



All the contents of this journal, except where otherwise noted, is licensed under a Creative Commons Attribution License

Adult mortality by education level in São Paulo: comparative analysis based on different methodological strategies*

Mirian Martins Ribeiro**
Cassio Maldonado Turra***
Cristine Campos de Xavier Pinto****

In this article, we estimate adult mortality by education level in São Paulo. We compare estimates based on deaths from the 2010 Census and the 2013 Mortality Information System (Sistema de Informação de Mortalidade – SIM) – DATASUS, and three different ways of measuring education level: recorded in the SIM, reported in the census for the household heads and imputed statistically in the census for individuals who died. For the statistical imputation, we use the Dempester (1977) method, which proposes using the expectation-maximization algorithm (EM algorithm) to deal with missing data. We consider three education levels (low, medium, and high) and estimate mortality rates based on Poisson models. The results indicate that between ages 25 and 59, more years of schooling are associated with mortality rates up to 77% lower. Secondary (medium) education level provides most of the mortality gains at adult ages (about 50%). The mortality differentials calculated with death records from the SIM and census deaths with education imputed statistically are similar. However, estimates based on the assumption that the deceased's education is equal to the household head's in the census resulted in atypical mortality patterns. We hope that the imputation model we propose in the current study can be used in future mortality analyses by SES using census deaths.

Keywords: Adult mortality. Missing data imputation. Socioeconomic differentials in mortality. Brazil. Education level.

* This study was supported by the Brazilian Graduate Studies Coordinating Board (Capes, Code 001), which funds the Demography Program at the Federal University of Minas Gerais; and the Brazilian National Research Council (CNPq). We thank Simone Wajnman and two anonymous referees for their insightful suggestions in earlier versions of the study.

** Instituto de Ciências Sociais Aplicadas, Universidade Federal de Ouro Preto (Ufop), Ouro Preto-MG, Brazil (miriannina@gmail.com; <https://orcid.org/0000-0002-4256-220X>).

*** Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte-MG, Brazil (cmturra@gmail.com; <https://orcid.org/0000-0003-4051-3567>).

**** Escola de Economia de São Paulo (EESP), Fundação Getúlio Vargas (FGV), São Paulo-SP, Brazil (cristine.pinto@fgv.br; <https://orcid.org/0000-0003-2955-9759>).

Introduction

This study aims to estimate adult mortality by education level in São Paulo based on different methods. In Brazil, data for accurately estimating socioeconomic differences in adult mortality at the individual level are rare. Therefore, our goal is to offer comparative estimates using three different methodological approaches: deaths reported in the 2010 demographic census assuming that educational attainment of the deceased household members is equal to the educational attainment of household heads; data on deaths from the 2010 demographic census with the education level of the deceased members imputed by statistical methods; and data on deaths by education level from death records, extracted from the Mortality Information System (SIM), Ministry of Health. Measuring SES differentials in mortality accurately is necessary for a country marked by high inequality levels such as Brazil.

The Mortality Information System (SIM) has been the primary data source for mortality studies in Brazil. Still, its records are not free from errors in coverage and content and must be compared with alternative data sources. The information about deaths in households, included in the 2010 Census, opened new doors in mortality analysis. Since then, scholars have gradually examined deaths reported in the Brazilian Census, despite the lack of the characteristics of the deceased, except for age and sex. For example, one study has investigated the role of household socioeconomic characteristics in youth mortality (PEREIRA; QUEIROZ, 2016). At least three others have looked at adult mortality differentials according to the household head's education level (RIBEIRO, 2016; SILVA *et al.*, 2016) and household income (MORAIS, 2019).

The present study contributes to the existing literature in at least three ways. The first innovation is the research on educational differentials of adult mortality using data from the 2010 Census, based on the education level of the deceased individuals, instead of using household proxies. We propose the imputation of missing education using the Expectation-Maximization technique (EOM algorithm) and the Classical Multivariate Regression Model in the census sample of São Paulo. We selected São Paulo to test the viability of the statistical imputations and the consistency of mortality estimates across data sources because it is the country's largest city. In the future, the same methodology can be reproduced in studies on SES differentials in mortality for the total Brazilian population to capture heterogeneity across municipalities, states, and regions. Also, by comparing the proposed imputation strategy with rates estimated based on the household heads' education level, a method used in at least one earlier study (SILVA *et al.*, 2016), our research contributes to the debate on how to measure SES differences in adult mortality when using census data. Finally, our study is innovative in comparing mortality estimates based on census data with rates calculated from deaths recorded in the 2013 SIM. Historically, there has been a large proportion of missing cases on education in the SIM, hindering its use for that purpose. In 2013, the

proportion of deaths with missing education was shallower, following improvements implemented by the Minister of Health since 2010.

Our results indicate that adult education and mortality are inversely associated in Brazil, confirming studies conducted in several other countries. However, the results vary depending on the definition of education level used. The use of household information as a *proxy* for individual characteristics can result in inconsistent estimates. The alternative imputation method that we propose independently generated mortality patterns consistent with SIM estimates, reinforcing its relevance for future analyses.

Education and adult mortality

Main theoretical models

According to Crimmins (1993), demographic studies lack a unified theory about socioeconomic differentials in mortality. Two of the most discussed views are the economic and psychosocial models (PRESTON; TAUBMAN, 1994). Many economic models, which are usually more structured to associate health and socioeconomic conditions (PRESTON; TAUBMAN, 1994), are inspired by Grossman's health demand model (1972). According to this approach, individuals make decisions that help them build health capital. Death occurs when stock reaches a value below critical level. There are two functions involved in this process: the budgetary function and the health production function. The health production function indicates how health capital varies according to different inputs, including the number of health goods consumed, the amount of medical knowledge or technology available to the population, the availability of medical services, genetic endowment, and the characteristics of the environment in which individuals live. While individual choices determine the consumption of health goods, the other inputs are defined exogenously (ROSENZWEIG; SCHULTZ, 1983; PRESTON; TAUBMAN, 1994). As education affects the level of human capital, individuals with higher education attainment have a higher income, which ensures greater resources for the consumption of goods and services related to health production: healthy food, greater living spaces, less polluted living areas, better quality health services, and others. Furthermore, education directly affects health levels throughout the life cycle, as it improves the ability of individuals to discern and the way they process information which, in turn, affects health conditions and the risk of death (PRESTON; TAUBMAN, 1994; SICKLES; TAUBMAN, 1997; GROSSMAN, 2008).

The prices of health products and services are also central to economic models, as they influence mortality differentials through at least two mechanisms. The first is that even if models assume that market prices are the same for everyone, access to health goods and services through public programs or health insurance can minimize differences in consumption between individuals with different education levels. The second mechanism is the opportunity cost for obtaining medical care, which depends on individuals' income

and waiting time for care. Even among low-income individuals, the opportunity cost can rise if the waiting time is too long. Some public policies that improve access to health services can change relative prices (PRESTON; TAUBMAN, 1994), reducing mortality differentials. One of the main limitations of economic models is that they do not incorporate environmental factors (such as the incidence of infectious diseases). The absence of these factors may alter the population's health capital and the role of social context in shaping individual preferences and habits (PRESTON; TAUBMAN, 1994; WILLIAMS, 1990).

Psychosocial models focus on factors neglected by economic models. Although the literature that investigates the role of social relations in health dates from the 19th century, interest was renewed in the 1970s with the emergence of the social support theory. The topic has been boosted by the report of an advisory committee of the United States Public Health Services (US PUBLIC HEALTH SERVICES, 1964). It associated morbidity and mortality due to several causes with cigarette consumption (HOUSE *et al.*, 1988) but considered individual behavior, lifestyle, and attitudes independent of social structure. Contrarily to that, psychosocial theories start from the premise that social structures shape values and influence individuals' behavior concerning the restrictions they must comply with (WILLIAMS, 1990).

Starting in the early 1990s, psychosocial models began to incorporate socioeconomic characteristics (PRESTON; TAUBMAN, 1994; HOUSE *et al.*, 1990; WILLIAMS, 1990) and showed that psychosocial factors and health services mediate the relationship between socioeconomic factors and health. According to the models, morbidity and mortality are associated with psychosocial and behavioral factors, such as stress, religious practice, marital status, eating habits and physical exercise, smoking, and social support. One example is marital status. Married persons, especially men, have a higher risk of dying than unmarried ones (ROGER, 1995; LILLARD; WAITE, 1995). Another example is religion. People who are heavily involved in religious precepts have lower mortality than those less involved, probably due to social ties and support (ROGERS, 1996). Stress, measured by variables such as unemployment, divorce, death of loved ones, also seems to have a positive relationship with several diseases. Finally, social ties and personal control are factors that can act on biological processes, affect vulnerability to disease and influence the chances of death (WILLIAMS, 1990).

