas. In: EUROSTAT – European Commission (Ed.). **Work session on demographic projections: methodologies and**

<sup>19</sup> It is also unclear whether it is better to smooth mortality schedules before correcting for undercount, or vice-versa. That remains an open research question. TOPALS regression would likely be useful in the smoothing stage, regardless of the order of smoothing and correction.

working papers. Luxembourg: Eurostat, 2010. p. 113-126. Available at: <<http://tinyurl.com/bravo-malta-eurostat>>. Accessed on: 9 Sep. 2016.

CAMARDA, C. G. Mortality smooth: an R package for smoothing Poisson counts with P-splines. **Journal of Statistical Software**, v. 50, p. 1-24, 2012.

CAMERON, A. C.; WINDMEIJER, F. A. G. R-squared measures for count data regression models with applications to health-care utilization. **Journal of Business and Economic Statistics**, v. 14, n. 2, p. 209-220, 1996.

CASTRO, M. C.; SIMÕES, C. C. S. Spatio-temporal trends of infant mortality in Brazil. In: XXVI IUSSP INTERNATIONAL POPULATION CONFERENCE. **Proceeding...** Marrakech: International Union for the Scientific Study of Population, 2009. Available at: <<http://iussp2009.princeton.edu/papers/92270>>. Accessed on: 01 Oct. 2015.

CONGDON, P. Life expectancies for small areas: a Bayesian random effects methodology. **International Statistical Review**, v. 77, n. 2, p. 222-240, 2009.

CURRIE, I. D.; DURBAN, M.; EILERS, P.H.C. Smoothing and forecasting mortality rates. **Statistical Modelling**, v. 4, n. 4, p. 279-298, 2004.

CURTIN, L. R.; KLEIN, R. J. Direct standardization (age-adjusted death rates). **Healthy People 2000 Statistical Notes**, n. 6, Mar. 1995.

DE BEER, J. Smoothing and projecting age-specific probabilities of death by TOPALS. **Demographic Research**, v. 27, n. 20, p. 543-592, 2012.

DE BOOR, C. **A practical guide to splines**. Berlin: Springer, 1978.

DIVINO, F.; EGIDI, V.; SALVATORE, M. A. Geographical mortality patterns in Italy: a Bayesian analysis. **Demographic Research**, v. 20, n. 18, p. 435-466, 2009.

EAYRES, D.; WILLIAMS, E. S. Evaluation of methodologies for small area life expectancy estimation. **Journal of Epidemiology and Community Health**, v. 58, n. 3, p. 243-249, 2004.

EILERS, P. H. C.; MARX, B. D. Flexible smoothing with B-splines and penalties. **Statistical Science**, v. 11, n. 2, p. 89-121, 1996.

FEENEY, G. Estimating infant mortality trends from child survivorship data. **Population Studies**, v. 34, p. 109-128, 1980.

FREIRE, F. H. M. A.; QUEIROZ, B. L.; LIMA, E. E. C.; GONZAGA, M. R.; SOUZA, F. H. Mortality estimates and construction of life tables for small areas in Brazil, 2010. In: ANNUAL MEETING OF THE POPULATION ASSOCIATION OF AMERICA. **Proceedings...** San Diego, 2015. Available at: <<http://paa2015.princeton.edu/abstracts/150854>>. Accessed on: 01 Dec. 2015.

FREIRE, F. H. M. A.; SOUZA, F. H.; QUEIROZ, B. L.; LIMA, E. E. C.; GONZAGA, M. R. Tabelas de sobrevivência para os municípios brasileiros em 2010: análise espacial do padrão e nível da mortalidade. In: XIX ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS. **Anais...** São Pedro (SP): Abep, 2014. Available at: <<http://abep.info/anais/anais.php?id=59#.VsyRk30rLcs>>. Accessed on: 01 Nov. 2015.

FRIAS, P. G.; SZWARCOWALD, C. L.; SOUZA JUNIOR, P. R. B de; ALMEIDA, W. S.; LIRA, P. I. C. Correcting vital information: estimating infant mortality, Brazil, 2000-2009. **Revista de Saúde Pública**, v. 46, n. 6, p. 1-10, jul., 2013.

GAWRYSZEWSKI, V. P.; KOIZUMI, M. S.; MELLO-JORGE, M. H. P. As causas externas no Brasil no ano 2000: comparando a mortalidade e a morbidade. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 20, n. 4, p.995-1003, jul./ago. 2004.

GOMES, M. M. F.; TURRA, C. M. The number of centenarians in Brazil: indirect estimates based on death certificates. **Demographic Research**, v. 20, n. 20, p. 495-502, 2009.

HILL, K. Approaches to the measurement of childhood mortality: a comparative review. **Population Index**, v. 57, n. 3, p. 368-382, 1991.

