Assessing the quality of self-reported education in Brazil with intercensal survivorship ratios

Marília R. Nepomuceno  I  nepomuceno@demogr.mpg.de
Cássio M. Turra

This working paper has been approved for release by: Alyson van Raalte (vanRaalte@demogr.mpg.de), Head of the Research Group: Lifespan Inequalities.

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.
Assessing the quality of self-reported education in Brazil
with intercensal survivorship ratios

Marília R. Nepomuceno*1 and Cássio M. Turra2

1Max Planck Institute for Demographic Research, Rostock, Germany
2Centre for Development and Regional Planning – CEDEPLAR - Federal University of Minas Gerais,
Belo Horizonte, Brazil

Abstract

In developing countries, improving access to schooling has been and remains a priority. At the same time, a growing body of research relates education to demographic variables. It is therefore essential to measure the education variable accurately. In Brazil, although the high degree of inaccuracy in age reporting is well known, previous research has neglected problems of misreporting which may affect other variables such as education. To fill this gap, we calculate mortality levels by education as implied by intercensal survivorship ratios, to investigate the quality of self-reported education among adults in Brazil between the 1991 and 2000 censuses. Our findings show evidence of education misreporting. Analysis by single year of schooling only barely reveals the expected educational gradient in mortality. After categorization of age and years of schooling into groups, a positive relationship between education and survival does appear, although some implausible patterns remain. This study is an important step in demonstrating and assessing potential errors in census education data in Brazil. Our results highlight the importance of efforts to reduce misreporting of data on education, particularly in countries where an educational expansion is underway, and deficiencies in data quality are a potential issue of concern.

Keywords: Census data; education misreporting; data quality; developing countries; Brazil; adult mortality.

*Corresponding author. E-mail: nepomuceno@demogr.mpg.de
1 Introduction

Education is one of the keys to development and economic growth and is connected to demographic variables in many different ways. In virtually all societies, individuals with higher levels of education enjoy better health and longer lives (Preston and Taubman 1994; Elo and Preston 1996; Mackenbach et al. 1999; Koch et al. 2007; Zhu and Xie 2007; Rosero-Bixby and Dow 2009; Turra et al. 2016; Lutz and Kebede 2018; Smith-Greenaway and Yeatman 2019). Highly educated couples have fewer children, mainly because they marry later, use contraception more effectively, and have more autonomy in reproductive decision-making (Singh and Casterline 1985; Potter et al. 2010; Bongaarts et al. 2017; Rios-Neto et al. 2018). Moreover, migration flows across regions with different developmental levels are usually associated with selection of migrants by education group (de Haas 2010; Lewis 1986). It is thus not surprising that as educational enrolments expand worldwide, there is a growing emphasis on population forecasts by educational level (KC et al. 2010; Lutz and KC 2011; Lutz et al. 2014).

As an essential variable in demographic analysis, accurate measurement of education is indispensable. Studies of the quality of reported education are not new in the literature (Folger and Nam 1964; Gustavus and Nam 1968; Black et al. 2003; Sorlie and Johnson 1996; Kane et al. 1999; Johnson-Greene et al. 1997; Battistin et al. 2014; Lerch et al. 2017). Some of these have documented the misreporting of educational levels in data systems. For example, in both the 1950 and 1960 U.S. census, there was a significant amount of net overreporting, particularly at older ages for high school and college levels (Folger and Nam 1964). Examination of the 1990 U.S. census detected misreporting of some doctoral and professional degree categories (Black et al. 2003).

One of the consequences of education misreporting in census data is the miscalculation of the denominator of demographic rates, which can bias the measurement of educational gradients. In mortality analysis, for example, where death rates usually derive from two independent data sources - vital statistics (death counts) and census data (exposures-at-risk) - disagreement between reported education in the numerator and denominator may affect conclusions about how mortality differentials by education vary across age, gender, race, and geographical groups. Several authors have documented disagreements between death counts and population census records by education in Europe and the U.S. The level of disagreement tends to increase with age in Lithuania (Shkolnikov et al. 2007), and for some racial/ethnic groups in the U.S. such as blacks and Hispanics (Rostron et al. 2010). However, even among population groups that
are less likely to report education levels differently across data sources, changes in the census questions can lead to inconsistent information over time (Lerch et al. 2017).