Psychosocial factors, in turn, vary according to individuals' socioeconomic characteristics, such as income, occupation, and education level. Obesity and the number of cigarettes smoked, for example, are negatively associated with education. Between 1974 and 1985, the prevalence of smokers in the United States declined five times faster among people with tertiary education than people with lower education attainment (NCHS, 1981). Indeed, smoking is considered one explanation for increased educational differentials in mortality between the 1980s and 1990s. Education explains about a quarter of the rise in differentials in the practice of smoking for women (MEARA *et al.* 2008). In addition, Pampel *et al.* (2010) indicated the influence of socioeconomic factors, especially education level,

on individuals' lifestyle, mainly affecting tobacco use, physical exercise, and eating habits. Although the consumption of alcoholic beverages is positively associated with education level, alcoholism is negatively associated with education level (NCHS, 1981).

Education level provides individuals with a stock of knowledge that influences how they interact with others while intervening in crises and emergencies, searching for medical goods and services, preventing diseases, and promoting health. More educated individuals are more likely to adopt healthy behaviors, seek medical care, perceive health problems, and communicate better with doctors, thus understanding what is said and following medical recommendations. In contrast, less educated individuals are more susceptible to negative psychosocial characteristics from the perspective of health. They are more likely to adopt risky behaviors, experience stress, social isolation, and lack of social control, which are important determinants of morbidity and mortality (ROGERS *et al.*, 1999b).

The main criticism of the psychosocial approach is the absence of a unified theory that explains the interactions between socioeconomic factors, psychosocial factors, and the determinants of mortality. The models indicate the patterns and trends but with insufficient data to go beyond speculation about causal relationships.

Differences in adult mortality by education level

The growing interest in the socioeconomic differences in mortality dates from the beginning of the 19th century. Until the middle of the 20th century, studies mainly used occupational status as an SES measure (ANTONOVSKY, 1967; ELO, 2009). From the 1960s onwards, education level became the socioeconomic dimension more frequently used. It is a variable with a lower degree of reverse causation with mortality, least affected by the proximity to death, which remains relatively stable throughout adult life, and plays a significant role in determining occupation, income, and wealth (KITAGAWA; HOUSE, 1963; PRESTON; TAUBMAN, 1994; ELO; PRESTON, 1996; MONTEZ *et al.*, 2012). In addition, data quality is higher for education than other SES variables (LIBERATOS *et al.*, 1988; MONTEZ *et al.*, 2012), although it is not error-free (NEPOMUCENO; TURRA, 2020). In high-income countries, especially the United States and the United Kingdom, education has been a central measure when monitoring socioeconomic inequalities in mortality (CUTLER; LLERAS-MUNEY, 2007).

The pioneering study by Kitagawa and Houser (1973), conducted in the United States, is a milestone in the field. Using data from the 1960 death registry, the authors found substantial adult mortality differentials by education, especially among the working-age population. In the following decades, other studies also found an educational gradient in mortality (ELO; PRESTON, 1996; ROGERS *et al.*, 1999b; ROSS *et al.*, 2012) and an increase in differentials over time (PAPPAS *et al.*, 1993, ELO; PRESTON, 1996).

In the 21st century, the topic remains relevant on the US research agenda. Until the end of the 2000s, the growth in mortality differences was attributed to a more accelerated

fall in mortality among the most educated than among the less educated (PAPPAS *et al.*, 1993; PRESTON; ELO, 1995). However, recent studies have also shown an increase in mortality for the less-educated population groups (MONTEZ *et al.*, 2011; HAYWARD *et al.*, 2015). Sasson (2016) indicated that, between 1990 and 2010, life expectancy at 25 years of age for less-educated women decreased by about three years, mainly due to increased mortality between ages 45 and 64. Conversely, there were considerable improvements in the life expectancy of the most educated. Similarly, Case and Deaton (2015) pointed to an increase in mortality among non-Hispanic white adults aged 45 to 54 years from 1999 to 2013, which is explained by the rise in mortality from external causes, especially intoxication by drugs, alcohol, and suicide. These deaths are concentrated among individuals with secondary or lower education.

In high-income European countries, such as Denmark, Norway, Sweden, Finland, England, and Wales, educational differentials in mortality are similar, although mortality levels vary across countries (VALKONEN, 1989). On the other hand, transition economies have the greatest educational differentials in mortality (MACKENBACH *et al.*, 2008; SHKOLNIKOV *et al.*, 1998; DENISOVA, 2009; LEINSALU, 2003) and experienced an increase in inequality since the 1990s, with the end of the socialist regime (SHKOLNIKOV *et al.*, 1998; LEINSALU, 2003). There is also ample evidence of educational differentials in mortality in other regions of the world, including Asia (LIANG *et al.*, 2000; SUBRAMANINAN *et al.*, 2006; HURT *et al.*, 2004), Africa (BERHANE *et al.*, 2002; WALQUE; FILMER, 2013), and Latin America and the Caribbean (KOCH *et al.*, 2007; MANZELLI, 2014; SANDOVAL; TURRA, 2015).

Data quality problems have been an obstacle to advancing studies on educational differentials in the adult population in Brazil. Despite the limited information, recent efforts generated advances based on indirect techniques, imputation methods, and education data reported at the individual level (RENTERIA; TURRA, 2008; RENTERIA, 2010; GUEDES *et al.*, 2011; TURRA; RENTERIA; GUIMARÃES, 2016; NEPOMUCENO; TURRA, 2020; TURRA *et al.*, 2018). Renteria (2010), for example, estimated mortality for the Brazilian female population aged 20 to 69 using respondents' mothers' survival status and education, collected by IBGE household surveys (PNAD and PPV). Following the orphanhood method, she found decreasing mortality rates with education level. The differences were significant, and non-educated women had a mortality rate up to 70% higher than those who studied between 1 and 8 years.

Additionally, Silva *et al.* (2016) investigated mortality differentials by education for people aged 15 to 59 and 60 to 80 years old using data from the 2010 Census. The authors assumed the education of the deceased household members was equal to household heads' education to fix the missing information on SES for deaths. They found that probabilities of death in the younger age groups were inversely associated with education level. In addition, where household heads had tertiary education, the estimated mortality was lower than for those with no schooling or incomplete primary education. However, the authors found a reversed gradient among individuals aged over 70: a positive association between mortality

and education, except for the Southeast region. This finding may be a consequence of the methodology for imputing missing education since it relies on the assumption of no intergenerational mobility in Brazil.

Despite the recent research efforts in Brazil, it is necessary to keep improving the information available by investing in data quality, matching different databases, and estimating new mortality measures. Only in this way can we measure mortality differences in Brazil correctly. This study's primary goal is to contribute by providing new methodological alternatives and estimates.

Methodology

We estimated adult mortality rates by education level in the city of São Paulo using Poisson regression models and three different sets of information: deaths reported in the 2010 demographic census assuming that the education level of the deceased is equal to the education level of household heads; data on deaths from the 2010 demographic census with the education level of the deceased imputed by statistical methods; and data on deaths by education level extracted from the Mortality Information System (SIM) from the Ministry of Health – DATASUS.

The analysis includes the population aged between 25 and 59 years living in the city of São Paulo in 2010. Adopting a lower age limit is necessary to ensure that most individuals have completed school life. It avoids the effects of educational attainment changes on mortality estimates. We excluded the population aged over 60 to mitigate data quality problems in the 2010 Census. In the following sections, we detail the data sources and methods used in each of the information sets.

Deaths in the 2010 Brazilian Census

We used data from the 2010 Demographic Census for the first two sets of estimates.¹ Census data are still an unusual source for directly estimating mortality in Brazil since IBGE collected household deaths in the main questionnaire only in 2010. One of the advantages of census data is that it is possible to work with the same source for deaths and population. Another advantage is that the Brazilian Census provides a wide range of socioeconomic and geographical information at household and community levels, which are necessary to estimate mortality differentials. However, in the census, nothing is known about the deceased's socioeconomic characteristics, at the individual level, except their age and sex. Also, there is a greater chance of enumeration issues if deaths come from the census rather than vital statistics. As the question is asked at the household level, deaths occurring in households that were extinct throughout the reference period, including single-person households, are not captured. Deaths may also be reported more than once if the deceased

¹Data were extracted using the *Stata 14.0* program, through the *Datazoom* platform, which was developed by the Department of Economics at PUC-Rio.

person has moved to more than one home during the reference period. Finally, reference period errors by respondents can also lead to underestimation of deaths. As mentioned above, we opted to exclude the population over 60 years old to minimize the enumeration problems partially since there is a greater concentration of single-person households among older people in Brazil (WAJNMAN, 2012).

