Adult mortality by education level in Brazil:
comparative findings from Census and Vital Registration Deaths

Cássio M Turra (Cedeplar, UFMG), Mirian Ribeiro (UFOP), Fernando Fernandes (Cedeplar, UFMG)

Extended Abstract

Brazil is a country marked by high inequality levels. Regrettably, there is still no high-quality data for correctly measuring socioeconomic differences in adult mortality at the individual level. Research on this topic is at least 50 years late in Brazil than in the U.S. and European countries. We try to fill some of the literature gap by providing new comparative estimates of adult mortality by education using two different methodological approaches: i) data on deaths from the 2010 demographic census with the education level of the deceased members imputed by statistical methods, and ii) data on deaths by education level from death records, extracted from the Mortality Information System (SIM), Ministry of Health.

The Mortality Information System (SIM) has been the primary data source for mortality studies in Brazil. However, there has been a large proportion of missing cases on education hindering the research of SES differentials in mortality. On the other hand, the information about deaths in households, included in the 2010 Census, opened new opportunities in mortality analysis (e.g., Pereira & Queiroz, 2016; Ribeiro, 2016; Silva et al., 2016; Morais, 2019). However, it collects only the sex and age of the deceased, making the study of mortality differentials at the individual level also tricky. Even so, census data must be considered in studies of SES differences in mortality for three reasons. First, deaths and population come from the same source reducing numerator-denominator inconsistencies. Second, it provides a wide range of socioeconomic and geographical information at household and community levels, which may help analyze mortality differentials among individuals. Third, when high-quality data are not available, any data source is valuable.

In a recent article for São Paulo, Ribeiro, Turra & Pinto (2021) offered innovative methodological solutions to solve these issues, which turned to show consistent results. First, following improvements implemented by the Minister of Health that reduced SIM deaths with missing education since 2010, the researchers used imputation methods to treat the remaining missing data. Second, they combined several demographic and SES variables with the Expectation-Maximization technique (E-M algorithm) and a classical Multivariate Regression Model to predict education for reported deaths in the 2010 Census. The results indicated that adult education and mortality are inversely associated in São Paulo, confirming studies conducted in other countries. Also, the alternative prediction method for the 2010 census generated mortality patterns consistent with SIM estimates. Brazil is a vast and heterogeneous country. Therefore, the success of the methodological efforts implemented by the authors reinforced the need for a more extensive study that includes all country regions. In the current article, we test the hypothesis that the patterns initially found for São Paulo vary across the territory. We use the two independent sources of mortality data to improve the robustness of our findings.

Methods

Data

Our analysis includes deaths and person-years of exposure for ages 25 to 59 years. A lower age limit ensures that most individuals have completed school life, avoiding the effects of educational attainment changes on mortality. On the other hand, we exclude the population aged over 60 to mitigate data quality problems in the 2010 Census.

In the first set of estimates, we use deaths reported in the 2010 Brazilian Census. The 2010 census offers nothing about the individual characteristics of the deceased except their age and sex. Also, there is a chance of enumeration issues since deaths are asked at the household level. For example, deaths occurring in households that were extinct throughout the reference period, including single-person households, are not captured. We minimize this issue by excluding the population 60 years and older from our analysis since these are the ages that concentrate most of the single-person households (Wajnman, 2012). Reference period errors and census coverage problems are also potential issues that may impact the quality of mortality estimates.

Since there is no golden standard data in Brazil, we also use deaths from the Mortality Information System (SIM) as a robustness check to verify the consistency of SES patterns in mortality. SIM records are reliable, covering virtually most deaths. However, education information in SIM has been limited due to the excess of missing data. In 2009, there was a change in data collection methodology on education (BRASIL, 2009, 2011). Death certificates began to collect information based on education stages, followed by the completed grade. Since then, the percentage of missing data has decreased gradually. After some assessment, we decided to use SIM data from August 1, 2012, to July 31, 2013. This way, we can minimize the proportion of missing data and keep the SIM dates close to the 2010 census reference period.