In Brazil, research has neglected the possible effects of education misreporting in demographic, social, and economic studies. Although several authors have looked at coverage errors, both in vital and census data (Paes and Albuquerque 1999; Paes 2007; Lima and Queiroz 2014; Cavennaghi and Alves 2016), demographers have paid less attention to content errors, and analysis has mostly been limited to age misreporting (e.g. Paes and Albuquerque 1999; Agostinho 2009; Gomes and Turra 2009; Turra 2012; di Lego et al. 2017; Nepomuceno and Turra 2019). Nevertheless, the high degree of inaccuracy in age reporting suggests that problems of misreporting may affect other variables, including education.

This article examines the quality of education information in Brazilian census data. To do this, we calculate mortality levels by education as implied by intercensal survivorship ratios, to investigate the quality of self-reported education data among adults in Brazil between the 1991 and 2000 censuses. Here, we hypothesize that measurement errors are more substantial during times of accelerated expansion of schooling, such as the period that characterized the education system in Brazil after 1980 (IBGE 2004; Rios-Neto et al. 2010). The rapid education transition may have influenced the reliability of education data in at least three different ways. First, if better-educated individuals tend to report their characteristics more accurately, mainly when information is retrospective, errors will be more substantial in earlier data collections and among older cohorts. Second, the expansion of education may have changed people’s perception of their relative social position. Older age groups, who lived their schooling years before the development of the education system, may feel inclined to overstate their educational levels to level-off any cohort differences. Finally, the expansion of schooling has been followed by changes in the educational system and census questions, affecting the comparability of responses over time. Given the global effort to improve access to schooling (United Nations 2015, 2016), our results aim to draw attention to the accuracy of reported education not only in Brazil but also in those countries where educational expansion is underway, and deficient data quality is a potential issue.
2 Data and Methods

We drew data for men and women 40–89 years old, from the 1991 and 2000 Brazilian Census (IPUMS-I 2019). Although the most recent census data available are from 2010, the exclusion from the questionnaire of information on the highest grade completed precluded us from using the data for that year. To have an exact 10-year interval between periods, we estimated the population for the year 1990 based on the set of age-specific growth rates between 1991 and 2000 and the population in 1991. We selected only individuals who were not attending school at the census time, representing about 99.5% of the population aged 40 to 89 in 1990 and 97.3% in 2000.

We measure educational attainment as the highest grade completed within the most advanced level attended in the education system. We used the IPUMS’ harmonized variable named “YRSCHOOL” which accounts for the number of years of schooling.

To evaluate the quality of self-reported education, we assessed the implicit mortality by education between the two censuses through intercensal survivorship ratios (ISR) by age, sex, and educational attainment:

\[
ISR_x^{k,i}(j) = \frac{N_{x+j}^{k,i}(t + j)}{N_x^{k,i}(t)}
\]

where \(N_x^{k,i}(t)\) is the cohort at age \(x\), sex \(k\), at educational level \(i\) in the year \(t\), and \(N_{x+j}^{k,i}(t + j)\) is the same cohort \(j\) years older at the educational level \(i\) and sex \(k\) in the year \(t + j\). In our first analysis, we used single year of schooling, from 0 to 12+. Next, we calculated the number of years of schooling categorized into four intervals: 0–3, 4–8, 9–11, and 12+, that correspond respectively to the first and second stages of the primary, secondary and tertiary education.

