Life expectancy, adult mortality and completeness of death counts in Brazil and regions: comparative analysis of IHME, IBGE and other researchers estimates of levels and trends

Bernardo Lanza Queiroz
lanza@cedeplar.ufmg.br

Marcos Roberto Gonzaga
mrcs.roberto@gmail.com

Ana Maria Nogales Vasconcelos
nogales@unb.br

Bruno Torrente
brunotorrente230198@gmail.com

Daisy Maria Xavier de Abreu
dmxa@medicina.ufmg.br

Short Abstract

Estimates of completeness of death registration are crucial to produce estimates of life tables, population projections and to the global burden of diseases study. They are an imperative step in quality of data analysis. In the case of state level data in Brazil, it is important to consider spatial and temporal variation in the quality of mortality data. In this paper, we compare and discuss alternative estimates of completeness of death registration, adult mortality (45q15) and life expectancy estimates produced by the National Statistics Office (IBGE), Institute for Health Metrics and Evaluation (IHME) and estimates presented in Queiroz, et.al (2017) and Schmertmann and Gonzaga (2018), for 1980 and 2010. We find significant differences in estimates that affect both levels and trends of completeness of adult mortality in Brazil and states. IHME and Queiroz, et.al (2017) estimates converge in 2010, but there are large differences when compared to estimates from the National Statistics Office (IBGE). Larger differences are observed for less developed states.

Structured Abstract

Background: Estimates of completeness of death registration are crucial to produce estimates of life tables, population projections and to the global burden of diseases study. They are an imperative step in quality of data analysis. In the case of state level data in Brazil, it is important to consider spatial and temporal variation in the quality of mortality data. There are three main sources of data quality evaluation in Brazil, but there are very little comparative studies across them and how they evolve over time and across states in the country.

Objectives: Compare and discuss alternative estimates of completeness of death registration, adult mortality (45q15) and life expectancy estimates produced by the National Statistics Office (IBGE), Institute for Health Metrics and Evaluation (IHME) and estimates presented in Queiroz, et.al (2017) and Schmertmann and Gonzaga (2018), for 1980 and 2010.

Methods: We provide a descriptive and comparative analysis of different estimates from four (4) data sources by state level (27 states) in Brazil in two different points in time. We focus on completeness of death counts, adult mortality and life expectancy at birth.

Results: We find significant differences in estimates that affect both levels and trends of completeness of adult mortality in Brazil and states. IHME and Queiroz, et.al (2017) estimates converge in 2010, but there are large differences when compared to estimates from the National Statistics Office (IBGE). Larger differences are observed for less developed states.

Conclusion: We show that the quality of mortality data in Brazil improved steadily overtime, but with large regional variations. However, we observed that IBGE estimates show the lowest levels of completeness to the Northern part of the country compared to other estimates. Choice of methods and approaches might lead to very unexpected results and conclusion.

Contribution: We produce a detailed comparative analysis of estimates of completeness of death registration by different sources and discuss the main results and possible explanations for the differences. We show that new improved methods are still in demand to study adult mortality in less developed economies and at the local level. More comparative studies are important to improve quality of estimates in Brazil.

1.Introduction

Adequate knowledge of mortality levels in a country and its regions is necessary for an efficient planning of public policies, especially health and social security, to better understand the impacts of regional (economic and social) differentials and also to carry out work of research in different areas of knowledge, specifically demography, economics and public health. However, studies on adult and general mortality in developing countries, such as Brazil, have been limited by the lack of good quality data, problems with completeness of death counts registration and quality of population information, especially for subnational and local areas. Obtaining mortality estimates becomes a challenge, since the quality of the information is generally not satisfactory (Dorrington, 2012a; 2012b; Hill and Queiroz, 2010; Hill, You and Choi, 2009, Hill, Timaeus and Choi, 2005). In general, the problems encountered in the data for the country are aggravated when dealing with population subgroups and / or at a subnational level (Ahmad and Hill, 2011; Schmertmann and Gonzaga, 2018; Alexander et.al, 2017). Even in developed countries, there are differences in the quality of vital statistics at subnational level (and population sub-groups). These differences can be attributed mainly to the degree of economic development of each region (Kulkarni, *et. al*, 2011).

