Assessing the quality of self-reported education among adults in Brazil, 1991-2000

Marilia R. Nepomuceno
Cássio M. Turra
Demography Department
Cedeplar, UFMG, Brazil

1. Introduction

There is a growing body of research on the relationship between education and demographic variables. In virtually all societies, better educated individuals have lower morbidity and mortality rates than the less educated (Preston and Taubman 1994; Elo and Preston 1996; Mackenbach et al. 1999; Koch et al. 2007; Perez 2010; Zhu and Xie 2007, Gomes 2011, Rosero-Bixby and Dow 2009, Ribeiro 2016). Highly educated couples also have fewer children, mainly because they marry later, use contraception more effectively, and have more autonomy in reproductive decision-making (Singh and Casterline 1985, Martin 1995, Bongaarts 2010, Rios-Neto 2000, Merrick and Berquó 1983). In addition, education affects migration patterns; studies have found positive selection of migrants by education when measuring flows from poorer to wealthier regions (De Haas 2010; Lewis 1986).

To the extent that education is an important variable in demographic analysis, it is essential to measure it properly. Several studies have already documented errors in education reporting (Folder and Nam 1964, Gustavus and Nam 1968, Black et al. 2003, U.S. Bureau of the Census 1963, Kane et al. 1999, Johnson-Greene et al. 1997, Battistin et al. 2014). Usually, education is overstated in data systems, particularly at older ages, at higher educational levels, and among men (Folder and Nam 1964, Gustavus and Nam 1968, Black et al. 2003, U.S. Bureau of the Census 1963). For example, according to a postcensal study in the U.S., 16% of adults overstated their educational level, whereas a lower proportion, about 10%, made mistakes in the opposite direction, underestimating their educational attainment in 1960 (U.S. Bureau of the Census 1963). More recently, Black et al. (2003) showed that about one-fifth of respondents in the 1990 U.S. census, who declared to hold a doctoral degree, and another third from those who reported to hold a professional degree, did not.

The deleterious consequences of education misstatement for population studies depend on the magnitude of the error and the type of analysis. For example, in regression analysis, when education is as an independent variable, education misstatement can bias

the estimated parameters if the measurement error is correlated with the error term (Kane et al. 1999, Black et al. 2003). This is the case of studies on returns to schooling, when individuals with poor cognitive skill, who are more likely to earn lower wages, also misreport education (Black et al. 2003). Another example is the estimation of educational disparity in mortality, particularly when death rates are calculated from two different data sources. Although education misstatement may affect any dataset, data quality tends to be worse in death certificates than in population estimates (Shai & Rosenwaike 1989, Sorlie & Johson 1996, Shkolnikov et al. 2007, Rostron et al. 2010, Lerch et al. 2017), since education is provided by a third party who often know little about the deceased, like a funeral director (Rostron et al. 2010). In addition, inconsistencies between two data sources is usually greater among the less educated, which tend to overstate the mortality of the less educated and understate mortality rates for the most educated (Shkolnikov et al. 2007).

In Brazil, the literature has neglected the possible effects of education misstatement in demographic, social and economic studies. Although several studies have looked at coverage errors, both vital and census data (Paes and Alburqueque 1999, Paes 2007, Lima and Queiroz 2014, Cavenaghi and Alves 2016), little attention has been paid to content errors. In addition, most of the analysis on content errors has been limited to age misreporting, which suggests, based on the high degree of inaccuracy that has been found (Popolo 2001; Horta 2005; Agostinho 2009; Gomes and Turra 2009; Horta 2012, Turra 2012), that problems of misreporting may affect other variables, including education.

Here, we hypothesize that education misstatement becomes even worse during times of an accelerated expansion of schooling. In Brazil, large public investments in education started only in the second half of the 20th century. In 1950, half of the population was illiterate and only 36% of the children aged 5 to 14 years were enrolled in school (IBGE 2001). However, after the promulgation of the 1988 constitution, school has been mandatory for children aged 7 to 14 years, and public school became a constitutional right. Hence, illiteracy rates reduced sharp from 25.5% in 1980 to 13.6% in 2000, and nearly all children aged 7 to 14 years attended school in the late-2000s (Rios-Neto et al. 2010, IBGE 2001). The gross enrollment rate for the secondary education level has also risen fast, from about 30% in 1980 to 80% in 2010 (INEP 1999/2006), and the tertiary

education experienced an expansion of about 3 million individuals between 2000 and 2010 (INEP 2011).