After the categorization of the years of schooling, values of the ISR at each age were translated into life expectancy at age 40 in the Coale-Demeny West model life-table (Coale et al. 1983). We used the United Nations version of the Coale-Demeny Model Life Tables, which extends the original mortality levels to include life expectancy at birth up to 100 years (United Nations 2017). This allowed us to assess the consistency of the adult level of mortality across age and educational groups.
The ISR capture changes in the cohort size during the intercensal period. In the absence of international migration and changes in census data quality, one should expect survivorship ratios to be less than one due to the impact of mortality. Here we assume that the population was closed to international migration at ages above 40, between 1990 and 2000. This assumption is reasonable because the international net migration rate was very low during the 1990s in Brazil, reaching less than half of one percent for the population aged ten years and older (Carvalho and Campos 2006; Campos 2011), and an even lower values at older ages (Garcia 2013). To mitigate the effect of changes in census coverage on the ISR, we also adjusted the census enumeration according to omission rates (1.04 and 1.02, respectively for males and females for 1990, and 1.03 for males and 1.01 for females in 2000) published by the Brazilian Institute of Geography and Statistics (IBGE 2003).

We expect survivorship ratios to increase with educational attainment. Earlier analyses that used surveys or data from the Mortality Information System of the Ministry of Health suggested a strong educational gradient in adult mortality in Brazil (Rentería and Turra 2009; Turra et al. 2016, 2018). Despite the existence of public programs to improve or supplement adult education, we expect only negligible gains in education over time, at ages above 40 in Brazil. Between the 1991 and 2000 censuses, the proportion of individuals who reported not attending school varied from 99.2% to 96.5% for the age group 40–49, and from 99.6% to 98.2% for the age group 50–59. For older age groups, these proportions were even higher and more similar in the two censuses: between 99.7% and 99% for the age group 60–69, 99.8% to 99.3% for the age group 70–79, and 99.8% and 99.4% for individuals aged 80 and older.

3 Results

Table 1 presents the education distribution by age groups and sex in Brazil. In both years, there is a substantial proportion of adults in the least educated groups, and a smaller percentage in the most educated categories. Between 1990 and 2000, the most considerable change in the education distribution is among the least educated (0–3 years of schooling), mainly for younger age groups because of the advance of the education transition in the country. For older age groups, proportions change only slightly over time.
Table 1: Education distribution (%) by age: Brazil, men and women 1990 and 2000.

Source: Census data (IPUMS-I 2019).

Figure 1 shows intercensal survivorship ratios by age, sex, and single year of schooling. The estimates show an irregular and unexpected pattern of the ISR by education. First, except for the oldest cohorts (70–74 and 75–79 years old in 1990), the ISR can get higher than one, suggesting that the size of cohorts increased between the censuses for some years of completed schooling, despite the impact of mortality. Second, the ISR do not increase monotonically with education. Although survivorship tends to increase at higher years of schooling, it fluctuates over the distribution, varying sharply, particularly at one, five and nine years of schooling. For example, the number of men and women aged 40–44 with one year of schooling increased by more than 30% between 1990 and 2000, and more than 40% for those with five years of schooling. For nine years of schooling, the number of women and men aged 55–59 reporting this level increased by 30% and 10% respectively between 1990 and 2000.
Figure 1: Intercensal Survivorship Ratio by age and years of schooling: Brazil, men and women, 1990–2000.

Source: Census data (IPUMS-I 2019).
One way to re-examine education reporting is by categorizing years of schooling into intervals (0–3, 4–8, 9–11, and 12+). Table 2 shows the results, including the ISR and the life expectancy at age 40 (e_{40}) from the West Model of the Coale-Demeny Life Tables implied by the ISR. To further improve our estimates, we also re-categorized the 5-year into 10-year age intervals, hoping to minimize the possible effects of age misreporting in the census, particularly at older ages (Agostinho 2009; Nepomuceno and Turra 2019).

Table 2 shows the educational gradient in mortality, i.e., the positive relationship between education and survival. Life expectancy at age 40 is 0.44 to 7.84 years lower for individuals with 0–3 years of schooling than for those with 12 or more years of schooling, depending on the cohort and sex.