Imputation of education level in Census data

In general, in a data matrix in which rows represent units of analysis and columns represent variables, the entries for which there are no data are considered missing data (LITTLE; RUBIN, 2002). However, many problems do not necessarily involve data loss. This is the case of the 2010 Census since the lack of information on education level of the deceased is a result of the data collection design. The estimation of mortality rates based on the data on deaths of the 2010 Census requires that education information be imputed for all deaths. But even in this case, the information can be reformulated and treated with methods initially developed to deal with missing data (HEITJAN; RUBIN, 1990; HEITJAN; RUBIN, 1991; SCHAFFER, 1997; ENDERS, 2010; MISTLER; ENDERS, 2012; LITTLE; RHEMUTULLA, 2013).

A more straightforward strategy to deal with the 2010 Census limitation is to assume that the deceased's educational level is equal to that of another household member. We used this strategy in our first set of estimates since it was previously applied in the literature. It is time to scrutinize it in the light of our novel results. 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). This categorization follows the methodology applied by the DATASUS in SIM, as discussed posteriorly.

Our second strategy was to impute the missing values statistically based on household and community information. The mechanisms of connection between the missing data and their determinants are essential for choosing the treatment method to avoid significant bias in the expected estimates. Rubin (1976)² created a classification system based on three hypotheses about the missing data factors. Missing data are classified as: i) missing completely at random (MCAR); ii) missing at random – MAR); and iii) missing not at random (NMAR) (LITTLE; RUBIN, 2002; ENDERS, 2010). We can consider the missing education level in the 2010 Census as MAR. The reasons for missing information for Y information in specific analysis units are related to other variables with complete answers, but not to Y itself.

Education (Y) imputations were performed using the Expectation-Maximization algorithm (EM algorithm). EM is an iterative method for estimating incomplete data, formalized and discussed by Dempster *et al.* (1977) and described and detailed by Schafer (1997), Little and Rubin (2002), and Enders (2010). It is one of the most effective methods

² Rubin (1976) developed a theory about missing data that has become a reference in the literature.

for dealing with missing data, even under the hypothesis of NMAR (RUBIN, 1996; SCHAFER; GRAHAM, 2002; ALISSON, 2009). Other imputation methods were tested by Ribeiro (2016), who confirmed that EM was the most appropriate method.

Each subsequent interaction involves an expectation estimation step (step E) and a maximization step (step M). Taking the log-likelihood function $l(\theta | Y)$ as the natural logarithm of the likelihood function $L(\theta | Y)$, the E-M steps consist of: i) replacing the missing values using some simpler method (imputation by the mean or the assumption of a uniform distribution, for example); ii) estimating the needed parameters by a likelihood function $l(\theta^{(0)} | Y_{\text{obs}})$; iii) re-estimating missing values using the parameters estimated in ii; iv) re-estimating parameters $l(\theta^{(1)} | Y_{\text{obs}})$ and so forth, $l(\theta^{(2)} | Y_{\text{obs}}) \dots l(\theta^{(t)} | Y_{\text{obs}})$ until the value of θ reaches convergence in the t -th interaction.³

We estimated the model for applying the EM method by multivariate normal regression (MVN), considering the explanatory variables in Table 1. The selection of variables does not follow a cause-and-effect relationship with the imputed variable. For imputation purposes, the main issue is to incorporate factors highly correlated with the missing data. As the sample of the 2010 Census is large enough for more inclusive models, the strategy was to include as many variables as possible that were significantly associated with education. The selection of variables is explained in greater detail in Ribeiro (2016). The MVN method uses data augmentation (DA) with Markov chain Monte Carlo (MCMC) iterations. Although data are estimated using maximum likelihood, we added a random component to create a bootstrap sample. First, we calculated the expected values for the parameters that describe the relationship between the independent variables and education, starting with a uniform distribution, using the EM algorithm. The uniform distribution assumes that all coefficient values are equally likely. Then, starting with bootstrap samples, we obtained the expected values for the missing data. This method is iterative, and imputations were performed using the Markov chain Monte Carlo (MCMC). The missing data are filled when 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 was not an issue. 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 education into the three categories already mentioned: low, medium, and high education.

³ For more details on the imputation procedure, see Ribeiro (2016).

TABLE 1
Variables selected for the regressions

Level/classification	School/description	Measure/grouping
Dependent variable	Education level	Years of study
Deceased individual	Age group	25-29 years
		30-34 years
35-39 years		
40-44 years		
45-49 years		
50-54 years		
	55-59 years	
	Sex	1 if man, 0 if woman
Characteristics of household heads	Sex of household heads	1 if man, 0 if woman
	Education level of household heads	Years of study
	Age of household heads	<30 years
		30-39 years
		40-49 years
		50-59 years
		>=60
	Marital status of household heads	1 if they have a spouse, 0 if not
Does the household head have at least one major disability?	1 if yes, 0 if not	
Race/ethnic group of household head	White	
	Black	
	Asian	
	Brown	
	Indigenous	
Difference between the age group of household head and the individual (adult aged between 25 and 59 years)	Difference > - 10 & < 10 years	
	Household head is 10 to 19 years younger	
	Household head is 20 years or more than 20 years younger than the individual	
	Household head is 10 to 19 years older	
	Household head is 20 years or more than 20 years older than the individual	

(continue)

(continued)

Level/classification	School/description	Measure/grouping
Household	Indicator of socioeconomic status of household composed of the variables: - presence of tv to in the household - presence of a car in the household - presence of motorcycle in the household - presence of washing machine in the household - presence of refrigerator in the household - presence of running water in at least 1 room - presence of a computer with internet in the household - presence of mobile phone in the household - predominant permanent material (masonry) - if the house is adequate (having water supply through a general network, sewerage through a general network or septic tank, direct or indirect garbage collection, and a maximum of two residents per bedroom) - at least 1 resident of the household is a beneficiary of an income transfer program (bolsa família; PETI or other)	4 categories – quartiles
	Dependency ratio (<14 or > 60)/(>15 and <59)	1 if >1, 0 if <1
	Occurrence of infant death in the household	1 if yes, 0 if not
	Households composed exclusively of foreigners or naturalized individuals or include Brazilian natives?	1 if include any Brazilian native; 0 if is composed of only foreigners or naturalized individuals
	Proportion (%) of people living in subnormal areas within the weighted area	None <50 >50 and < 90 >90
Community	Metropolitan area	1 if yes, 0 if not
	Location of household	1 if urban, 0 if rural
	Education level (educ) of a randomly Selected individual (bootstrap sampling)	

Deaths from the Mortality Information System (SIM)

In the third set of estimates, we used data from the Mortality Information System (SIM). The goal was to verify the consistency of the two sets of results estimated based on the census data. We used deaths that occurred between August 1, 2012, and July 31, 2013.

SIM records for São Paulo are exceptionally reliable, covering virtually 100% of deaths (QUEIROZ; SAWYER, 2012). However, education level information in SIM has been considered low quality, mainly due to the excess of missing data. In 2009, there was a change in the methodology of data collection on education. Death certificates began to collect information based on education stages (no education, elementary school, middle school, high school, higher education), followed by the completed grade. Although the new methodology was implemented in 2011 (BRASIL, 2009, 2011), the health system still

uses both declaration models (old and new). Therefore, the data continue to be publicly available only in the old format: 0, 1 to 3, 4 to 7, 8 to 11, and 12 or more years of study. Since then, the percentage of missing data remained relatively high at the state level, but it has decreased for some state capitals. In São Paulo city, the education level was reported for 92.3% of adults who died between 25 and 59 years of age, from August 2012 to July 2013. The low proportion of missing data makes SIM a viable data source for investigating mortality differentials by education level and assessing census estimates' consistency in the city. To deal with the 7.7% cases with missing data on education in the SIM, we performed a simple hot-deck imputation. We assumed that missing data were MCAR and used age and sex to redistribute missing cases. More details about applying the hot-deck with SIM data can be found in Ribeiro (2016). As mentioned earlier, after the data corrections, we grouped education into low (0-3 years of education), medium (4-11 years), and high (at least one year of tertiary education) levels.