The rapid education transition may have influenced the reliability of education data in different ways. First, if better educated individuals tend to report their own characteristics more accurately, particularly when information is retrospective, errors will be larger in earlier data collections and among older cohorts. Second, the expansion of education may have changed people's perception about their relative social position. Older cohorts, who lived their schooling years before the expansion of the education system, may feel inclined to overstate their educational levels to level-off any cohort difference. Third, the expansion of schooling has been followed by changes in the educational system, affecting the comparability of responses over time. From 1960 to 2000, three education reforms occurred in Brazil (Rigotti 2004), two of them greatly change the education structure (Law Nº 5.692/71, and Law Nº 9.394/96). Before 1971, the structure of education system was divided into preschool, primary school (1-4 grades) with the possibility to add more two years of schooling, first cycle of high school (5-8 grades), second cycle of high school (1-3 grades), and higher education (at least 12 years of education). In 1971, the primary school (1-4 grades) and the first cycle of high school (5-8 grades) merged into what was called first level (1-8 grades). In addition, the terminology of the second cycle of high school changed to second level (1-3 grades). Later, in 1996, the first level changed again to Elementary education (1-8 grades), and the second level started to be called Secondary education (1-3 grades).

Not only that, census questions about educational attainment in Brazil have been modified over time. For instance, in the 1991 census the main question was "What was the last level passed?", while in 2000 it became "What is the highest course attended, in which you completed at least one grade?". Studies for other countries suggest that even a small change on question wording across census collections can make people answer differently about their own educational attainment (Lerch et al. 2017).

Therefore, our goal in this paper is to offer evidence of education misstatement by evaluating education reporting among adults in Brazil with data from the 1991 and 2000 censuses. In what follows, we first examine the quality of education data, directly, by looking at three different indicators: the completeness of education data, education misstatement, and educational levels by sex. Next, to provide further indirect evidence

of education misstatement both in 1991 and 2000, we calculate the implicit mortality measures by education from intercensal survivorship ratios.

2. Direct Evidence of Education Misstatement

2.1. Data

We draw data for men and women, 40-89 years old, from the 1991 and 2000 census in Brazil (IPUMS-I 2015). We select only individuals who were not attending school at the census time, which represent about 99.5% of the population aged 40 to 89 years in 1991 and 97.3% in 2000.

We measure educational attainment according to the highest grade completed within the most advanced level attended in the education system. In the 1991 census, people were first asked about the last graded passed, and then the last level passed. In 2000, the first question was about the highest course attended, in which you completed at least one grade, and next people were asked about the last grade passed. Changes in the questionnaire may be a source of data inconsistency between the two censuses.

2.2. The completeness of education information

An important source of data inaccuracy is the proportion of the data that is of missing. In both the 1991 and 2000 censuses, the Brazilian bureau of census imputed the missing cases on education by using the Felling-Holt model (IBGE 2002). Although is not possible to identify the imputed cases in the 1991 census, data from the 2000 census reveal that only 8% of the information about the highest course attended was imputed for the whole population. The apparent high response rate suggests that missing in not a major issue for education. However, compared to missing data for age, the proportion of education data that is missing is 16 times larger in the 2000 census, suggesting that lack of knowledge about individual characteristics can be a critical quality issue.

For the population aged 40-89 years, the proportion of imputed data about the highest course attended is similar for both sexes (about 4.2%). This proportion reduces with age: 6.2% for the age group 40-49, and 4.1%, 3.8%, 3.3%, and 2.7%, respectively, for the age groups 50-59, 60-69, 70-79, and 80-89. According to Reyes (2016), response rates for questions about personal characteristics tend to be higher for older people. In the case of Brazil, one explanation is that older people are more likely to be at home

during the census interview, which of course may reduce missing but it is not a guarantee that the information provided is accurate. Indeed, in the 2000 census, about 45% of people aged 40-89 years provided their own education. The proportion of people who declared their own information increases from 43.3% to 49.5% between the age groups 40-49 and 60-69, but then, it reduces to 37.34% for persons ages 80 to 89.

To cast some light on the effect of changes in the structure of the education system on data quality, we calculate the proportion of imputed data about the highest course attended according to the old (before the laws Nº 5.692/71 and Nº 9.394/96) and the current education terminologies. This is possible because in the 2000 census, response choices for this question were provided for both terminologies. Table 1 reveals that the proportion of imputation is substantially larger for individuals who reported education according to the current terminology. For example, at the age group 40-49, the proportion of imputed data is 18 times larger for individuals who declared according to the current terminology compared to those who reported according to the old one. This discrepancy greatly increases with age. Therefore, the results presented in Table 1 suggest a high degree of inaccuracy for individuals who declared their highest course attended according to the current terminology, particularly at older ages. One possible explanation is that people are more likely to remember their educational attainment according to the education system at the time when they were attending school. Fortunately, 74% of people ages 40 to 89 years reported their education according to the old terminology, which seems to have better data quality than information for the current terminology.

Table 1: Proportion of imputed data about the highest course attended by education terminology, both sexes, 2000, Brazil.

Age-Group	Education Terminology	
	Old *	Current **
40-49	0,92%	16,62%
50-59	0,68%	21,05%
60-69	0,66%	31,08%
70-79	0,73%	38,25%
80-89	0,74%	42,73%

Notes: * primary school, 1st cycle of high school, and 2nd cycle of high school.
** Elementary and secondary education
Source: Census data (IPUMS I – 2015)

We also calculate the proportion of the data that is missing regarding the question about the last grade passed; about 3.6% of people ages 40 to 89 years did not provide information about their last grade passed, a quite similar proportion of that regarding the highest course attended (4.1%).