Some implausible results remain after the categorization of years of schooling. Table 2 reveals that the survival advantage of the most educated does not decrease monotonically with age. For instance, the difference in e_{40} between the least and the most educated drops from 7.8 years for women aged 40–49 to 5.2 to those aged 50–59 in 1990. But it then increases to 6.1 years at the age group 60–69, reducing to 4.0 at 70–79 years in 1990. By comparing the individuals with 9–11 and those with 12 years or more of schooling, we identify other implausible results that cast doubt on the quality of self-reported education in the Brazilian census. For most male and female cohorts, life expectancy reduces for those with 12+ years of schooling compared with 9–11 years of schooling, contradicting the expected gradient. Table 2 also shows that mortality levels implied by the ISR vary significantly across cohorts/age. For all educational categories and both sexes, the e_{40} implied by the ISR increases for the older cohorts, and the difference can get as high as twelve years for the lowest educated group.
### Table 2: Intercensal Survivorship Ratio by educational groups and decennial age groups: Brazil, men and women, 1990–2000.

Note: * life expectancy at age 40 from the Model West implied by the ISR.

Source: Census data (IPUMS-I 2019).

<table>
<thead>
<tr>
<th>Age in 1990</th>
<th>Age in 2000</th>
<th>Years of schooling</th>
<th>Survival ratio</th>
<th>Implied $e_{40}$</th>
<th>Survival ratio</th>
<th>Implied $e_{40}$</th>
<th>Survival ratio</th>
<th>Implied $e_{40}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40–49</td>
<td>50–59</td>
<td>0–3</td>
<td>0.8410</td>
<td>26.66</td>
<td>0.8603</td>
<td>27.94</td>
<td>0.9570</td>
<td>36.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4–8</td>
<td>0.8398</td>
<td>33.52</td>
<td>0.7925</td>
<td>30.67</td>
<td>0.8867</td>
<td>36.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–11</td>
<td>0.6747</td>
<td>34.94</td>
<td>0.6358</td>
<td>33.17</td>
<td>0.7287</td>
<td>37.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12+</td>
<td>0.4350</td>
<td>37.46</td>
<td>0.4229</td>
<td>37.70</td>
<td>0.4785</td>
<td>38.46</td>
</tr>
<tr>
<td>50–59</td>
<td>60–69</td>
<td>0.8671</td>
<td>28.51</td>
<td>0.8974</td>
<td>30.76</td>
<td>0.9600</td>
<td>37.64</td>
<td>0.9523</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4–8</td>
<td>0.8798</td>
<td>35.51</td>
<td>0.8931</td>
<td>36.49</td>
<td>0.9607</td>
<td>44.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–11</td>
<td>0.7090</td>
<td>35.87</td>
<td>0.7739</td>
<td>38.74</td>
<td>0.8881</td>
<td>45.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12+</td>
<td>0.5082</td>
<td>40.20</td>
<td>0.5646</td>
<td>42.04</td>
<td>0.6545</td>
<td>44.94</td>
</tr>
<tr>
<td>60–69</td>
<td>70–79</td>
<td>0.7090</td>
<td>35.87</td>
<td>0.7739</td>
<td>38.74</td>
<td>0.8881</td>
<td>45.15</td>
<td>0.8388</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4–8</td>
<td>0.7909</td>
<td>36.49</td>
<td>0.8113</td>
<td>39.46</td>
<td>0.9145</td>
<td>47.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9–11</td>
<td>0.4350</td>
<td>37.46</td>
<td>0.4229</td>
<td>37.70</td>
<td>0.4785</td>
<td>38.46</td>
</tr>
<tr>
<td>70–79</td>
<td>80–89</td>
<td>0.5082</td>
<td>40.20</td>
<td>0.5646</td>
<td>42.04</td>
<td>0.6545</td>
<td>44.94</td>
<td>0.6241</td>
</tr>
</tbody>
</table>

4 Discussion and Conclusion

Our study shows evidence of education misreporting in Brazil. The ISR by single year of schooling only barely reflects the known educational gradient in mortality described in the literature. After the categorization of the years of schooling, the positive relationship between education and survival shows up, although some implausible results remained.