The difficulties in the data sources used in the estimation of mortality are related to the level of completeness of death registration and errors in the declaration of age, for both deceased and population. In order to circumvent these data problems, different methodologies have been developed to measure adult mortality, but these methods are considered limited for subnational estimates (Queiroz, et.al, 2017; Murray, et.al, 2010; Luy, 2010; Hill et al., 2009; Agostinho and Queiroz, 2010;). More recent research, combine demographic methods to Bayesian models to estimate completeness of death counts and produce mortality estimates,

but mostly to developed countries or one point in time (Alexander, et.al, 2017; Schmertmann and Gonzaga, 2018).

In Brazil, there is an array of estimates of completeness of death registration and adult mortality. Different agencies and authors use alternative methods and approach to produce those estimates that are extremely relevant to public health policies, population forecasts and general planning. Even though there are great differences among those estimates, there is little comparative analysis among them (Schmertmann and Gonzaga, 2018; Queiroz, et.al, 2017). Thus, it is relevant to perform a systematic analysis of different estimates and compare and contrast the results, providing some guidance on how methods and data can explain the observed differences.

In this paper, we review the evolution of completeness of death counts registration in Brazil, adult mortality, 45q15, and life expectancy at birth over the last four decades by state level, based on three different sources: i) Queiroz, et.al (2017); ii) Brazilian National Statistics Office (IBGE); iii) the Institute of Health Metrics and Evaluation (IHME). We also compare, in more detailed, estimates from 2010 from the previous three sources to the ones produced by the Bayesian model proposed by Schmertmann and Gonzaga (2018). The analysis of sub-national estimates in Brazil, annex 1 shows the list of states and some basic information for each area, is very relevant due to the large socioeconomic differences across states in the Northern part of the country compared to the southern part, that are closely related to the evolution of mortality and population data (França, et.al, 2017; Amaral, Queiroz and Calazans, 2015). We also discuss the differences in data and methods used in each and discuss the potential implications of the estimates to demographic studies. The main challenge is that we do not know the true level of mortality, but it is important to have transparency and replicability when using this information. We believe that this paper makes an important contribution by

discussion different estimates from different agencies in a comparative perspective. Specially estimates at the regional level that have large implications for public health planning.

2. Data and Methods

2.1. General Overview of Data and Methods

We provide a descriptive and comparative analysis of different estimates from four (4) data sources by state level (27 states) in Brazil in two different points in time. The comparative analysis is also developed by sex, so we can discuss possible differences across groups. We compare the evolution from 1980 to 2010, both censuses years in Brazil, to avoid the use of estimates population. We focus on completeness of death counts, adult mortality and life expectancy at birth. The first issue we focus on is the quality of mortality data, which is measured by the completeness of death counts coverage. From their estimates of completeness, it is possible to adjust mortality data and obtain estimates of adult mortality (45q15) and life expectancy at birth using traditional demographic methods (Wachter, 2014). This analysis is motivated by observed differences in state level estimates by the two different agencies and authors, as well as by the importance of mortality estimates to other studies and public health planning.

2.2. Death Distribution Methods

Since our first interest is on the completeness of death counts coverage, we discuss the main methods behind all the estimates analyzed in this pap, named Death Distribution Methods – DDM. These methods based on the population dynamics equations, compare the distribution of deaths by age with the age distribution of the population and provide the age standard of mortality for a defined period of time (Timaeus, 1991). There are three main methods of deaths

record evaluation: General Growth Balance (GGB) proposed by Hill (1987), Synthetic Extinct Generation (SEG) proposed by Bennett and Horiuchi (1981; 1984) and adjusted synthetic extinct generations, proposed by Hill, You and Choi (2009). The methods have very strong assumptions: 1) population is closed to migration, 2) the degree of completeness of deaths is constant by age, 3) there are no differences in completeness of the census data (population) by age and over time, and 4) there are no errors in the declaration of ages of the living or the deaths.

Generalized Growth Balance (GGB) (Hill, 1987) is a generalization of the Brass Balance Equation (1975) method and is applied to non-stable populations, but maintains the other assumptions. The GGB method is derived from the basic demographic balance equation, which defines the population growth rate as the difference between the entry rate and the exit rate of the population. When two or more censuses are available, the growth rate for each age group can be calculated, and the stability assumption is no longer necessary. In this method, both the completeness of the death registration and the relative coverage of the enumeration of two censuses are estimated at the same time.