Methods

The lack of information on education in the 2010 Census is a result of the data collection design. One straightforward strategy to deal with this limitation is to assume that the deceased's educational level is equal to that of another household member. However, earlier analysis suggests this approach leads to biased results in a country characterized by rapid education transition, such as Brazil (Ribeiro, Turra & Pinto, 2021). Therefore, we predict the education of the deceased statistically based on household and community variables. We consider that missing education in the 2010 Census is at random (MAR). Thus, the reasons for missing information for education in specific analysis units are related to other variables with complete answers, but not to itself (Little & Rubin, 2002; Enders, 2010).

We first fit a model to estimate education by multivariate normal regression (MVN), considering several explanatory variables. To do this, we use census data for surviving respondents. The MVN method uses data augmentation (DA) with Markov chain Monte Carlo (MCMC) iterations. We perform education (Y) imputations for the deceased using the Expectation-Maximization algorithm (EM algorithm). EM is an iterative method for estimating incomplete data (Dempester et al. 1977; Schafer, 1997; Little and Rubin, 2002; Enders, 2010) and is one of the most effective methods for dealing with missing data (Rubin, 1996; Schafer and Graham, 2002; Alisson, 2009). Each subsequent interaction involves an expectation estimation step and a maximization step. The E-M steps consist of i) replace the missing values using the coefficients estimated from the MVN model and assuming a uniform distribution; ii) fit the MVN model again, including the imputed cases, by a likelihood function. We add a random component to create a bootstrap sample of the new coefficients; iii) re-estimate missing values using the parameters estimated in ii; iv) re-estimate coefficients for the MVN model and so forth until the parameter values converge to a stationary distribution. There is no stopping rule that guarantees stationary distribution, and serial dependency may occur. However, as the convergence rate depends on the fraction of missing data, which is low in our study, this is not an issue. Other imputation methods were tested by Ribeiro (2016) using data from São Paulo city, who confirmed that EM was the most appropriate method.

Considering the assumption of normality for the distribution of the dependent variable, our MNV imputation model used education measured in years. After the imputations, we grouped years of schooling into three categories to make education levels in the 2010 Census and the SIM consistent. They are low education level (none to incomplete stage 1 of primary education), medium (complete stage 1 of primary education or more; up to complete secondary education), and high (at least one year of tertiary education).

Concerning deaths from SIM, education data was originally filled for 76.6% of the deceased 25-59 years of age in 2012 and 2013. To deal with the 23.4% of cases with missing information, we perform a simple hot-deck imputation. We assume missing data are completely at random (Rubin, 1996) and use age and sex to redistribute the missing cases.

After getting the data ready, we estimate mortality rates by running Poisson regression models for each data source (Census and SIM). The dependent variable is the logarithm of the mortality rate (incidence rate). The independent variables are age in five-year groups (ages 25 to 59), education level (low, medium, and high), and person-years of exposure (measured as the mid-year period). We fit separate models for men and women.

To increase consistent between the two data sources, we correct the 2010 Census mortality levels using the crude mortality rate by sex estimated based on the SIM. The purpose of this correction was to reduce the under enumeration of deaths in the 2010 Census (coverage and declaration errors). We correct mortality levels but maintain the age and educational patterns of mortality captured by the census collection.

Preliminary Results

Table 1 shows preliminary Poisson models' results, controlling, simultaneously, by age and education level. We offer estimates separately for men and women based only on the 2010 census data. Our preliminary results are limited to the Southeast region of Brazil, which includes Minas Gerais, São Paulo, Rio de Janeiro, and Espírito Santo states. It is the richest (60% of the GDP) and most populous region of Brazil (about 90 million people, 40%). Not surprisingly, according to Table 1, mortality levels increase with age. Also, confirming earlier studies (Ribeiro, Turra, and Pinto 2021), the most significant relative gains are concentrated in the group with medium education. Compared to the lowest education group, mortality rates are about 54 and 39% lower for men and women, respectively. Among women, higher education provides additional mortality protection than among men, but the association is shallow. The results need to be extended for the whole country. More importantly, they need to be scrutinized against SIM records. The final paper will provide all the comparative analysis, improving our knowledge about mortality differentials by SES in Brazil.