The Poisson model for estimating mortality rates

The Poisson regression model is often used to calculate mortality rates (COLEMAN, 1964). It allows for the estimation of the number of deaths according to the risk of death and exposure. The incidence rate at which the event occurs, once multiplied by the exposure time, results in the number of occurrences. The model is given by

$$\log Y = \beta X_i + \log[\text{exposition}] \quad (1)$$

where the dependent variable is the logarithm of the mortality rate (incidence rate), β is a vector of coefficients estimated by maximum likelihood, and X is a matrix of independent variables. In the present study, X is represented by age in five-year groups (ages 25 to 59) and by education level measured according to the three categories mentioned above (low, medium, and high).

We did not include an interaction term between age and education because of the low number of deaths for groups with higher education and younger ages (see Table 2). Therefore, we ran separate models by sex, considering only the main effects of age and education level. The models allow us to examine differences in mortality levels but hinder the investigation of age patterns of mortality by education.

After fitting the Poisson models separately for men and women, we calculated the mortality rates. As indicated at the beginning of this section, our first model uses data on deaths from the 2010 Census, assuming that the deceased's education level was equal to the household head's education level. The second model also uses data on deaths from the 2010 Census, but with the deceased's education level imputed by the statistical methods described previously. Finally, the third model uses data on deaths by education level extracted from SIM-DATASUS.

We used the mid-period population in all estimates to approximate person years lived, assuming that deaths were evenly distributed throughout the calendar year. Although it

would be possible to calculate time of exposure of the deceased from the month the death occurred, our models' results would be similar whatever the strategy adopted. Thus, we opted for the most parsimonious form of calculation that also partially offsets the migration effects in São Paulo during the analysis period.

For the estimates using SIM deaths, we used aggregated data to estimate the exposed population since population and deaths are not in one single data source. Following the same line of thought used for the Census, we projected the population of July 1, 2012, estimated by IBGE, for the mid-period. We applied in 2013 the 2010 Census distribution by sex, age, and education level, assuming that three years is sufficiently short to ensure there have been no significant changes in these distributions.

As a final step, we corrected 2010 Census mortality levels using the crude mortality rate by sex estimated based on the SIM. The purpose of this correction was to reduce any problems with the enumeration of deaths in the 2010 Census (coverage and declaration errors). We corrected levels but maintained the age and educational patterns of mortality as captured by the census collection.

Results

Table 2 presents a descriptive analysis of the data used, through crude mortality rates (CMR) by sex, calculated from both the 2010 Census and SIM-2013. The overall mortality level is higher in SIM-2013 than in the 2010 Census, respectively 14% and 18% for men and women. The estimated confidence intervals for the 2010 Census data suggest that the two databases' estimates are statistically different and reinforce the need to correct the Census's mortality level.

TABLE 2
Crude mortality rates (CMR) per thousand, population aged 25-59 years
Municipality of São Paulo – 2010-2013

Sex	Census – sample		Census – weighted sample	SIM	SIM/Census ratio
	CMR	CI (95%)	CMR	CMR	
Women	1.69	[1.50 ; 1.92]	1.70	2.01	1.18
Men	3.79	[3.47 ; 4.14]	3.82	4.35	1.14
Total	2.68	[2.50 ; 2.88]	2.70	3.11	1.15

Source: IBGE. 2010 Population Census; Mortality Information System (SIM), 2013.

Table 3 shows the distribution of deaths according to the different sources and methods used. The distribution of deaths by education level statistically imputed in the 2010 Census is like the SIM distribution. The number of deaths in the 2010 Census is higher at older ages, making the two distributions approximate. This result is related to the variation in the consistency of imputations by age. Conversely, the distribution of the 2010 Census

deaths imputed according to household heads' education level is more concentrated in the middle education level group. At younger ages, the distribution of education measured by the education level of household head favors the lowest level when compared to the other two methodologies. The pattern is the opposite at older ages, giving more relative weight to the medium educational level. These patterns are explained by changes in the household composition throughout the life cycle and the distribution of education level across generations in Brazil, as will be argued in section Discussion.

TABLE 3
Percentage distribution of deaths by age, sex, education level, and estimation method
Municipality of São Paulo – 2010-2013

Age	Men						Women					
	Low	Medium	High	Total	N	Weighted sample	Low	Medium	High	Total	N	Weighted sample
Education of household heads, deaths from the 2010 Census												
25-34	57.1	33.3	9.5	100.0	42	897	41.2	52.9	5.9	100.0	17	335
30-34	42.1	55.3	2.6	100.0	38	754	33.3	38.9	27.8	100.0	18	358
35-39	55.1	32.7	12.2	100.0	49	1,030	35.7	64.3	0.0	100.0	14	293
40-44	54.3	38.6	7.1	100.0	70	1,466	31.3	50.0	18.8	100.0	32	642
45-49	52.1	38.0	9.9	100.0	71	1,514	45.7	41.3	13.0	100.0	46	932
50-54	50.5	40.2	9.3	100.0	107	2,205	43.3	45.0	11.7	100.0	60	1,269
55-59	43.3	39.2	17.5	100.0	120	2,536	41.3	38.1	20.6	100.0	63	1,286
Total	49.9	39.2	10.9	100.0	497	10,403	40.4	44.4	15.2	100.0	250	5,115
Education statistically imputed, deaths from the 2010 Census												
25-34	33.3	33.3	33.3	100.0	42	897	17.6	41.2	41.2	100.0	17	335
30-34	44.7	34.2	21.1	100.0	38	754	27.8	44.4	27.8	100.0	18	358
35-39	51.0	28.6	20.4	100.0	49	1,030	64.3	7.1	28.6	100.0	14	293
40-44	58.6	30.0	11.4	100.0	70	1,466	37.5	31.3	31.3	100.0	32	642
45-49	56.3	35.2	8.5	100.0	71	1,514	45.7	39.1	15.2	100.0	46	932
50-54	63.6	22.4	14.0	100.0	107	2,205	56.7	31.7	11.7	100.0	60	1,269
55-59	72.5	21.7	5.8	100.0	120	2,536	69.8	20.6	9.5	100.0	63	1,286
Total	58.8	27.6	13.7	100.0	497	10,403	51.2	30.4	18.4	100.0	250	5,115
Education and deaths from the 2013 SIM (death records)												
25-34	43.7	45.8	10.5	100.0	911	-	32.8	48.1	19.1	100.0	293	-
30-34	49.4	38.5	12.1	100.0	1,014	-	36.6	42.2	21.2	100.0	415	-
35-39	61.2	28.6	10.2	100.0	1,127	-	43.0	38.1	18.9	100.0	535	-
40-44	63.5	28.8	7.7	100.0	1,490	-	54.1	33.1	12.8	100.0	712	-
45-49	61.3	28.2	10.5	100.0	1,892	-	53.9	30.6	15.5	100.0	1,066	-
50-54	61.7	26.9	11.4	100.0	2,576	-	58.0	27.7	14.3	100.0	1,486	-
55-59	61.8	27.1	11.1	100.0	3,341	-	63.9	22.8	13.3	100.0	1,930	-
Total	59.5	29.9	10.6	100.0	12,351	-	54.9	30.0	15.1	100.0	6,437	-

Source: Based on data from the 2010 Census and SIM/DATASUS 2012-2013.

Table 4 shows the Poisson models' results, controlling, simultaneously, by age and education level. We offer estimates separately for men and women and each data set. In all models, mortality levels increase with age. However, the ratios between the rates in SIM-2013 suggest relatively higher mortality levels among men and women aged 55-59 than

in the 2010 Census. The IRR estimators are less accurate when measured by age because of the small number of cases, particularly for women (Table 3). Although the sample size in the 2010 census data hinders the estimation of correct mortality levels, we believe the gradient by education is not affected. As previously mentioned, differences between the two data sources (SIM and Census) may be also a consequence of coverage issues or the under-enumeration of deaths in the 2010 Census. This type of error would be more frequent among individuals who live alone, more prevalent at older ages. It may also happen at other ages if households were omitted during the data collection.