Finally, we investigate the proportion of imputed data by single year of education. This analysis, not shown here, reveals very high proportion of imputed information for the 6th grade of the former primary education (about 30%). This grade was facultative until the 1971 education reform, and possibly, the imputation system does not consider this aspect, which may explain the higher proportion of imputation.

2.3. Education Overstatement

Another relevant source of data inaccuracy is education misstatement. We identify both in 1991 and 2000, the case when individual answered his/her last grade passed above the possible higher grade covered by the course attended, such as when a person declared 4th grade of the Secondary education (and 3th is the maximum possible). Unfortunately, we cannot identify individuals who made the mistake in the opposite direction, since those who possibly understate their grade of education were included in a regular grade of the course attended. Fortunately, the proportion of education overstated looks small (Table 2), but it is higher for women. For example, the proportion of women ages 40 to 49 years who overstated their last grade passed of the 2nd cycle or Secondary education (grades 1-3) was 2.4 times higher than that of men in 1991. Some studies that have looked at the quality of data on age in Brazil, have also found that women are more likely to exaggerate their age than men (Popolo 2001; Agostinho 2009, Gomes and Turra 2009).

Table 2 –Proportion of individuals who overstated their last grade passed by the highest course attended, both sexes, 1991-2000, Brazil.

Age group	1st cycle of high school		2nd cycle of high school or Secondary Education	
	1991	2000	1991	2000
Male				
40-49	0,00%	0,00%	0,12%	0,24%
50-59	0,00%	0,00%	0,12%	0,20%
60-69	0,05%	0,00%	0,10%	0,18%
70-79	0,19%	0,32%	0,08%	0,19%
80-89	0,17%	0,66%	0,08%	0,13%
Female				
40-49	0,00%	0,00%	0,29%	0,36%
50-59	0,00%	0,00%	0,21%	0,31%
60-69	0,09%	0,00%	0,19%	0,22%
70-79	0,20%	0,34%	0,19%	0,23%
80-89	0,17%	0,54%	0,22%	0,29%

Source: Census data (IPUMS I – 2015)

In addition, Table 2 shows higher proportions of education overstatement in 2000 than in 1991. The change in education system structure, hold in 1996, may contribute to this aspect, making people misstate their grade of education.

2.4. Sex Ratio by Educational Attainment

In the absence of migration, sex ratio declines with age as women outlive men. Table 3 presents the sex ratios by educational categories in 1991 and 2000. They reduce by age for all educational groups, except for individuals with 12 years of schooling or more. Moreover, among the most educated, men outnumber women at all ages, except at the age group 40-49 in 2000. One explanation for this pattern may be the gender gap in education, since women were less likely to attend college in the past. According to Gustavus and Nam (1968), in the U.S., in 1960, men with college slightly outnumbered women at the same educational attainment after age 45; however, the sex ratio declined at older ages as a result of mortality differential by sex. In Brazil, the magnitude of the education gap is too large, especially at older ages, to be true. For example, the sex ratio for those with 12 years or more of schooling at age group 80-89 is about two times higher than that of the age group 40-49 in both census years.

For the educational categories 4-7 and 8-11 years of schooling, men somewhat outnumbered women at younger age groups in 1991, but in 2000, probably because of

the expansion of education system, the sex ratios decline to below 1. It is also important to mention the irregular age pattern of sex ratio for persons with 8-11 years of schooling and among persons in the adult literacy category. An example is the increase in sex ratio between the age group 40-49 and 50-59, followed by a decrease between the age groups 50-59 and 60-69 at educational category 8-11 years of schooling, in both censuses. This unexpected age pattern may indicate data inaccuracy; female education overstatement among women at age group 50-59 may account for it.

Table 3 – Sex Ratio by educational attainment, Brazil, 1991 and 2000

Age-group	Educational Attainment					
	0-3	4-7	8-11	12+	Adult Literacy	Total
	1991					
40-49	0,883	1,000	0,974	1,178	1,381	0,953
50-59	0,849	0,994	1,004	1,510	1,096	0,931
60-69	0,834	0,922	0,918	1,913	0,986	0,884
70-79	0,816	0,824	0,748	2,491	1,052	0,833
80-89	0,706	0,663	0,589	2,802	0,992	0,706
	2000					
40-49	0,950	0,965	0,942	0,927	1,177	0,937
50-59	0,864	0,968	0,946	1,132	0,995	0,921
60-69	0,808	0,883	0,909	1,387	0,981	0,862
70-79	0,786	0,761	0,750	1,716	0,764	0,804
80-89	0,690	0,614	0,552	1,848	0,633	0,687

Source: Census data (IPUMS I – 2015)

3. Indirect Evidence of Education Misstatement

In this section, we provide additional evidence of education misstatement by calculating survivorship ratios (*SR*) by age, sex and educational attainment. The *SR*, which is a measure of changes in the cohort size during the intercensal period, should provide a reasonable measure of mortality if data quality does not vary across censuses, and net migration is close to zero.