The method assumes that the relationship between entrance and exit rates of the population occurs for any open age $x +$ interval (people aged x and over). In other words, in a closed population the entries occur as birthdays in the x -age. Thus, the difference between the entry rate at $x +$ and the population growth rate at $x +$ yields a residual estimate of the mortality rate at $x +$. The calculated difference between the entry rate in this age segment and the growth rate of this segment gives us an estimate of the mortality rate of the segment 15 years and over assuming that there are no significant migratory flows. This residual mortality estimate can be compared with the mortality rate for the same age group, calculated from census data or vital

records, and the differences between them can be analyzed as errors in in age declaration for both population and deaths counts

The Extinct Generations method developed by Bennett and Horiuchi (1981), part of the method developed by Preston et al. (1980), which is used when the population deviates from stability. Intuitively, the method shows that the population of a given age today should equal all the deaths that will occur to that population in the future; current deaths above a certain age can be adjusted by the specific growth rate of that age group by making the idea of a stable population more flexible. In this method, the estimate of underreporting of deaths is also given by the ratio between the estimated number of people in the age x ($P'(x)$) and the observed number of people in the age x ($P(x)$). The method uses age-specific growth rates to convert a distribution of death by age into an age distribution of population. Since in a population the observed deaths from a given age x are equal to the population of the age x , adjusted by the population growth rate by age interval, the deaths of a population in the age $x +$ provide an estimate of the population of that age x . The degree of coverage of the death registration system will be given by the ratio between the population estimated by the deaths above age x and the population observed above age x .

Hill, You and Choi (2009) and Murray, et.al (2010) based on a series of simulation estimates argue that the combination of both methods produce more robust results. Their main argument is that the GGB has a good ability to estimate the quality of a census relative to the other and would allow for the best fit of the method, and SEG is less sensitive to variation in data quality. The main caveat of the method is that they assume closed population for better use of estimates. Although there are methodologies in the literature that allow us to deal with this problem (Bhat, 2002 and Hill and Queiroz, 2010). as when working with aggregated country data and regional data, it is important to consider the possible effects of migratory

flows on the assessment estimates of data quality. Another alternative applied in regional studies using death distribution methods is to use an age range to estimate the degree of coverage that suffers little or no influence from migratory flows (Queiroz et al., 2009), methodology proposed by Bhat (2002) or Hill and Queiroz (2010).

2.3. Queiroz, et.al (2017)

Queiroz and collaborators (2017) evaluate the completeness of reported deaths using death distribution methods. They use the R-package (DDM), developed by Everton Lima, Tim Riffe and Bernardo Queiroz, and their estimates focus on censuses years (1980, 1991, 2000 and 2010). Population data, by age and sex, are obtained directly from the National Statistics Office (IBGE) (www.ibge.gov.br) and mortality data are obtained from the Mortality Information System (MIS) of Ministry of Health (available at www2.datasus.gov.br). MIS provides information on deaths by age, sex and causes of death at state level, since 1979. MIS data come from death certificate where causes of death are registered according to the international form recommended by the International Classification of Diseases. Data on causes of death are code using the Revision of the International Classification of Diseases (ICD) codes (9th, from 1980 to 1995, and 10th, from 1996 onwards).

2.4. IHME – Global Burden of Disease

The method used by IHME is available at healthdata.org and in Naghavi, et.al (2017). A more detail of the method used to estimate life tables is available in Dicker, et.al (2018). In summary, death reporting from vital registration and censuses were evaluated for completeness using improved death distribution methods. However, it is not very clear from the description of the method and the material available what is the improvement they added to the death distribution methods – the paper mentions developed based on simulations from Murray, et.al (2009), but does not specify age ranges used or what was the procedure when completeness

was above 100%. We were also not able to identify the codes and program used to estimate completeness for Brazil and regions. They also applied an improved sibling survival method to survey data on sibling survival modules to correct for survivor bias, zero-survivor bias, and recall bias. Estimates of under-5 and adult mortality were generated using the state-of-the-art data synthesis methods: a combination of spatio-temporal and Gaussian Process regressions.

2.5 IBGE

According to methodological reports and working papers, IBGE uses different methods and data to produce mortality estimates for state levels in Brazil since 1980 (IBGE, 2013; Albuquerque and Senna, 2005). For 1980, 1991 and 2000 censuses years, despite of the limited applicability of the method in non-stable populations and in the context of large migration flows between states in Brazil, IBGE used the Growth Balance Method (Brass, 1975) to estimate the completeness of death counts registration in most states. However, there are some specificities in their application. For females in Northeast region in 1980 and in North and Northeast regions in 2000, IBGE used other methods instead of Growth Balance (Albuquerque and Senna, 2005). Based on expert opinions, a reduction factor was applied on the undercount of deaths estimate in each year in order to consider a best quality of the deaths of the adult and elderly population in relation to the young and young adult population (Albuquerque and Senna, 2005). For infant deaths IBIGE used indirect demographic methods suggested by Brass (1975) and Trussel (1975).