TABLE 4
IRR – Incidence Rate Ratio by sex according to sources or methods
Municipality of São Paulo – 2010-2013

Variable	Model 1 – Deaths from the 2010 Census, education level of household heads				Model 2 – Deaths from the 2010 Census, education statistically imputed				Model 3 – Deaths from the 2013 SIM
	IRR	Standard error	CI-95%		IRR	Standard error	CI-95%		
Men									
<i>Age</i>									
25-29	-		-	-	-		-	-	-
30-34	1.09	0.26	0.68	1.74	1.06*	0.25	0.66	1.69	1.12*+
35-39	1.41	0.32	0.91	2.19	1.35*	0.30	0.87	2.09	1.31*+
40-44	2.25	0.47	1.50	3.38	2.12*	0.43	1.41	3.16	1.82*+
45-49	2.46	0.51	1.64	3.69	2.29*	0.47	1.53	3.41	2.62*+
50-54	4.60	0.92	3.11	6.82	4.22*	0.85	2.85	6.26	3.92*+
55-59	5.93	1.16	4.04	8.70	5.34*	1.02	3.68	7.75	6.12*+
<i>Education level</i>									
Low	-		-	-	-		-	-	-
Medium	0.73	0.08	0.59	0.90	0.43	0.05	0.34	0.55	0.45+
High	0.31	0.05	0.23	0.42	0.32*	0.05	0.25	0.43	0.23
<i>Constant</i>	0.0000	0.00	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0018+
Women									
<i>Age</i>									
25-29									
30-34	0.99	0.36	0.49	2.01	0.96*	0.35	0.47	1.95	1.40*+
35-39	0.78	0.30	0.36	1.68	0.74*	0.29	0.35	1.59	1.93*
40-44	2.25	0.75	1.17	4.33	2.10*	0.69	1.10	4.01	2.67*+
45-49	3.08	0.97	1.67	5.70	2.84*	0.88	1.54	5.22	4.19*+
50-54	5.49	1.98	2.70	11.14	4.98*	1.79	2.46	10.06	6.19*+
55-59	5.81	1.86	3.10	10.88	5.17*	1.61	2.81	9.50	9.21*+
<i>Education level</i>									
Low	-		-	-	-		-	-	-
Medium	0.83	0.17	0.88	1.52	0.46	0.09	0.32	0.67	0.50+
High	0.40	0.10	0.39	0.82	0.43*	0.09	0.28	0.65	0.37+
<i>Constant</i>	0.0000	0.00	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0005*+

Source: 2010 Population Census; Mortality Information System (SIM) 2013.

* Indicates that the Model IRRs are within the CI of Model 1.

+ Indicates that the Model IRRs are within the CI of Model 2.

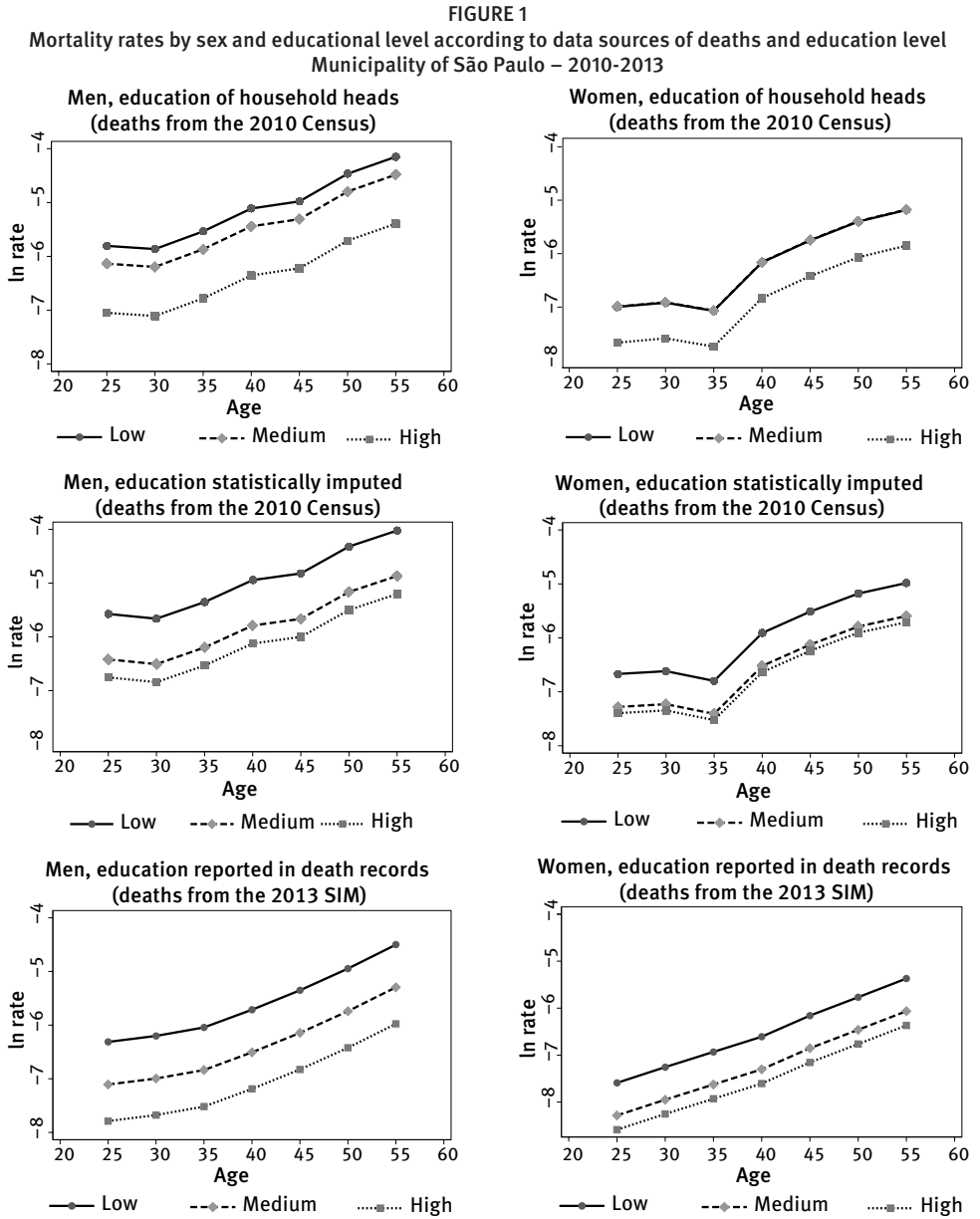
Note: IRR in bold: p-value <0.05. Deaths without correction in the 3 Models.

The differences in mortality by education levels are evident in the three models, indicating lower mortality rates for the more educated groups. However, the ratios between the rates in each data source reveal different gradients. In Model 1, which considers deaths

from the 2010 Census and education of household head, most of the difference occurs on the highest education level. Compared to the lowest education group, mortality rates are about 69% and 60% lower for men and women, respectively. However, the gap between low and medium education levels is relatively small for men and not statistically significant for women.

In Model 2, which is also based on data from the 2010 Census, with education imputed by statistical methods, the largest relative gains are concentrated in the group with medium education. Compared to the lowest education group, mortality rates are about 57% and 54% lower for men and women, respectively. The mortality gradient by education level estimated in Model 3, calculated using data from SIM-2013, is similar to that of Model 2. However, the marginal effect of higher education is slightly more pronounced in Model 3 than in Model 2, particularly among men.

Based on the Poisson Models coefficients, we calculated the mortality rates for men and women by age, education level groups, and the methodology adopted. Mortality rates estimated from Models 1 and 2 (data from the 2010 Census) were adjusted using the ratio between CMR by sex from SIM-2013 and Census-2010, as shown in Table 1. Figure 1 shows four main results. Not surprisingly, mortality rates increase with age and are higher among men than among women. Using the household head's education level as a *proxy* for the deceased's education level (Model 1) results in an atypical gradient not found in the alternative methodologies. Although mortality differences calculated from Models 2 and 3 are not identical, they suggest the important role of secondary education as a protective factor for death. Finally, the association between education and survival of adults is so high that women with low education have mortality rates similar to those of men of medium and high education at the same ages. This result occurs despite the well-known female advantage in adult mortality in Brazil, suggesting that lower education reduces biological and behavioral protections that occur more frequently among women.



Discussion

This study offered new estimates of educational differences in adult mortality in Brazil, contributing to the still scarce literature on the topic. Given the absence of a database with recognized superior quality, we compared three different sets of information. The purpose of this comparison was to offer a critical analysis of adult mortality differences by education level in Brazil. The results indicate that education can reduce mortality rates by up to 77% between ages 25 and 59. Moreover, in a country characterized by low education attainment, graduating from secondary education is associated with significant survival gain. However, the type of data and the need for imputing missing data on education affect the results. Therefore, methodological issues remain central to the analysis of mortality differentials in Brazil, generating uncertainties about the actual patterns and requiring researchers' attention.