The *SR* is very simple to calculate, requiring the distribution of population by age, sex and education from successive census:

$$SR_x^{k,i}(j) = \frac{N_{x+j}^{k,i}(t+j)}{N_x^{k,i}(t)} \qquad (1)$$

where $N_x^{k,i}(t)$ is the cohort x years old of 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 educational level i and sex k in the year $t+j$.

Although unexpected patterns for *SR* can be caused by variation in census coverage and in age misreporting, they also reflect inconsistencies in education reporting across the censuses. Therefore, *SR* can serve as indirect evidence of education misreporting.

To mitigate the effect of changes in census coverage on the *SR*, we adjust census enumeration according to the omission rates published by the Brazilian Institute of Geography and Statistics (IBGE 2016). Assuming international migration is not important at ages 40 and older (Carvalho 1996 and Carvalho & Campos 2006), any remaining distortions in *SR* by educational attainment should be mostly related to both age and education misstatement.

Tables 4 and 5 present a series of *SR* by educational attainment (years of schooling) between 1990[1] and 2000 for men and women. When net migration is close to zero, *SR* should decline by age and be lower than 1, reflecting the effect of mortality on cohort size. However, *SR* is higher than 1 for individuals with 1, 5, 6, 9 and 10 years of schooling, meaning that the size of cohorts increased between censuses for these educational levels. In some cases, the increase in cohort size was substantial. For example, there were 36% more individuals ages 50 to 54 in 2000 with 1 year of schooling than ten years earlier in the age group 40-44. It is very unlikely that only variation in age misreporting explains this pattern. Therefore, these findings can be a result of *i*) individuals improving their schooling during the intercensal period, or *ii*) education misstatement in 1991 and/or 2000.

As aforementioned, we selected only individuals who were not attending school during the census interview in order to reduce the effect of education mobility on the *SR*. In addition, we did not include adults younger than 40 years old to further minimize the number of cohort members who acquire additional schooling between censuses. Indeed, between 1991 and 2000, the proportion of persons not attending school declined only slightly from 99.0% to 95.7% in the age group 40-44, and from 99.3% to 96.7% in the age group 45-49. At older ages, the proportions were quite similar in both censuses.

[1] We interpolated the population from 1991 to 1990 by applying intercensal growth rates.

In any event, the expansion and changes in the education system may play a role for the inconsistencies found. The educational levels 1, 5 and 9 correspond, respectively, to the beginning of the former primary education (1 to 4 years of schooling), the former 1st cycle of high school (5 to 8 years of schooling), and the secondary education (9 to 11 years of schooling). The implausible high *SR* at these educational levels suggests that there is a group of individuals in the first census that moved from 0,4 and 8 years of schooling to a higher level at the second census. In presence of education expansion, some individuals may refuse to classified themselves as illiterate, overstating their educational level, by falsely declaring their educational attainment as 1 year of education or more. A similar reason may apply for persons who had completed the former primary school and declared their educational attainment as having started the former 1st cycle of high school; and for individual who had completed the elementary school (1-8 years of schooling) and declared as having some secondary school.

Table 4 – Intercensal Survivorship Ratio by educational attainment for males 1990/2000, Brazil.

Age in 1990	Age in 2000	Educational Attainment (in years of schooling)														
		0	1	2	3	4	5	6	7	8	9	10	11	12+	Adult Literacy	Total
40-44	50-54	0.6199	1.3641	0.8824	0.8309	0.8137	1.3926	0.9772	0.8736	0.9237	1.0592	1.0042	0.9547	0.9277	0.3052	0.8915
45-49	55-59	0.6649	1.3109	0.8861	0.8232	0.7888	1.3932	0.9564	0.9232	0.9013	0.9884	0.9555	0.9266	0.8985	0.2704	0.8729
50-54	60-64	0.7045	1.2550	0.8857	0.8215	0.7640	1.3629	0.9636	0.8564	0.8645	0.9368	0.9824	0.8896	0.8674	0.3022	0.8585
55-59	65-69	0.7033	1.1963	0.8258	0.7580	0.7280	1.2755	0.9143	0.8229	0.8198	1.0614	1.1169	0.8073	0.8425	0.2599	0.8167
60-64	70-74	0.6480	1.0330	0.7076	0.6943	0.6469	1.1002	0.7512	0.7543	0.7151	1.0233	0.9861	0.7744	0.8064	0.2295	0.7366
65-69	75-79	0.5250	0.8232	0.5766	0.5715	0.5418	0.7211	0.7549	0.7833	0.5633	1.2794	0.7636	0.6064	0.6804	0.1803	0.6046
70-74	80-84	0.4305	0.6537	0.4680	0.4719	0.4465	0.6868	0.4147	0.5631	0.4695	0.7562	0.6240	0.5256	0.5308	0.1591	0.4890
75-79	85-89	0.3257	0.4700	0.3580	0.3197	0.3395	0.5209	0.3761	0.4109	0.3627	0.4453	0.3312	0.3443	0.3960	0.0988	0.3636