HORTA, C. J. G.; NOGUEIRA, O. J. O.; WONG, L. L. R.; CARVALHO, J. A. M. Estimativas de mortalidade para os municípios brasileiros: uma proposta metodológica para resultados comparativos. In: XI ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS. **Anais...** Caxambu: Abep, 1998. Available at: <<http://www.abep.org.br/?q=publicacoes/anais/anais-1998-sess%C3%B5es-tem%C3%A1ticas-especiais>>. Accessed on: 03 Nov. 2015.

HUMAN MORTALITY DATABASE. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2015. Available at: <<http://www.mortality.org/>>. Accessed on: 01 Oct. 2015.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico do Brasil**. Rio de Janeiro: IBGE, 2010 (Resultados da Amostra).

JONKER, M. F.; VAN LENTHE, F. J.; CONGDON, P. D.; DONKERS, B.; BURDOF, A.; MACKENBACK, J. P. Comparison of Bayesian random-effects and traditional life expectancy estimations in small-area applications. **American Journal of Epidemiology**, v. 176, n. 10, p. 929-937, 2012.

JUSTINO, J. R.; FREIRE, F. H. M. A.; LUCIO, P. S. Estimação de sub-registros de óbitos em pequenas áreas com os métodos bayesiano empírico e algoritmo EM. Rio de Janeiro: **Revista Brasileira de Estudos de População**, v. 29, n. 1, p. 87-100, jan./jun. 2012.

LIMA, E. E. C.; QUEIROZ, B. L. Evolution of the deaths registry system in Brazil: associations with changes in the mortality profile, under-registration of death counts, and ill-defined causes of death. **Cadernos de Saúde Pública**, v. 30, n. 8, p. 1721-1730, 2014.

\_\_\_\_\_. A evolução do sub-registro de mortes e causas de óbitos mal definidas em Minas Gerais: diferenciais regionais. **Revista Brasileira de Estudos de População**, v. 28, n. 2, p. 303-320, jul./dez. 2011.

LIMA, E. E. C.; QUEIROZ, B. L.; MISOV, T.; LENART, A. Methods to estimate mortality curves in small areas: an application to municipality data in Brazil. In: ANNUAL MEETING OF THE POPULATION ASSOCIATION OF AMERICA. **Proceedings...** Washington, 2016. Available at: <<https://paa.confex.com/paa/2016/mediafile/ExtendedAbstract/Paper6532/Version%20Final%206532%20PAA.pdf>>. Accessed on: 11 Sep. 2016.

LIMA, E. E. C.; QUEIROZ, B. L.; SAWYER, D. O. Método de estimação de grau de cobertura em pequenas áreas: uma aplicação nas microrregiões mineiras. **Cadernos de Saúde Coletiva**, v. 22, n. 4, p. 409-18, 2014.

MACNAB, Y. C.; DEAN, C.B. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. **Biometrics**, v. 57, n. 3, p. 949-956, Sep. 2001.

MCKINNON, S. A. **Municipal-level estimates of child mortality for Brazil: a new approach using Bayesian statistics**. (Dissertation). University of Texas at Austin, Faculty of the Graduate School, Texas, 2010. Available at: <<https://repositories.lib.utexas.edu/handle/2152/ETD-UT-2010-08-1866>>. Accessed on: 20 Sep. 2015.

MELO-JORGE, M. H. P. M.; GAWRYSZEWSKI, V. P.; LATORRE, M. R. D. O. Análise dos dados de mortalidade. **Revista de Saúde Pública**, v. 31, n. 4, p. 5-25, 1997.

MELO-JORGE, M. H. P. M.; LATORRE, M. R. D. O. Acidentes de trânsito no Brasil: dados e tendências. **Cadernos de Saúde Pública**, v. 10, n. 1, p. 19-44, 1994.

MELO-JORGE, M. H. P. M.; LAURENTI, R.; GOTLIEB, S. L. D. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do Sinasc. **Ciência & Saúde Coletiva**, v. 12, n. 3, p. 643-654, 2007.

MOULTRIE T. A.; DORRINGTON R. E.; HILL, A. G.; HILL, K.; TIMAEUS, I. M.; ZABA, B. **Tools for demographic estimation**. Paris: International Union for the Scientific Study of Population, 2013 Available at: <<http://demographicestimation.iussp.org/content/introduction-tools-demographic-estimation>>. Accessed on: 6 Sep. 2016.

MURRAY, C. J. L.; FERGUSON, B. D.; LOPEZ, A. D.; GUILLOT, M.; SALOMON, J. A.; AHMAD, O. Modified logit life table system: principles, empirical validation, and application. **Population Studies**, v. 57, n. 2, p. 165-182, 2003.

OCAÑA-RIOLA, R.; MAYORAL-CORTÉS, J. M. Spatio-temporal trends of mortality in small areas of Southern Spain. **BMC Public Health**, v. 10, n. 26, p. 1-12, 2010.