Despite all the limitations in the census data, the study also showed that the information from the 2010 Census could be helpful in research on mortality differentials, provided that due care is taken when imputing missing variables. Previously, Silva *et al.* (2016) measured the probabilities of death and life expectancy for the population between 15 and 60 years old according to the household head's education level. Our analysis suggests that the education level of household heads is not a good *proxy* for the education of deceased household members. It results in mortality gradients significantly different from those found when using death records or census data imputed by a more robust statistical model, such as the one proposed here.

One possible explanation for the differences in the results is the educational transition underway in Brazil. There has been an increase in education access since the 1990s with universal coverage of primary education. Between 1986 and 2008, the average education level of the population aged 7 to 25 years went from 4 to 6 years of studies (RIOS-NETO; GUIMARÃES, 2010). Therefore, although individuals' education is associated with household heads' education, variations in education attainment by generation are inevitable in Brazil. Moreover, for each phase of the life cycle, measured according to the age of the deceased, household composition changes. This pattern by age, combined with variations in education attainment between generations, causes the distortions shown in this study. For example, we found (in results not shown) that heads of households are, on average, older than the deceased in households where young individuals have died. Therefore, in these cases, household heads tend to have a lower education level than the deceased members. The association of age and education level reverses in households where deaths occurred at older ages, favoring an overestimation of the average education level. As a result of these combined effects for all ages, we found that a fraction of the deaths that should be classified as belonging to individuals with low education was erroneously identified as belonging to individuals with medium education. Consequently, the mortality rates for the two educational levels have become similar, reducing the mortality gap. Silva *et al.*

(2016) recognized that some intergenerational mobility could underestimate the mortality differentials by education level.

Conversely, education imputed by statistical methods offers mortality estimates based on the individual's educational attainment, making the mortality differentials measured through this strategy more consistent with those obtained with SIM data. The success of the statistical imputation proposed here is due to the large sample size and the availability of many individuals (survivors) and household characteristics in the 2010 Census. We still need to apply it in different geographic regions before confirming it works for Brazil as a whole. Also, it is not free of bias, particularly if we consider the existence of education misreporting among the living residents (NEPOMUCENO; TURRA, 2020), an essential variable in our imputation model. Nevertheless, the similarity of mortality patterns by education level estimated using data from the 2010 Census (imputed) and death records from SIM-2013 suggests that the methodological solution we proposed can be quite valuable for future analyses based on data from deaths in the Brazilian censuses.

Despite the statistical imputation's success, SIM data are still preferable to data from the 2010 Census in mortality studies. However, mortality rates obtained using SIM deaths are also not free from errors. We do not yet know the quality of education information reported on a Brazilian death certificate, especially if someone who does not know the deceased person well has filled it. Moreover, as there are still two types of education information in the SIM, the compatibility process for systematizing the data can be a source of error. Since education may not always be consistently informed in death records and census data, there are also potential numerator-denominator discrepancies when estimating rates using two different data sources.

Despite possible data errors, the mortality differentials estimated in this study, when calculated according to the individuals' education level, are comparable to those presented in earlier studies. Sandoval and Turra (2015) estimated adult mortality by education in Chile for 2001-2003. Although the authors use slightly different education groups (≥ 8 , ≥ 9 and ≤ 12 , ≥ 13 years of education level), results estimated for a neighboring country may help detect unrealistic Brazilian patterns. As an example, individuals living in Chile aged between 55 and 59 years have rates of 13.1 (low education), 8.5 (average), and 3.7 per thousand (high education), indicating relative differences like those found in our study for men living in São Paulo, based on data from SIM-2013. For women in the same age group, the differences between the medium and low education levels are similar. Still, the survival gains at higher education levels are higher in Chile compared to São Paulo. The association of education and mortality is expected to vary in each country, depending on economic, social, and cultural contexts. Confirming our mortality patterns by sex, Gomes *et al.* (2013) also found a negative association between education level and mortality higher for men than for women using SABE panel data for elderly residents in São Paulo.

Due to the data limitations on deaths from the 2010 Census, we did not estimate the educational differentials in mortality for people over 60 years of age. However, the

international literature suggests that mortality differentials decrease with age in other countries (Chile and United States, for example). Some explanations for this pattern include selectivity effects from mortality by education, variations in the relationship between education and mortality by birth cohort, and the existence of public health and other social insurance programs for the elderly in many countries (LAUDERDALE, 2001; PRESTON; ELO, 1995). In the case of Brazil, the increase in access to education among the younger cohorts (RIOS-NETO; GUIMARÃES, 2010) may have important implications for the age patterns of mortality differentials. The education system's expansion may have changed the relationship between education, income, and occupational status across generations, modifying health and mortality differences by cohorts. The consolidation of the public health system (SUS) may have also played an essential role in intermediating health and education. As a universal and comprehensive public system, it may reduce mortality differentials at older ages when the prevalence of morbidities is more significant and individuals retire, reducing their access to supplementary health. Consequently, regional variations in health service provision can be related to mortality differentials by education level in Brazil, a factor not examined in the present study.

Finally, future research must go beyond the measurement of educational differentials in all-cause mortality if the goal is to forecast longevity gains in Brazil. It is necessary to identify the causes of death responsible for the gradients observed and consider other socioeconomic measures in addition to education, such as occupation, income, and socioeconomic conditions before adulthood, experienced in childhood.

References

- ALLISON, P. D. Missing data. *In*: MILLSAP, R. E.; MAYDEU-OLIVARES, A. **The SAGE handbook of quantitative methods in psychology**. SAGE Publications. 2009. p 72-89.
- ANTONOVSKY, A. Social class life expectancy and overall mortality. **The Milbank Quarterly**, v. 45, n. 2, p. 31-73, 1967.
- BERHANE, Y., HOGBERG, U.; BYASS, P.; WALL, S. Gender, literacy, and survival among ethiopian adults, 1987-96. **Bulletin of the World Health Organization**, New York, v. 80, n. 9, p. 714-720, Sept. 2002.
- BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Portaria nº 116, de 11 de fevereiro de 2009**. Regulamenta a coleta de dados, fluxo e periodicidade de envio das informações sobre óbitos e nascidos vivos para os Sistemas de Informações em Saúde sob gestão da Secretaria de Vigilância em Saúde. Brasília, 2009.
- BRASIL. Ministério da Saúde. Coordenação Geral de Informações e Análise Epidemiológica – CGIAE. Sistema de Informações sobre Mortalidade – SIM: Consolidação da base de dados de 2011. Disponível em: http://tabnet.datasus.gov.br/cgi/sim/Consolida_Sim_2011.pdf.
- CASE, A.; DEATON, A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. **Proceedings of the National Academy of Sciences**, v. 112, n. 49, p. 15078-15083, Dec. 2015.

CRIMMINS, E. Demography: the 30 past years, the present, and the future. **Demography**, v. 30, n. 4, p. 579-591, 1993.

CUTLER, D. M.; LLERAS-MUNEY, A. Education and health: evaluating theories and evidence. In: HOUSE, J. S.; SCHOENI, R. F.; KAPLAN, G. A.; POLLACK, H. (ed.). **The health effects of social and economic policy**. New York: Russel Sage Foundation, 2007.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 39, n. 1, p. 1-38, 1977.

DENISOVA, I. **Mortality in Russia: microanalysis**. Center for Economic and Financial Research and New Economic School, March, 2009. (Working Paper, 128).

ELO, I. T.; PRESTON, S. H. Educational differentials in mortality: United States, 1979-1985. **Social Science and Medicine**. v. 42, p. 47-57, 1996.

ELO, I. T. Social class differentials in health and mortality: patterns and explanations in comparative perspective. **Annual Reviews Of Sociology**, v. 35, p 553-572, 2009.

ENDERS, C. K. **Applied missing data analysis**. New York: The Guilford Press, 2010.

GOMES, M. M. F.; TURRA, C.; FÍGOLI, M. G. B.; DUARTE, Y. A. O.; LEBRÃO, M. L. Associação entre mortalidade e estado marital: uma análise para idosos residentes no Município de São Paulo, Brasil, Estudo SABE, 2000 e 2006. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 29, n. 3, p. 566-578, mar. 2013.