Note: The 1991 population was moved to September 1, 1990, and the 2000 population was moved to September 1, 2000.
Source: Census data (IPUMS I – 2015)

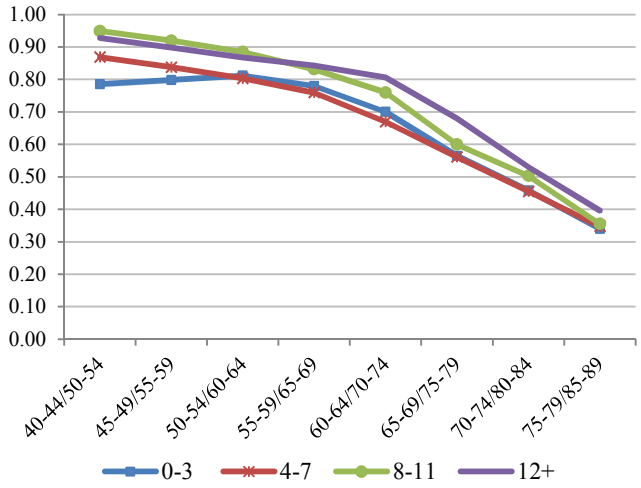
Table 5– Intercensal Survivorship Ratio by educational attainment for females 1990/2000, Brazil.

Age in 1990	Age in 2000	Educational Attainment (in years of schooling)														
		0	1	2	3	4	5	6	7	8	9	10	11	12+	Adult Literacy	Total
40-44	50-54	0.6333	1.3648	0.9330	0.8662	0.8321	1.4636	1.0192	0.9434	0.9853	1.0414	1.0125	0.9778	0.9795	0.3756	0.9240
45-49	55-59	0.7047	1.3644	0.9352	0.8475	0.8235	1.4928	1.0645	0.9584	0.9353	1.0723	1.0623	0.9370	0.9198	0.4302	0.9113
50-54	60-64	0.7582	1.4270	0.9409	0.9160	0.8702	1.6034	1.0496	0.9868	0.9671	1.1711	1.1564	0.9677	0.9543	0.3473	0.9436
55-59	65-69	0.7332	1.2096	0.8949	0.8300	0.8149	1.5035	1.0153	0.9486	0.8961	1.2151	1.1665	0.9307	0.9103	0.2849	0.8724
60-64	70-74	0.6618	1.1082	0.8193	0.7801	0.7679	1.3576	0.9942	0.9078	0.8662	1.1868	1.1809	0.8828	0.8404	0.2924	0.7979
65-69	75-79	0.5618	0.8957	0.6997	0.7013	0.6902	0.9193	0.9480	0.8432	0.7753	1.4946	0.8663	0.8178	0.8628	0.2394	0.6863
70-74	80-84	0.5104	0.8391	0.5831	0.5813	0.6045	0.7046	0.9189	0.5701	0.6571	0.7871	0.6282	0.7067	0.7509	0.2233	0.5968
75-79	85-89	0.3946	0.5690	0.4318	0.4144	0.4735	0.5118	0.6026	0.5242	0.4458	0.5627	0.5407	0.5526	0.4744	0.2273	0.4473

Note: The 1991 population was moved to September 1, 1990, and the 2000 population was moved to September 1, 2000.
Source: Census data (IPUMS I – 2015)

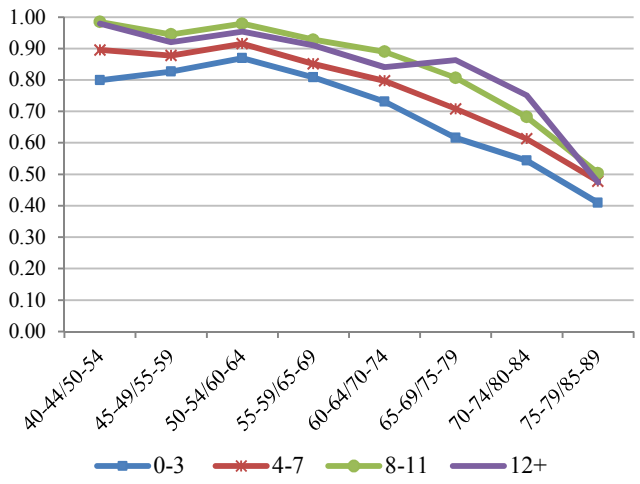
Some of these effects of education misstatement should be mitigated when schooling years are classified into education categories rather than schooling years. Hence, we re-calculate *SR* for individuals with 0-3 years of schooling, 4-7 years of schooling, 8-11 years of schooling, and 12 years of schooling or more (Figures 1 and 2). Not surprising, the *SR* series are now smoother and there are no values higher than 1. In addition, the educational gradient in mortality, i.e. the negative relationship between education and mortality, shows up. But there are still some implausible results, including similar values of *SR* for the educational categories 0-3 and 4-7 years (men), and for the categories 8-11 and 12+ (both men and women). Also, *SR* do not decline monotonically with age, particularly at ages younger than 60, which again could be explained, at least partially, by the true intercensal education mobility among young adults or intercensal reporting inconsistencies.