In 2010 IBGE made some changes on the methodology to estimate life tables for Brazilian states. For South's states plus São Paulo, Rio de Janeiro and Distrito Federal deaths records were used without any correction due to undercount. In all other states, the General Growth Balance Method (Hill, 1987) was used. A top-down estimation procedure was applied in order

to make sure that the sum of deaths between states was equal to national total estimated deaths (IBGE, 2013)¹.

3. Results

First, we analyze the estimates of the completeness of death counts registration in Brazil and states from 1980 to 2010 produced by IHME. Figure 1 shows the estimates of completeness for both sexes for Brazil and states. The results show a process of convergence over time for all states in Brazil reaching above 90% of completeness in 2010. However, some of their estimates for the less developed areas of the country are unexpected when analyzed in light of development level and other estimates, as we observed states with similar levels of development with large differences in completeness since the 1980s. For instance, Piauí and Maranhão show the most impressive progress as they improve from 60% of completeness in 1980 to 100% in 2017. But, some states with very similar social and economic development, for instance Sergipe, Bahia and Paraíba, have estimates close to 100% since 1980s.

Figure 1 – Level of completeness of death registration, Brazil and states, both sexes, 1980 – 2010, IHME estimates

¹ More details about the IBGE to produce their life table estimates in 2010 are available at www2.ibge.gov.br/home/estatistica/populacao/tabuadevida/metodologia.shtm.

Source: IHME (2017)

Second, we use life table estimates from IHME to obtain life expectancy at birth overtime in Brazil and study the evolution of this metric by state. We can compare the evolution of life expectancy at birth since 1950 and identify which state, for males and females, had the highest life expectancy over time. Figure 2 shows states with highest life expectancy separated by sex. The most striking and surprising period is from 1975 to 2010, when Amapá, Piauí and Rio Grande do Norte present the highest life expectancy compared to more developed regions of the country. Other studies show that these three states are amongst the lowest life expectancy in Brazil overtime (IBGE, 2013; Souza et al, 2018). We argue that the result is related to the estimate of completeness presented in Figure 1 showing that some regions had surprising high levels of completeness in the 1980s and 1990s and as the correction factor for registered deaths depends on the completeness this led to very low mortality rates and high life expectancy. In the following sections, we presented a comparison of these estimates to other studies in Brazil.

Figure 2 – States with highest life expectancy in Brazil, by sex, 1980 to 2010 – based on IHME estimates.

We now focus our analysis in the comparison of the levels of completeness of death counts coverage produced by IHME to those produced by IBGE and Queiroz, et.al (2017). We focus our analysis on periods 1980/1991 and 2000/2010, those are intercensal censuses periods and it is possible to capture an important variation overtime and space using those two periods. Figure 3 shows estimates for males and females and by states in the two intercensal periods. There are several interesting patterns: a) IBGE shows the lowest estimates of completeness for states in the Northeast during the whole period of analysis, and there is hardly an improvement in data quality from 1980 to 2010. For instance, Maranhão has completeness estimated as below 80% in 2000/2010; b) IHME presents high level of completeness for almost all states since 1980, even for states with very low development levels; c) there is a pattern of slight underestimation between IHME and Queiroz et.al (2017), and the states that had the largest discrepancies were the less developed ones, like Acre, Amazonas and Maranhão (Figure 3); d)

estimates in 2000/2010 for IHME and Queiroz, et.al. (2010) are very similar for males and females and all states, but they are quite different from estimates produced by IBGE.

Another relevant issue while comparing Queiroz, et.al (2017) to IHME's is the fact that their estimates to the completeness of death counts never passed 1 (Figure 3), while we had estimates that passed even 1,2. This indicates that IHME estimates might be assuming that the impreciseness of the vital record can only be related to the deaths that were not recorded. However, situations with poor data might produce estimates of completeness above 100% indicating that underlying data is bad and not that registration of deaths are complete. Also, estimates above 100% might be indicating that the methods applied to those context are not stable and the strong assumptions one has to make do not hold there. The problem might be even worse, since they might a combination of both issues. We believe that this is one major point of future discussion and research.