An ISR higher than one, mainly at one, five and nine years of education, suggests an unlikely increase in the cohort size between censuses. In the Brazilian education system, these grades corresponded to the beginning of stages 1 and 2 of primary (1–4, 5–8 years of schooling) and the beginning of secondary (9–11 years of schooling) in 1990. Therefore, the implausible high ISR at these points of the distribution suggests two possible (complementary) explanations. First, some adults may have genuinely moved between levels of educational attainment over the intercensal period, more than compensating the mortality effect. Second (and more likely), a fraction of the cohorts may have misclassified themselves as literate or having some stage 2 of primary or secondary education in the second census, because of memory error and other self-reporting inconsistencies caused by changes in the social context, the census questionnaire, and education reforms.
The cohorts we analyzed experienced at least two education reforms after their schooling years that greatly changed the structure and the terminology of the education system, one in 1971, and another in 1996. The first reform, for instance, merged two formerly levels that corresponded, respectively, to 1–4 grades and 5–8 grades into a new level with 1–8 grades (Rigotti and Cerqueira 2004). Individuals who lived their schooling years before these reforms may have problems to classify themselves according to the new education system, resulting in misclassification of their educational level. Further, census questions to measure educational attainment changed over time, which may be another source of inconsistency. In 1991 census, the questions to calculate the highest grade completed within the most advanced level attended were “What was the last grade passed?” and “What was the last level passed”; while in 2000 the questions were “What was the highest course attended, in which you completed at least one grade?” and “What was the last grade passed?”. These mentioned changes point out some challenges in the comparison of the years of schooling over time, by age and cohorts, as revealed by our findings.

Education gains may also contribute to the implausible higher ISR. Since the number of individuals aged 40–49 with one year of schooling increased by more than 30% during the intercensal period, and part of this may be due to education gains, we checked the proportion of illiterates who were attending school in the 1991 census, and could potentially achieve one year of schooling between 1991 and 2000. This proportion was very low, reaching less than 0.05% in 1991, and less than 0.06% in the 2000 census. This small proportion is not enough to explain the striking increase in the population aged 40—49 with one year of schooling between the censuses.

To minimize the effects of age and education misreporting, we presented results by 10-year age groups and four intervals of years of schooling. After the re-categorization, the educational gradient in mortality became more explicit, and the estimated ISR were all lower than one. The categorization of the years of schooling is an alternative to reduce the effect of education misreporting. However, since there are substantial socio and economic differences among individuals within groups, by categorizing, we cannot measure important educational differences in mortality due to the inconsistencies revealed by the ISR by single years of schooling.

Our results also show some inconsistencies after the re-categorization, including lower ISR for individuals with 12+ than 9–11 years of schooling. One important finding is the lack of consistency in mortality levels by age. According to the $e_{40}$ implied by the ISR, older cohorts experienced lower levels of mortality at older ages. The difference is particularly strong for the least educated. The lower levels of mortality at older ages could be explained by the unusual
adult age pattern of mortality in Brazil, which has been shown to increase with age at a slower rate than in countries with high-quality population and mortality data (Turra 2012).

Since the estimates provided here show evidence of education misreporting in the census, our findings increase concerns about the true educational distribution of the adult population. Further, since the data on education attainment seem to be differentially misreported by age and educational level, the validity of age-specific demographic rates by educational level should be interpreted cautiously in Brazil. Furthermore, if education misreporting is different across censuses, trend analysis using these data will reflect erroneous patterns. Global studies that use census education data from developing countries to project population should be aware of this weakness. Lastly, our findings draw attention to the importance of investigating the potential bias in demographic rates by educational levels in Brazil and in other developing countries where educational expansion is underway.

This study is just the first step in revealing potential errors in census education data, and we still do not know nearly enough. Efforts need to be made to measure the magnitude and the direction (overreporting and underreporting) of misreporting of census education data, for the whole population and for subgroups. Different sources of education data could be used to estimate the amount of education misreporting and provide adjusted figures. Linkage with administrative records may help, although this type of data is rarer at older ages. At the same time, attempts to reduce misreporting of education should rely on improvements in census data collection, such as in phrasing of relevant and comparable questions over time and the reduction of omission rates.
5 Acknowledgement

We thank Alyson van Raalte, Ugofilippo Basellini, Simone Wajnman, Jose Alberto de Carvalho, Bernardo Queiroz, Mirian Ribeiro and Marco Gonzaga for their helpful comments and suggestions. This present study was supported by the European Research Council [grant number 716323] and 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).
References