PAES, N. A.; ALBUQUERQUE, M. E. E. Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. **Revista de Saúde Pública**, v. 33, n. 1, p. 33-43, fev. 1999.

PLETCHER, S. D. Model fitting and hypothesis testing for age-specific mortality data. **Journal of Evolutionary Biology**, v. 12, n. 3, p. 430-439, 1999.

PRATA, P. R. A transição epidemiológica no Brasil. **Cadernos de Saúde Pública**, v. 8, n. 2, p.168-175, abr./jun. 1992.

QUEIROZ, B. L.; LIMA, E. E. C.; FREIRE, F. H. M. A.; GONZAGA, M. R. Adult mortality estimates for small areas in Brazil, 1980-2010: a methodological approach. **Lancet** (British edition), v. 381, p. S120, 2013.

RAJARATNAM, J. K.; TRAN, L. N.; LOPEZ, A. D.; MURRAY, C. J. L. Measuring under-five mortality: validation of new low-cost methods. **PLOS Med**, v. 7, n. 4, e1000253, 2010. Available at: <<http://dx.doi.org/10.1371/journal.pmed.1000253>>. Accessed on: 7 Sep. 2016.

RAM, U.; JHA, P.; GERLAND, P.; HUM, R. J.; RODRIGUEZ, P.; SURaweera, W.; KUMAR, K.; KUMAR, R.; DIKSHIT, R.; XAVIER, D.; GUPTA, R.; GUPTA, P. C.; RAM, F. Age-specific and sex-specific adult mortality risk in India in 2014: analysis of 0-27 million nationally surveyed deaths and demographic estimates from 597 districts. **Lancet Glob Health**, v. 3, p.767-75, Dec. 2015.

RIGGAN, W.; MANTON, K.; CREASON J.; WOODBURY, M.; STALLARD, E. Assessment of spatial variation of risks in small populations. **Environ Health Perspectives**, v. 96, p. 223-238, 1991.

RUUD, P. A. **Classical econometric theory**. New York: Oxford University Press, 2000.

SCHERBOV, S.; EDIEV, D. Significance of life table estimates for small populations: simulation-based study of standard errors. **Demographic Research**, v. 24, n. 22, p. 527-550, 2011.

SCHRAMM, J. M. A.; OLIVEIRA, A. F.; LEITE, I. C.; VALENTE, J. G.; GADELHA, A. M. J.; PORTELA, M. C.; CAMPOS, M. R. Transição epidemiológica e o estudo de carga de doença no Brasil. **Ciência & Saúde Coletiva**, v. 9, n. 4, p. 897-908, 2004.

SOARES FILHO, A. M.; SOUZA, M. F. M.; GAZAL-CARVALHO, C.; MALTA, D. C.; ALENCAR, A. P.; SILVA, M. M. A.; MORAIS NETO, O. L. de. Análise da mortalidade por homicídios no Brasil. **Epidemiologia e Serviços de Saúde**, v. 16, n. 1, p. 7-18, jan./mar. 2007.

SOUZA, E. R.; LIMA, M. L. C. Panorama da violência urbana no Brasil e suas capitais. **Ciência & Saúde Coletiva**, v. 11, supl., p. 1211-1222, 2007.

STEPHENS, A. S.; PURDIE, S.; YANG, B.; MOORE, H. Life expectancy estimation in small administrative areas with non-uniform population sizes: application to Australian New South Wales local government areas. **BMJ Open**, v. 3, n. e003710, doi: 10.1136/bmjopen-2013-003710, 2013.



SULLIVAN, J. M. Models for the estimation of the probability of dying between birth and exact ages of early childhood. **Population Studies**, v. 26, n. 1, p. 79-97, 1972.

SZWARCWALD, C. L.; LEAL, M. C.; ANDRADE, C. L. T.; SOUZA, P. R. B. Estimaco da mortalidade infantil no Brasil: o que dizem as informaes sobre bitos e nascimentos do Ministrio da Sade? **Cadernos de Sade Pblica**, v. 18, n. 6, p. 1725-1736, nov./dez. 2002.

THATCHER, R.; KANNISTO, V.; ANDREEV, K. The survivor ratio method for estimating numbers at high ages. **Demographic Research**, v. 6, n. 1, p. 1-18, 2002.

TOSON, B.; BAKER, A.; Life expectancy at birth: methodological options for small populations. **National Statistics Methodological Series**, v. 33, p. 1-27, 2003. Available at: <<http://tinyurl.com/toson-baker-e0-small-pops>>. Accessed on: 9 Sep. 2016.

TSIMBOS, C.; KALOGIROU, S.; VERROPOULOU, G. Estimating spatial differentials in life expectancy in Greece at local authority level. **Population, Space and Place**, v. 20, n. 7, p. 646-663, 2014.