Figure 3 – Completeness of Death Counts, Brazil – Queiroz, et.al, (2017), IHME and IBGE, 1980/1991 and 2000/2010.

Source: IHME (2016); Queiroz, et.al (2017), IBGE (2013) and Albuquerque and Senna (2005)

Based on estimates of completeness, all three sources also provide information on adult mortality, $_{45}q_{15}$, as the probability of dying between ages 15 and 60. Adult mortality is a simple measure and has the possibility of comparing estimates of adult mortality with other studies. We consider that the entry into adulthood occurs at age of 15 years; at that age, there is the inflection point in which the declining of childhood mortality risks is replaced by increased mortality risks for young adults and adults. In addition, this measure covers a substantive age

interval - up to the age of 60 years - and avoids problems inherent in estimates of mortality at more advanced ages. We present estimates comparing mostly estimates from Queiroz, et.al (2017), IBGE and IHME (Figures 4). As we did in Figure 3, we focus the analysis on estimates for males for the same earlier and later periods. The results for females show similar pattern in both periods.

In relation to the estimates from IHME compared to Queiroz, et.al, (2017), we observed largest differences in the 1980-1991 period and a convergence in the more recent periods (2000-2010). The main differences observed in the earlier period are in the northeastern states, the same one that presented largest differences in the completeness of death counts coverage estimates. As estimates of completeness converge overtime from the two sources, the mortality estimates also converge.

The main differences on the adult mortality probability (45q15) between the National Statistics Office (IBGE, 2016) and others (IHME and Queiroz, et.al, 2017) are from the period 1980/1990 (second and third panels from Figure 4). IBGE estimates of adult mortality were much higher than others in most Brazilians states during this period for males (Figure 4), the behavior of female estimates follows proportionally the same pattern in all comparisons. The IBGE estimates, combined with 5q0, would result in biased estimates for the life expectancy at birth for some states, and even more, some of those biased estimates were attributed to states that were known for their lack of development in their health systems, like Piauí, Maranhão and Tocantins. Estimates from Queiroz et al (2017), obtained from the several papers cited below, are closer to what is used by the UN Population Division and IHME. However, even in 2010 we observe large differences between IBGE estimates and other sources.

Figure 4 – Males Adult Mortality Probabilities, 45q15, Brazil – Queiroz, et.al (2017), IHME, and IBGE, 1980/1991 and 2000/2010

Source: IHME (2016); Queiroz, et.al (2017), IBGE (2013) and Albuquerque and Senna (2005)

The estimates are in a process of convergence with the ones produced by the National Statistics Office (IBGE) in the 2000/2010 period, but even though we can still notice a pattern of slight overestimation of adult mortality for IBGE compared to IHME and Queiroz et al (2017). One thing that can be pointed out observing the estimates from both sources is the fact that the most developed states, like São Paulo and Rio de Janeiro, are usually the ones that have both estimates pointing on the same direction, meanwhile some less developed states still have

a large discrepancy between estimates from Queiroz and collaborators and the ones produced by IBGE.

In order to perform a more detailed analysis of those estimates, we focus on results for 2010 for life expectancy at birth and adult mortality. Figure 5 shows a summary of estimates of life expectancy at birth from different sources for Brazil and states: direct calculations from MIS (grey open circle), estimates from IBGE (grey solid dot), IHME (red solid dot) and a Bayesian model (horizontal bars) proposed by Schmertmann and Gonzaga (2018), estimates for males in 2010. Queiroz, et.al (2017) do not have estimates of life expectancy at birth. The horizontal bars represent 80% credibility interval from the posterior distribution of life expectancy, from Schmertmann and Gonzaga (2018) and the states names abbreviation represents the median of the distribution. The effect of under-registration of death counts is observed when looking at estimates using direct data from MIS – no correction – that presents highest life expectancy for all states, mostly in the North and Northeast regions of the country. The regions with better data quality have life expectancy estimates very similar to that obtained using observed data. The second most striking result is the comparison of IBGE with other estimates. They are much lower in the North and Northeast regions and much higher in the South and Southeast.