GOMES, M. M. F.; TURRA, C.; FÍGOLI, M. G. B.; DUARTE, Y. A. O.; LEBRÃO, M. L. Passado e presente: condições de vida na infância e mortalidade de idosos. **Revista de Saúde Pública**, v. 49, 2015.

GUEDES, G. R.; SIVIERO, P. C. L.; QUEIROZ, B. L.; MACHADO, C. J. Approximating the educational differences in mortality: demographic indirect techniques. **Cadernos de Saúde Coletiva**, Rio de Janeiro, v. 19, n. 2, p. 240-3, 2011.

GROSSMAN, M. On the concept of health capital and the demand for health. **The Journal of Political Economy**, Chicago, v. 80, n. 2, p. 223-55, 1972.

GROSSMAN, M. The relationship between health and schooling. **Eastern Economic Journal**, v. 34, n. 3, p. 281-292, 2008.

HAYWARD, M. D.; HUMMER, R. A.; SASSON, I. Trends and group differences in the association between educational attainment and U.S. Adult mortality: implications for understanding education's causal influence. **Social Science & Medicine**, v. 127, p. 8-18, 2015.

HEITJAN, D. F.; RUBIN, D. B. Inference from coarse data via multiple imputation with application to age heaping. **Journal of the American Statistical Association**, v. 85, n. 410, p. 304-314, 1990.

HEITJAN, D. F.; RUBIN, D. B. Ignorability and coarse data. **Annals of Statistics**, v. 19, n. 4, p. 2244-2253, 1991.

HOUSE, J. S.; LANDIS, K. R.; UMBERSON, D. Social relationship and health. **Science**, v. 241, p. 540-545, 1988.

HOUSE, J. S.; KESSLER, R. C.; HERZOG, R. Age, socioeconomic status and health. **The Milbank Quartely**, v. 68, n. 3, p. 383-411, 1990.

HURT, L. S.; RONSMANS, C.; SAHA, S. Effects of education and other socioeconomic factors on middle age mortality in rural Bangladesh. **Journal of Epidemiology & Community Health**, v. 58, n. 4, p. 315-20, 2004.

KITAGAWA, E. M.; HAUSER, P. M. **Differential mortality in the United States: a study in socioeconomic epidemiology.** Cambridge, MA: Harvard University Press, 1973.

KOCH, E. *et al.* Desigualdad educacional y socioeconómica como determinante de mortalidad en Chile: análisis de sobrevivencia en la cohorte del proyecto San Francisco. **Revista Médica de Chile**, v. 135, n.11, p. 1370-1379, 2007.

LAUDERDALE, D. S. Education and survival: birth cohort, period, and age effects. **Demography**, v. 38, n. 4, p. 551-561, 2001.

LEINSALU, M.; VAGERO, D.; KUNST, A. E. Estonia 1989-2000: enormous increase in mortality differences by education. **International Journal of Epidemiology**, v. 32, n. 6, p. 1081-87, 2003.

LIANG, J. *et al.* Socioeconomic gradient in old age mortality in Wuhan, China. **Journal of Gerontology: Social Sciences**, v. 55B, n. 4, p. S222-S233, 2000.

LIBERATOS P.; LINK, B. G.; KELSEY, J. L. The measurement of social class in epidemiology. **Epidemiologic Review**, v. 10, n. 1, p. 87-121, 1988.

LILLARD, L. A.; WAITE, L. J. 'Til death do us part': marital disruption and mortality. **American Journal of Sociology**, v. 100, p. 1131-56, 1995.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data.** 2. ed. Hoboken, NJ: Wiley, 2002.

LITTLE, T. D.; RHEMTULLA, M. Planned missing data designs for developmental researchers. **Child Development Perspectives**, v. 7, n. 4, p. 199-204, 2013.

MACKENBACH, J. P.; STIRBU, I.; ROSKAM, A.-J, R.; SCHAAP, M. M.; MENVIELLE, G.; LEINSALU, M.; KUNST, A. E. Socioeconomic inequalities in health in 22 European countries. **New England Journal of Medicine**, v. 358, p. 2468-81, 2008.

MANZELLI, H. Educational attainment and adult mortality differentials in Argentina. **Revista Latinoamericana de Población**, v. 8, n. 14, p. 129-163, 2014.

MEARA, E.; RICHARDS, S.; CUTLER, D. The gap gets bigger: changes in mortality and life expectancy by education, 1981-2000. **Health Aff (Millwood)**, v. 27, n. 2, p. 350-360, 2008.

MISTLER, S. A.; ENDERS, C. K. Planned missing data designs for developmental research. *In*: LAURSEN, B.; LITTLE, T. D.; CARD, N. A. (ed.). **Handbook of developmental research methods.** New York, NY: Guilford Press, 2012. p. 742-754.

MONTEZ, J. K.; HUMMER, R. A.; HAYWARD, M. D.; WOO, H.; ROGERS, R. G. Trends in the educational gradient of U.S. adult mortality from 1986 through 2006 by race, gender, and age group. **Research on Aging**, v. 33, n. 2, p. 145-71, 2011.

MONTEZ, J. K.; HUMMER, R. A.; HAYWARD, M. D. Educational attainment and adult mortality in United States: a systematic analysis of functional form. **Demography**, v. 49, p. 315-336, 2012.

MORAIS, C. C. B. **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, 2019.

NCHS – National Center for Health Statistics. **Highlights from Wave 1 of the National Survey of Personal Health Practices and Consequences: United States, 1979.** Vital and Health Statistics, Series 15(1), 1981.

NEPOMUCENO, M. R.; TURRA C. M. Assessing the quality of education reporting in Brazilian censuses. **Demographic Research**, v. 42, p. 441-460, 2020.

OLSHANSKY, S. J. *et al.* A potential decline in life expectancy in the United States in the 21st century. **New England Journal of Medicine**, v. 352, n. 11, p. 1138-1145, 2005.

PAMPEL, F. C.; KRUEGER, P. M.; DENNEY, J. T. Socioeconomic disparities in health behaviors. **Annual Review of Sociology**, v. 36, p. 349-70, 2010.

PAPPAS, G.; QUEEN, S.; HADDEN, W.; FISHER, G. The increasing disparity on mortality between socioeconomic groups in the United States, 1960 and 1986. **New England Journal of Medicine**, v. 329, n. 2, p. 103-109, 1993.

PEREIRA, F. N. A.; QUEIROZ, B. L. Diferenciais de mortalidade jovem no Brasil: a importância dos fatores socioeconômicos dos domicílios e das condições de vida nas localidades. **Cadernos de Saúde Pública**, v. 32, n. 9, 2016

PEREZ, E. R.; TURRA, C. M. Desigualdade social na mortalidade no Brasil: diferenciais por escolaridade entre mulheres adultas. *In*: XVI ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS. **Anais [...]**. Caxambu: Abep, 2008.

PEREZ, E. R. **Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil**. 2010, 146f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

PRESTON, S. H.; TAUBMAN, P. Socioeconomic differences in adult mortality and health status. *In*: MARTIN, L.; PRESTON, S. H. **The demography of aging**. Washington, DC: National Academy Press, 1994. p. 279-318.

PRESTON, S. H.; EIO, I. T. Are educational differentials in adult mortality increasing in the United States? **Journal of Aging Health**, v. 7, p. 476-496, 1995.

QUEIROZ, B. L.; SAWYER, D. O. T. O que os dados de mortalidade do Censo de 2010 podem nos dizer? **Revista Brasileira de Estudos de População**, v. 29, n. 2, p. 225-238, 2012.

RIBEIRO, M. M. **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, 2016.

RIOS-NETO, E.; GUIMARÃES, R. R. M.; PIMENTA, P. S.; MORAES, T. A. Análise da evolução de indicadores educacionais no Brasil: 1981 a 2008. Belo Horizonte: Cedeplar, Universidade Federal de Minas Gerais, 2010. (Working Paper, n. 386).

RIOS-NETO, E.; GUIMARÃES, R. R. M. The demography of education in Brazil: inequality of educational opportunities based on Grade Progression Probability (1986-2008). **Vienna Yearbook of Population Research**, v. 8, n. 1, p. 283-312, 2010.

ROGER, R. G. Marriage, sex and mortality. **Journal of Marriage and the family**, v. 57, n. 2, p. 515-526, 1995.

ROGER, R. G. The effects of family composition, health, and social support linkages on mortality. **Journal of Health and Social Behavior**, v. 37, n. 4, p. 326-338, 1996.