UNDP. **Atlas do Desenvolvimento Humano no Brasil**. United Nations Development Programme, 2013. Available at: <<http://www.pnud.org.br/atlas>>. Accessed on: 07 Dec. 2015.

UNITED NATIONS. Population Division. **Manual X: indirect techniques for demographic estimation**. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81, 1983.

VASCONCELOS, A. M. N. A qualidade das estatísticas de bitos no Brasil. **Revista Brasileira de Estudos de Populao**, v. 15, n. 1, p. 115-124, 1998.

WASELFISZ, J. J. **Mapa da violncia 2013: homicdios e juventude no Brasil**. Braslia: Flacso, 2013.

WILMOTH, J. R.; ZUREICK, S.; CANUDOS-ROMO, V.; INOUE, M.; SAWYER, C. A flexible two-dimensional mortality model for use in indirect estimation. **Population Studies**, v. 66, n. 1, p. 1-28, 2012.

## About the authors

*Marcos Roberto Gonzaga* is Ph.D., Demography, Center for Development and Urban Planning (Cedeplar), Federal University of Minas Gerais (UFMG). Professor, Department of Demography and Actuarial Science, and Researcher Associate, Demography Graduate Program (PPGDEM), University of Rio Grande do Norte (UFRN).

*Carl Paul Schmertmann* is Ph.D., Economics, University of California-Berkeley. Professor, Department of Economics, and Research Associate, Center for Demography and Population Health, Florida State University.

## Contact address

*Marcos Roberto Gonzaga*

Universidade Federal do Rio Grande do Norte

Centro de Cincias Exatas e da Terra

Departamento de Demografia e Cincias Atuariais, Campus Universitrio

59078-970 – Natal-RN, Brasil

*Carl Paul Schmertmann*

FSU Population Center

601 Bellamy Building

113 Collegiate Loop

32306-2240 – Tallahassee-FL, United States

## Resumo

*Estimativa de taxas de mortalidade por idade e sexo para pequenas áreas com regressão de TOPALS: uma aplicação para o Brasil em 2010*

A alta variabilidade dos dados nos registros vitais, em razão do baixo número de pessoas expostas, impõe sérios problemas para estimação da mortalidade por idade e sexo em pequenas áreas. Muitas abordagens atuais, incluindo as mais utilizadas no Brasil, estimam as taxas específicas de mortalidade assumindo pressupostos matemáticos rígidos sobre o verdadeiro padrão etário da mortalidade. Padronização indireta, por exemplo, assume que todas as áreas dentro de uma área maior (microrregiões em uma mesorregião, por exemplo) possuem um padrão de mortalidade idêntico, com diferença constante no nível das taxas logarítmicas por idade. Propomos um método estatístico mais flexível que combina regressão Poisson com um modelo relacional denominado TOPALS (DE BEER, 2012). Usamos o novo método para estimar as taxas específicas de mortalidade em pequenas áreas no Brasil (estados, mesorregiões, microrregiões e municípios) em 2010. Resultados para o estado de Minas Gerais mostram diferenças notáveis no padrão de mortalidade por idade entre pequenas áreas adjacentes, demonstrando as vantagens do uso de um método de estimação mais flexível.

**Palavras-chave:** Mortalidade. Pequenas áreas. Método TOPALS. Regressão Poisson.

## Resumen

*Estimación de las tasas de mortalidad por edad y sexo de las pequeñas áreas con regresión de TOPALS: una aplicación a Brasil en 2010*

La alta variabilidad de los datos en los registros vitales, debida al bajo número de personas expuestas al riesgo de morir, plantea serios problemas para la estimación de la mortalidad por edad y sexo en pequeñas áreas. Muchos enfoques recientes, incluyendo los más utilizados en Brasil, estiman las tasas de mortalidad por edad con presupuestos matemáticos rígidos acerca del verdadero padrón etario de la mortalidad. La estandarización indirecta, por ejemplo, asume que todas las áreas dentro de una área mayor (microrregiones de una mesorregión) tengan una idéntica estructura de la mortalidad, con diferencia constante en los niveles de las tasas logarítmicas por edad. Proponemos un método estadístico más flexible que combina la regresión de Poisson con un modelo relacional llamado TOPALS. Utilizamos el nuevo método para estimar las tasas de mortalidad específicas en pequeñas áreas en Brasil (estados, mesorregiones, microrregiones y municipios) en 2010. Los resultados para el estado de Minas Gerais muestran diferencias notables en la estructura de mortalidad entre áreas pequeñas adyacentes, lo que demuestra las ventajas de usar un método de estimación más flexible.

**Palabras clave:** Mortalidad. Áreas pequeñas. Método TOPALS. Regresión Poisson.

Recebido para publicação em 12/06/2016

Recomendado para publicação em 06/12/2016

Aceito para publicação em 10/12/2016