The main explanations for the differences observed are related to the strategies and methods used to estimate the completeness of deaths and mortality rates. Schmertmann and Gonzaga (2018) show that the Bayesian method combines a relational model for mortality schedules with probabilistic prior information on death registration coverage derived from Death Distribution Methods (as estimated by Queiroz et al, 2017, and from field audits done by public health experts (Szwarcwald et al., 2010; Frias et al., 2013). It is important to point out that the Bayesian approach also included the IBGE estimates of undercount of deaths as

prior information for probability of death registration. We notice that for Southeast and South and some Center-West states the 80% credibility interval for life expectancy at birth overlap the estimates without any correction, which was waited since that in most of those state the probability of death registration is almost 100%.

In general, IHME estimates are more conservative in all North and Northeast states and in some Sul/Southeast states and closer to Bayesian estimates than IBGE ones. In most states, results from Bayesian model represent a balance between IBGE and IHME estimates. If we compare all estimates without any correction (open circle) we see that official life expectancy estimates (IBGE ones) for Southern and Southeastern states are implausibly high, even higher than life expectancy estimates using direct data from MIS. As point out by Schmertmann and Gonzaga (2018), it indicates that the IBGE estimates are plausible only if the vital registration system substantially overcounts deaths in these states.

One possible explanation of that overestimation of IBGE's life expectancy at Southern and Southeastern states would be related to the top-down estimation procedure they proposed. According to IBGE (2013), since the completeness of deaths estimate for each state and Brazil were made independently, the sum of estimated deaths across states was higher than the total estimated deaths for the whole country. Then, a top-down estimation procedure was applied in order to make sure that the sum of deaths between states was equal to national total estimated deaths. However, that procedure resulted in highest life expectancy at older ages in North and Northeast regions, compared to South and Southeast regions, which should be an implausible result since life conditions and health care system to the older population are better found and in greater quantity in South and Southeast regions (IBGE, 2013). Then, the IBGE solutions was to exclude North and Northeast regions of the top-down procedure. As a consequence, the

excess of estimated deaths across states was eliminated only in South and Southeast states leading to a higher life expectancy on those states.

Figure 5 – Males life expectancy at birth, 2010 – Bayesian estimate from Schmertmann and Gonzaga (2018), IHME, IBGE and MIS (Ministry of Health)

Source: IBGE (2013), IHME (2018), SIM/Ministry of Health (2018), Schmertmann and Gonzaga (2018).
Note: Shaded bars in different colors represent 80% posterior probability interval for Life expectancy at birth – $e(0)$ - from the Bayesian model; state abbreviations appear at posterior median; red solid dot represent IHME estimates, grey solid dot represent IBGE estimates and open circles represent unadjusted estimates from deaths registered by the Mortality Information System (MIS).

Figure 6 shows different estimates of adult mortality ($_{45q15}$) from alternative sources in 2010. The analysis of the results highlights the relevance of comparing and analyzing different mortality estimates in Brazil since they have large impacts on the estimation of the global burden of disease, population forecasting and general public health planning. Normally, it is argued that adult mortality estimates are more stable and consistent, however, the comparison indicates large differences across states in Brazil. Figure 6 follows the same structure of Figure

5, but now includes estimates from Queiroz, et.al (2017). Shaded bars in different colors represent 80% posterior probability interval for males adult mortality probability (45q15) from Bayesian model; state abbreviations appear at posterior median for 45q15; red solid dot represent IHME estimates, grey solid dot represent IBGE estimates, blue solid dot represent Queiroz et al (2017) estimates and open circles represent unadjusted estimates from deaths registered by the Mortality Information System (MIS).

The results show that for the North and Northeast states of the country IBGE estimates for adult mortality are much higher than any of the other studies in the analysis. The difference is very striking for some states such as Maranhão where IBGE estimates adult mortality around 0.29 for males compared to 0.17 from direct estimates (without correction), 0.18 from IHME, 0.19 from Queiroz, et.al (2017) and 0.22 from Schmertmann and Gonzaga (2018). The last estimate also presents a very large probability interval. On the other side, for states in the South and Southeast, where completeness of death counts coverage is close to 100%, IBGE estimates show adult mortality to be lower than direct computations from MIS. Queiroz, et.al (2017) show estimates a little bit higher than the ones from IHME and Schmertmann and Gonzaga (2018), but with relatively small differences and very close to what is obtain using direct data.

Figure 6 – Males Adult Mortality Probabilities ($_{45}q_{15}$), 2010 – Bayesian estimate from Schmertmann and Gonzaga (2018), Queiroz et al (2017), IHME, IBGE and MIS (Ministry of Health)

Source: IBGE (2013), IHME (2018), SIM/Ministry of Health (2018), Queiroz et al (2017), Schmertmann and Gonzaga (2018).