Figure 1 - Intercensal Survivorship Ratio by educational groups for males, 1990/2000, Brazil.



Source: Census data (IPUMS I – 2015)

Figure 2 - Intercensal Survivorship Ratio by educational groups for females, 1990/2000, Brazil.



Source: Census data (IPUMS I – 2015)

Therefore, as a final step to reduce additional errors, we categorize age in decennial intervals, hoping to minimize the effect of age misreporting particularly at older ages, and the effect of (true or false) education mobility at ages younger than 60. To better capture any remaining data inconsistency across subgroups, we estimate the West mortality level and life expectancy at age 40 (Coale & Demeny 1983) implied by the *SR* for each educational category (Table 6). The estimates show that for individuals with 0-

3 and 4-7 years of education, mortality levels implied by the *SR* vary significantly across cohorts. Among the lowest educated women, the implicit life expectancy at age 40 increases by almost 10 years when comparing the cohorts 40-49/50-59 and 70-79/80-89; similar results are found for men and for the group 4-7 years of education. These results suggest that younger cohorts experienced higher overall mortality levels in the 1990-2000 period than older cohorts; a finding that is unlikely since it contradicts the mortality transition in Brazil.

Table 6 – Intercensal Survivorship Ratio by educational groups and decennial age groups, both sexes, 1990/2000, Brazil.

Age in 1990	Age in 2000	0-3			4-7			8-11			12 +		
		Survival Ratio	Implied West Level	Implied e_{40}	Survival Ratio	Implied West Level	Implied e_{40}	Survival Ratio	Implied West Level	Implied e_{40}	Survival Ratio	Implied West Level	Implied e_{40}
Male													
40-49	50-59	0,7914	7,88	22,48	0,8554	12,75	26,13	0,9375	20,69	31,76	0,9161	19,44	30,59
50-59	60-69	0,7966	11,83	27,03	0,7845	13,23	27,11	0,8630	20,93	32,12	0,8574	20,09	31,56
60-69	70-79	0,6381	16,50	29,48	0,6261	14,03	28,76	0,6968	21,21	32,55	0,7579	23,66	35,50
70-79	80-89	0,4106	21,76	33,40	0,4151	22,01	33,78	0,4492	23,82	36,55	0,4775	24,03	36,04
Female													
40-49	50-59	0,8125	5,91	23,31	0,8878	8,34	26,37	0,9690	20,17	34,40	0,9573	19,59	34,37
50-59	60-69	0,8403	12,46	29,45	0,8867	18,18	32,89	0,9577	23,42	38,01	0,9375	22,92	37,21
60-69	70-79	0,6795	13,02	30,81	0,7609	19,79	34,71	0,8573	24,56	39,93	0,8483	24,3	39,51
70-79	80-89	0,4875	21,89	42,21	0,5600	24,36	39,42	0,6117	> 25	> 40.64	0,6464	> 25	> 40.64

Note: The 1991 population was moved to September 1, 1990, and the 2000 population was moved to September 1, 2000.
Source: Census data (IPUMS I – 2015)

Nevertheless, the results are consistent in showing that the group with highest education (12 years or more of schooling) has higher life expectancy at age 40 than the less educated group. For instance, for the cohort 60-69/70-79, men with 12 or more years of schooling can expect to live about 6 years more at age 40 than men with 0-3 years of education. Among women, the differences are even higher, of about 9 years. Consistently, the difference to the highest educated group reduces for individuals with 4-7 years of education, but disappears for individuals with 8-11 years of schooling which may be true or reflect some data artifact that we were not able to avoid with data categorization.

4. Conclusion

The relationship between education and demographic variables is increasingly documented. However, the validity of education data is often taken for granted. In data

deficient countries, as the case of Brazil, most of analysis on content errors has been focused on age misstatement, and little is known about the accuracy of other variables. Therefore, we assessed the quality of census education data, in Brazil, by showing direct and indirect evidences of education misstatement. Both results suggested that education census data is not accurate, particularly among women, supporting analysis on content errors that revealed inaccuracy of reported characteristics such as age. Indeed, based on the proportion of imputed cases, our results suggested that education data seems to be less accurate than age information.

Our analysis of imputed data also revealed a high degree of inaccuracy for individuals who reported education according to the current terminology, particularly at older ages, for whom more than 40% of information was imputed, greatly reducing the reliability of data for this subgroup. However, as most of adults (about 74%) chose the old terminology at declaring education in 2000, a way to reduce the measurement error in demographic analysis is to select the subgroup of people who declared their education according to the old terminology.