ROGERS, R. G.; HUMMER, R. A.; NAM, C. B. **Living and dying in the USA**: behavioral, health, and social differences of adult mortality. Academic Press, 1999. Cap 7. The effects of basic socioeconomic factors on mortality, p. 115-139.

ROSENZWEIG, M. R.; SCHULTZ, T. P. Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birthweight. **Journal of Political Economy**, v. 91, n. 5, p. 723-46, 1983.

ROSS, C. E.; MASTERS, R. K.; HUMMER, R. R. A. Educational and the gender gaps in health and mortality. **Demography**, v. 49, p. 1157-1183, 2012.

RUBIN, D. B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581-592, 1976.

RUBIN, D. B. Multiple imputation after 18+ years. **Journal of the American Statistical Association**, v. 91, n. 434, p. 473-489, 1996.

SANDOVAL, M. H.; TURRA, C. M. El gradiente educativo en la mortalidad adulta en Chile. **Revista Latinoamericana de Población**, v. 17, n. 9, p. 7-35, 2015.

SASSON, I. Trends in life expectancy and lifespan variation by educational attainment: United States, 1990-2010. **Demography**, v. 53, p. 269-293, 2016.

SCHAFFER, J. L. **Analysis of incomplete multivariate data**. London: Chapman & Hall, 1997.

SCHAFFER, J. L.; GRAHAM, J. W. Missing data: our view of the state of the art. **Psychological Methods**, v. 7, p. 147-177, 2002.

SHKOLNIKOV, V. M. *et al.* Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. **Social Science & Medicine**, v. 47, n. 3, p. 357-369, 1998.

SICKLES, R. C.; TAUBMAN, P. Mortality and morbidity among adults and the elderly. *In*: ROSENZWEING, M.; STARK, O. **Handbook of population and family economics**. Elsevier: 1997.

SILVA, L. E.; FREIRE, F. H. M. A.; PEREIRA, R. H. M. Diferenciais de mortalidade por escolaridade da população adulta brasileira. **Cadernos de Saúde Pública**, v. 32, n. 4, p. 1-12, 2016.

SUBRAMANIAN, S. V. *et al.* The mortality divide in India: the differential contributions of gender, caste, and standard of living across the life course. **American Journal of Public Health**, v. 96, n. 5, p. 818-825, 2006.

TURRA, C. M.; PEREZ, E. R.; GUIMARÃES, R. The effect of changes in educational composition on adult female mortality in Brazil. **Research on Aging**, v. 38, n. 3, p. 283-298, 2016.

US PUBLIC HEALTH SERVICES. **US Surgeon General's Advisory Committee on Smoking and Health, Smoking and Health**. Washington DC, 1964.

VALKONEN, T. Adult mortality and level of education: a comparison of six countries. *In*: FOX, J. **Health inequalities in European countries**. Aldershot: Gower, 1989. p. 142-162.

WAJNMAN, S. **Demografia da família e dos domicílios brasileiros**. Tese (Professor titular) – Cedeplar, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

WALQUE, D.; FILMER, D. Trends and socioeconomic gradients in adult mortality around the developing world. **Population and Development Review**, v. 39, n. 1, p. 1-29, 2013.

WILLIAMS, D. F. Socioeconomic differentials in health: a review and redirection. **Social Psychology Quarterly**, v. 53, n. 2, p. 81-99, 1990.

About the authors

Mirian Martins Ribeiro holds a BA in Economics (UFMG), and a MSc. and a Ph.D. in Demography from Cedeplar (UFMG). She is currently an Associate Professor at the Economics Department of the Institute of Applied Social Sciences, Universidade Federal de Ouro Preto.

Cassio Maldonado Turra holds a BA in Economics (UFMG), a MSc. in Demography from Cedeplar (UFMG), and a M.A. and a Ph.D. in Demography from the University of Pennsylvania. He is currently an Associate Professor at the Demography Department, Cedeplar, UFMG.

Cristine Campos de Xavier Pinto holds a BA in Economics (UFMG), a MSc. in Economics from PUC-RIO, a MSc. in Statistics and a Ph.D. in Economics from the University of California at Berkeley. She is currently Assistant Professor of Economics at the São Paulo School of Economics, Fundação Getúlio Vargas (FGV).

Contact address

Mirian Martins Ribeiro

Universidade Federal de Ouro Preto, Instituto de Ciências Sociais Aplicadas
Rua do Catete, 166, Centro
35420-000 – Mariana-MG, Brazil

Cassio Maldonado Turra

Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional
Av. Antônio Carlos, 6626, sala 3006, Pampulha
31270-901 – Belo Horizonte-MG, Brazil

Cristine Campos de Xavier Pinto

Fundação Getúlio Vargas
Rua Itapeva, 474, 12º andar, Bela Vista
01332-000 – São Paulo-SP, Brazil

Resumo

Mortalidade adulta por nível de escolaridade em São Paulo: análise comparativa a partir de diferentes estratégias metodológicas

Neste artigo, são estimados os diferenciais educacionais de mortalidade de adultos residentes em São Paulo. É realizada uma análise comparativa de estimativas a partir de dados do Censo 2010 e do Sistema de Informação de Mortalidade (SIM) – Datasus, e de três formas distintas de mensuração da escolaridade: registrada no SIM; declarada no Censo para o responsável pelo domicílio; e imputada estatisticamente no Censo para indivíduos que morreram. Para as imputações da escolaridade, utilizou-se o método de Dempester (1977), que propõe o uso do algoritmo esperança-maximização (algoritmo E-M) para lidar com dados faltantes. Foram considerados três níveis de escolaridade (baixo, médio e alto) e estimadas as taxas de mortalidade com base em modelos *Poisson*. Os resultados indicam que a obtenção de escolaridade pode reduzir em até 77% as taxas de mortalidade entre 25 e 59 anos de idade. Além disso, em um país em que a população tem baixa escolaridade, obter ensino médio representa um ganho significativo do ponto de vista da sobrevivência adulta (cerca de 50%). Encontraram-se padrões de mortalidade por escolaridade semelhantes para as estimativas obtidas com dados registrados no SIM e aqueles imputados no Censo 2010. Além disso, a análise sugere que estimativas assumindo a escolaridade do responsável pelo domicílio resultam em diferenciais de mortalidade atípicos, provavelmente distorcidos pela transição de educação no Brasil. Espera-se que o modelo de imputação proposto aqui possa ser utilizado em futuras análises dos dados de mortalidade a partir do Censo 2010.

Palavras-chave: Mortalidade adulta. Dados faltantes. Imputação. Diferenciais socioeconômicos em mortalidade. Brasil. Escolaridade.

Resumen

Mortalidad adulta por nivel de escolaridad en San Pablo: análisis comparativo a partir de diferentes estrategias metodológicas

En este artículo estimamos los diferenciales educativos de la mortalidad de adultos en San Pablo. Ofrecemos un análisis comparativo de estimaciones con base en datos del censo de 2010 y el Sistema de Información de Mortalidad (SIM) – Datasus, y tres formas diferentes de medir la escolaridad: registrada en el SIM, declarada en el censo por el jefe de hogar e imputado estadísticamente en el censo para las personas fallecidas. Para las imputaciones de escolaridad se utilizó el método de Dempester (1977), que propone el uso del algoritmo de maximización de esperanza (algoritmo E-M) para tratar los datos faltantes. Consideramos tres niveles de educación (bajo, medio y alto) y estimamos las tasas de mortalidad con base en los modelos de Poisson. Los resultados indican que la escolarización puede reducir las tasas de mortalidad entre los 25 y 59 años hasta en un 77 %. Además, en un país donde la población tiene bajo nivel de educación, completar la educación secundaria representa una ganancia significativa desde el punto de vista de la supervivencia de los adultos (alrededor del 50%). Encontramos patrones similares de mortalidad por educación para las estimaciones obtenidas con datos registrados en el SIM y datos imputados en el Censo de 2010. Además, nuestro análisis sugiere que las estimaciones asumiendo la educación del jefe de hogar dan como resultado diferenciales de mortalidad atípicos, probablemente distorsionados por la transición de educación en Brasil. Esperamos que el modelo de imputación propuesto aquí se pueda utilizar en futuros análisis de mortalidad del Censo de 2010.

Palabras clave: Mortalidad de adultos. Datos faltantes. Imputación. Diferenciales socioeconómicos en la mortalidad. Brasil. Educación.

Received for publication in 09/08/2020
Approved for publication in 29/12/2020