Note: Shaded bars in different colors represent 80% posterior probability interval for males adult mortality probability ($_{45}q_{15}$) from Bayesian model; state abbreviations appear at posterior median for $_{45}q_{15}$; red solid dot represent IHME estimates, grey solid dot represent IBGE estimates, blue solid dot represent Queiroz et al (2017) estimates and open circles represent unadjusted estimates from deaths registered by the Mortality Information System (MIS).

Lastly, an important issue of mortality estimates in Brazil arise when one studies the single-age mortality profile. Using data from Brazilian states in 2010, without any corrections, we notice an additional data problem at older ages that could explain the differences on ranking

of $e(0)$ from different methods. Comparisons of the schedule of mortality rates at older ages between states in 2010 highlight that age miss-reporting on both data sources, deaths and population, combined with undercounting of deaths in vital registration system, lead to a crossover of mortality rates at older ages between South and North/Northeast states.

Figure 7 shows the observed and smoothed males mortality rates by single ages for five selected states in 2010, two states in North regions (Tocantins and Piauí), two states in Northeast region (Alagoas and Maranhão) and Santa Catarina (located in the South region). The smoothed male mortality rates by age (solid dot) comes from the posterior distribution for mortality rates according to the Bayesian model proposed by Schmertmann and Gonzaga (2018). The observed rates by single ages (open circle) comes from vital registration system without any correction. Open and solid triangles come from IBGE estimates for five-years age interval². Then, comparisons of males age-specific mortality rates between Santa Catarina in the North/Northeast states shows a clear crossover of mortality rates starting around 50 or 60 years old. The crossover does not disappear even after corrections due to undercounting of deaths by age based on Bayesian model results, even in Maranhão state where the adjustment due to undercount of deaths is more evident.

The crossover of mortality rates at advanced ages between South and North/Northeast states has at least three possible causes. First, it would be possible that the completeness of deaths at advanced ages in North/Northeast states are lower than at adult and young ages. Second, it would be a consequence of different age misreporting patterns between South and North/Northeast states. Preston et al (1999) investigated how different types of age misreporting can affect estimates of mortality rates at older ages. The authors analyzed the effects of three patterns of age misreporting: net age overstatement, net age understatement,

² IBGE did not published UF estimates of mortality probabilities by single years of age.

and symmetric age misreporting. Based those three types of age misreporting and on five types of mortality estimates, they found that age misstatement biases mortality estimates downwards at advanced ages. The third cause would not be related to data quality but to the hypotheses of mortality selection. In the context of higher mortality rates at young ages, survivors to older ages would be physiologically stronger and then live longer than others. Finally, it would be possible that those crossover of mortality rates between South and North/Northeast states comes from different data problems and also from the hypothesis of mortality selection.

Figure 7 – Males log mortality rates by singles ages from Bayesian model and Mortality Information System and by five-years age interval from IBGE, for selected states in Brazil in 2010.

Source: IBGE (2013), Schmertmann and Gonzaga (2018).

Discussion

In this paper, we compared the IHME, IBGE, Queiroz, et.(2017) and Schmertmann and Gonzaga (2018) estimates of completeness of death counts coverage, adult mortality (45q15) and life expectancy at birth from 1980 to 2010 by state level in Brazil. We find that the estimates for all authors are very similar in the country level, confirming that Brazil has made significant progress in improving mortality data quality in the last three decades. However, differences between the four sources of estimates of completeness and mortality exist at the state level and during all period of analysis.

Mortality estimates are the central core of the Global Burden of Disease analysis – what they called mortality envelope. In countries with a long-time series and good data quality, mortality estimates are more stable and demand less adjustments. Estimates in adult mortality remains a challenge for demographers and public health researchers in most less developed countries, and one of those challenges is to overcome the lack of quality of vital data. In countries, such as Brazil and its states, vital registration systems are recent and still face many limitations. Thus, it is not a straightforward exercise to obtain mortality levels and trends. It is necessary to perform a series of adjustments before having levels and trends of mortality in the country and states. In this paper, we revised and compared estimates from IHME to two (2) different sources of mortality estimates in Brazil.