To provide further support to our results, the indirect evidences of education misstatements, based on survivorship ratios, gave clear evidences that census education data is not accurate; in addition, we showed that education data is not consistent between censuses. Although the *SR* by educational attainment suggested data quality issues, they are not immune to effects of migration and education mobility during intercensal period. On the other hand, it is very unlikely that only intercensal migration produced the *SR* distortions presented in our results, which suggested that people misstated their educational attainment, or individuals improved their schooling during intercensal period.

We also suggested the categorization of educational attainment in order to mitigate the effects of educational misstatement on the *SR*. After that, our results became more plausible; but, we still found some data inconsistencies. Thus, this study is an important step to gain an understanding of the quality of education data in Brazil; however, further analysis is needed. It remains unknown the magnitude of various types of errors on education data, and the relevance of these issues since they contribute to bias in population statistics. Therefore, the following steps of this research will be the

comparison of education information among different sources of data, like Education Census and Brazilian National Household Sample Survey.

References

Agostinho,C. S. (2009) *Estudo sobre a mortaliade adulta, para o Brasi entre 1980 e 2000 e unidades da Federação em 2000: uma aplicação dos métodos de distribuição de mortes*. [PhD thesis]. Federal University of Minas Gerais, Demography Program.

Black, D., Sanders, S., and Taylor, L. (2003) Measurement of higher education in Census and Current Population Surveys. *Journal of the American Statistical Association*. 98(463), 545-554

Battistin, E., De Nadai, M., & Sianesi, B. (2014). Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics*, 181(2), 136–150

Bongaarts, J. (2010). The causes of educational differences in fertility in Sub-Saharan Africa. *Vienna Yearbook of Population Research*, *8*, 31–50.

Carvalho, J.A.M. (1996) *O saldo dos fluxos migratórios internacionais no Brasil na década de 80: uma tentativa de estimação*. In: Migrações Internacionais: herança XX, agenda XI. Campinas: FNUAP.

Carvalho, J.A.M. & Campos (2006) A variação do saldo migratório internacional do Brasil. Estudos Avançados, São Paulo: Universidade de São Paulo - USP, Instituto de Estudos Avançados, 20(57), 55-58.

Cavenaghi, S. M., & Alves, J. E. D. (2016). Qualidade das informações sobre fecundidade no Censo Demográfico de 2010. Revista Brasileira de Estudos de População, 33(1), 189-205.

Coale, A. J., Demeny, P.,& Vaughan, B. (1983) Regional Model Life Tables and Stable Populations (Second Edition). New York/London: Academic Press

De Haas, H. (2010). Migration and Development: A Theoretical Perspective 1. *International Migration Review*, *44*(1), 227–264.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979-85. *Social science & medicine (1982)*, *42*(1), 47–57.

Folger, J. & Nam, C.(1964) Educational Trends from Census Data. *Demography*, 1(1), 247-257.

Fundação Instituto Brasileiro de Geografia e Estatística - IBGE (2001). Tendências demográficas no período de 1950/2000.

Fundação Instituto Brasileiro de Geografia e Estatística – IBGE (2002). Censo Demográfico 2000. Metodologia do Censo Demográfico 2000.

Fundação Instituto Brasileiro de Geografia e Estatística – IBGE (2016). Retroprojeção da população do Brasil por sexo e idade: 2000-1980.

Gomes, M. M. F., & Turra, C. M. (2009). The number of centenarians in Brazil: Indirect estimates based on death certificates. *Demographic Research*, *20*, 495–502.

Gustavus, S. & J. & Nam, C.(1968) Estimates of the "true" educational distribution of the adult population of the United States from 1910 to 1960. *Demography*, 5(1), 410-421

Horta, C.J.G. (2005) *Á procura dos padrões de mortalidade no Brasil*. [PhD thesis]. Federal University of Minas Gerais, Demography Program.

Horta, C.J.G. (2012) Idade Declarada nos Censos Demograficos e a Qualidade da Informação. Paper presented at XVIII Encontro Nacional de Estudos Populacionais, Águas de Lindóia, Brazil, November 19-23 2012.

INEP. 2011. Censo da Educação Superior 2010.Divulgação dos principais resultados do Censo da Educação Superior 2010. Brasilia: Instituto Nacional de Estudos e Pesquisas Educacionais.

IPUMS-I (2015) Minnesota Population Center. Integrated Public Use Microdata Series. Minneapolis: University of Minnesota.

Johnson-Greene, D., Dehring, M., Adams, K. M., Miller, T., Arora, S., Beylin, A., Brandon, R. (1996) Accuracy of self-reported educational attainment among diverse patient populations: a preliminary investigation. *Achieves of Clinical Neuropsychology.* 12(7), 635-643 .

Kane, T. J., Rouse, C. E., Staiger, D. (1999) Estimating returns to schooling when schooling is misreported. Cambridge: NBER Working Paper Series (WP 7235).

Koch, E., Romero, T., Manríquez, L., Paredes, M., Ortúzar, E., Taylor, A., et al. (2007). Desigualdad educacional y socioeconómica como determinante de mortalidad en Chile: análisis de sobrevida en la cohorte del proyecto San Francisco. *Rev Méd Chile*, *135*, 1370–1379.