Table 1 - IRR - Incidence Rate Ratios, Brazil Southeast region by sex, Census 2010

Variable	Men					Women				
	IRR	Standard error	p-value	95% Confidencial Interval		IRR	Standard error	p-value	95% Confidencial Interval	
age										
30-34	0.990	0.0496	0.8380	0.897	1.092	1.292	0.1082	0.002	1.096	1.522
35-39	1.114	0.0551	0.0290	1.011	1.228	1.550	0.1266	0.000	1.321	1.819
40-44	1.422	0.0669	0.0000	1.297	1.559	2.530	0.1907	0.000	2.183	2.933
45-49	1.847	0.0835	0.0000	1.691	2.018	3.285	0.2407	0.000	2.846	3.792
50-54	2.522	0.1102	0.0000	2.315	2.748	4.950	0.3514	0.000	4.307	5.689
55-59	3.365	0.1447	0.0000	3.093	3.661	6.044	0.4291	0.000	5.259	6.947
education										
medium	0.458	0.0124	0.0000	0.434	0.483	0.611	0.0215	0.000	0.570	0.655
high	0.549	0.0202	0.0000	0.511	0.590	0.566	0.0273	0.000	0.515	0.622
constant	0.004	0.0001	0.0000	0.004	0.004	0.001	0.0001	0.000	0.001	0.001

References

Allison P. D. Missing Data. 2009. Millsap R. E., In: Maydeu-Olivares A. The SAGE Handbook of Quantitative Methods in Psychology. SAGE Publications. P. 72-89.

Brasil, 2009. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde. Legislação Relativa aos Sistemas de Informações sobre Mortalidade (SIM) e sobre Nascidos Vivos (SINASC). Brasilia: 2009.

Brasil, 2009. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde. Manual de Instruções para o Preenchimento da Declaração de Óbito. Série A. Normas e Manuais Técnicos. Brasilia: 2011.

Dempster A. P., Laird N. M, Rubin D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 39(1): 1-38.

Enders, C. K. 2010. Applied Missing Data Analysis. The Guilford Press, New York.

Little R. J. A., Rubin D. B. 2002. Statistical Analysis with Missing Data. 2 ed. Hoboken, NJ: Wiley, 2002.

Morais C. C. B. 2019. Diferenciais de mortalidade adulta por nível de renda no estado de São Paulo: variações por sexo e fase do ciclo de vida, a partir de três estratégias de imputação. Dissertação (Mestrado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte.

Pereira F. N. A., Queiroz, B. L. 2016. Diferenciais de mortalidade jovem no Brasil: a importância dos fatores socioeconômicos dos domicílios e das condições de vida nas localidades. Cadernos de Saúde Pública.

Ribeiro, M. M. 2016. Mortalidade adulta por níveis de escolaridade no estado e no município de São Paulo: uma proposta de estimação a partir do Censo Demográfico de 2010. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte.

Ribeiro M. M, Turra, C. M., Pinto, C. C. X. 2021.Adult mortality by education level in São Paulo: comparative analysis based on different methodological strategies. Revista Brasileira de Estudos de População. In press.

Rubin, D. B. 1996. Multiple imputation after 18+ years. Journal of the American Statistical Association. 91(434): 473-489.

Schafer J. L. 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall: London, 1997.

Schafer J. L., Graham J. W. Missing data: our view of state of the art. Psychological Methods, 7: 147-177, 2002.

Silva, L. E., Freire, F. H. M. A, Pereira, R. H. M. 2016. Diferenciais de mortalidade por escolaridade da população adulta brasileira. Cad. Saúde Pública, Rio de Janeiro, 32(4): 1-12.

Wajnman, S. 2012.Demografia da família e dos domicílios brasileiros. Dissertation (Full Professor). Belo Horizonte: Cedeplar, 2012.