Lerch, M., Spoerri, A., Jasilionis, D., & Viciana Fernandèz, F. (2017). On the plausibility of socioeconomic mortality estimates derived from linked data: a demographic approach. *Population Health Metrics*, 15(1), 26-41.

Lewis, J. R. (1986). International labour migration and uneven regional development in labour expoting countries *Tijdschrift voor Economische en Sociale Geografie*, *77*(1), 27–41.

Lima, E. E. C., & Queiroz, B. L. (2014). Evolution of the deaths registry system in Brazil: associations with changes in the mortality profile, under-registration of death counts, and ill-defined causes of death. *Cadernos de Saúde Pública*, 30(8), 1721–30.

Mackenbach, J. P., Kunst, A. E., Groenhof, F., Borgan, J. K., Costa, G., Faggiano, F., et al. (1999). Socioeconomic inequalities in mortality among women and among men: an international study. *American journal of public health*, *89*(12), 1800–6.

Martin, T. C. (1995). Women's Education and Fertility: Results from 26 Demographic and Health Surveys. *Studies in Family Planning*, *26*(4), 187.

Merrick, T.W. & Berquó, E.S. (1983) The determinants of Brazil's recent rapid decline in fertility. Washington: National Academy

Paes, N.A. & Albuquerque, M.E. (1999) Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. Revista de Saúde Pública. 33(1), 33-43

Paes, N.A. (2007) Qualidade das estatísticas de óbitos por causas desconhecidas dos estados brasileiros. Revista de Saúde Púlica. 41, 436-445

Perez, E. R. (2010). *Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil*. [PhD thesis]. Federal University of Minas Gerais, Demography Program.

Perez, E. R., & Turra, C. M. (2009). Measuring educational differences in mortality among women living in highly unequal societies with defective data: the case of Brazil (No. 348). Belo Horizonte.

Popolo, F. Del. (2001). Problemas en la declaración de edad de la población adulta mayor en los censos. *Notas de Población*, *72*, 73–122.

Preston, S., & Taubman, P. (1994). Socioeconomic Differences in Adult Mortality and Health Status. In L. G. Martin & S. H. Preston (Eds.), *Demography of Aging* (pp. 279–318). Washington, D.C.: National Academy Press.

Rigotti, J. I. R. (2004) *Variáveis de educação dos censos demográficos brasileiros de 1960 a 2000.* In: Introdução á demografia da educação. Associação Brasileira de Estudos Populacionais.

Rios-Neto, E. L. G.(2000) Passado, presente e futuro da fecundidade: uma visão de idade, período e coorte. *Revista Brasileira de Estudos da População*. 17(2), 5-15

Rios-Neto, E. L. G., Guimarães, R. R. de M., Pimenta, P. S. F., & Moraes, T. de A. (2010). *Análise da evolução de indicadores educacionais no Brasil: 1981 A 2008* (No. 386). Belo Horizonte.

Ribeiro, M. M. (2016). *Mortalidade adulta por níveis de escolaridade no estado e no município de São Paulo: uma proposta de estimação a partir do Censo Demográfico de 2010*. [PhD thesis]. Federal University of Minas Gerais, Demography Program.

Rosero-Bixby, L., & Dow, W. H. (2009). Surprising SES Gradients in Mortality, Health, and Biomarkers in a Latin American Population of Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 64B(1), 105–117.

Rostron, B.L., & Arias, E. (2010) Education reporting and classification on death certificates in the United States. National Center for health Statistics. *Vital Health Stat*

2(151), 1-14.

Shai, D., & Rosenwaike, I. (1989).Errors in reporting education on death certificates: some findings for oldr male decendents from New York State and Utah. *American Journal of Epidemiology*, 130(1), 188–192.

Shkolnikov, V. M., Jasilionis, D., Andreev, E. M., Jdanov, D. A., Stankuniene, V., & Ambrozaitiene, D. (2007). Linked versus unlinked estimates of mortality and length of life by education and marital status: Evidence from the first record linkage study in Lithuania. *Social Science & Medicine*, 64(7), 1392–1406

Singh, S & Casterline, J. (1985) The socioeconomic determinants of fertility. In: *Reproductive change in Developing Countries - Insights from World fertility Surveys*. Oxford University Press.

Sorlie, P. D., & Johnson, N. J. (1996). Validity of education information on the death certificate. *Epidemiology* (Cambridge, Mass.), 7(4), 437–9.

Turra, C.M. (2012) Os limites do corpo: A longevidade em uma perspectiva demográfica. *Revista da UFMG*. 19(2), 156-181

U.S. Bureau of the Census, U.S. Census of Population: 1960, (1963) Subject Reports, Educational Attainment. Final Report PC (2)-5B. U.S. Government Printing Office, Washington, D.C.

Zhu, H., & Xie, Y. (2007). Socioeconomic Differentials in Mortality Among the Oldest Old in China. *Research on Aging*, *29*(2), 125–143.