This issue is even more important as the recent version of the Brazil Burden of Disease included estimates at the state level and there are plans to go to even smaller areas. Small-area estimates face additional problems in addition to low levels of completeness: fluctuations due to small number of events and random variation. In general, demographic methods have a series of limitations to deal with small area estimates and, in recent years, several studies tried to

incorporate statistical models to obtain more robust estimates of mortality and life expectancy (Gonzaga and Schmertmann, 2016; Schmertmann and Gonzaga, 2018; Alexander, et.al, 2017). Since we do not know the true level of mortality and there is an array of estimates in Brazil, it is important to compare and contrast these estimates.

Our analysis show that data quality has improved in recent decades but the quality of data are still deficient for a lot of states, mostly the less developed ones (Queiroz et al 2017). The possible causes for an imprecise vital record can be many, and even more, this lack of data quality can compromise even the most precise and robust estimates on mortality. Our results are very similar to what Schmertmann and Gonzaga (2018) observed when comparing their estimates to IBGE. On the other hand, we find some differences in estimates from IHME and Brazilian researchers that affect both levels and trends of completeness of adult mortality in Brazil and states, but there is a clear convergence trend in the most recent period.

Substantial differences between estimates may exist because of differences in data source, methods and/or modelling assumptions used. In the case of IHME estimates, considering that represent important effort to produce mortality estimates for all countries, regions of the country for a long period, and population estimates and envelope mortality are the basis of the model, trends and levels in causes of deaths might not be what we actually see. Then, it is very important that the data and methodology used by different researches be clear and reproducible by others. In most cases, because methodology in papers and results in websites are not very clear and methods and data are not available, it is not easy to reproduce results.

Thus, we should continuously incorporate new estimates in mortality analysis. Brazilian subnational mortality estimates represent a breakthrough for mortality studies, producing proper estimates of mortality to subsidize the understanding of population dynamics. One of next step would be analyze mortality estimates for those aged 60 and above because a good

part of the divergence in the estimates may be in the data quality at older ages. Furthermore, it is important to produce estimates with confidence interval that it would be an important gain for mortality studies.

In the case of Brazilian data, Lima and Queiroz (2011, 2014) showed how the degree of coverage of deaths in the micro-regions in the state of Minas Gerais and Brazil is characterized by a strong spatial pattern between 1980 and 2007, -number of deaths in the North and Northeast of the state. The same spatial pattern is observed in the Brazilian microregions between 1980 and 2010 (Queiroz and Lima, 2014). Although the results are concise, the authors emphasize the fragility of underreporting estimates given the limitations that traditional demographic methods present (Hill et al., 2009, Timaeus, 2001). Thus, the search for new means of estimation of the degree of coverage for the evaluation of death data in Brazil emerges as a matter of great importance in Brazilian demography. It is believed that a proposed combination of traditional demographic methods with estimates produced using Bayesian methods appears to be a good answer to this question (Bolstad, 2007).

There are other methodological alternatives to exploit the quality of mortality data and produce estimates of adult mortality and life expectancy at birth. The Preston Integrated Method (Preston, 1983) provides an alternative way to estimate a life table from two census-age distributions. This approach requires, in addition to two age distributions, a rough estimate of the probability of survival of the population from birth to age x , $ps(x)$, function (using a logit transformation of a parameter) and, if possible, an estimate independent of the probability of surviving at age 5 years, $p(5)$. The immediate question, however, is what to use as the standard survival function $ps(x)$, which should represent the true age pattern of mortality as much as possible. In a population with death record that is 80% or more complete, one possibility is to use a life table derived from recorded deaths and population; if the death register does not vary

greatly by age, the resulting survival function will not be very distorted in the age pattern. An alternative suggested by Preston is to use a model life table with a life expectancy close to what is believed to be appropriate for the population in question.

Adair and Lopez (2018) suggest an empirical model to generate estimates of sub-registration correction with applicability for small areas. The method demonstrates sufficient flexibility to predict a wide range of completeness levels at a given gross registered mortality rate. The method can be applied using data readily available at the subnational level. The model assumes that the degree of coverage of the obese has a positive relation with the crude mortality rate recorded, a negative relation with the level of mortality and a negative relation with the older age structure of the population. The model was adjusted for data from more than 100 countries and the estimated parameters can be used to obtain estimates of the degree of coverage in other countries or localities around the world. Schmertmann and Gonzaga (2018) proposed the combination of demographic methods, rate smoothing based on a relational model and use of Bayesian statistics to obtain estimates of data quality and thus corrected mortality curves for small areas. The proposed model combines a relational model for mortality curves with probabilistic prior information on death record coverage derived from demographic estimation techniques such as Death Distribution Methods and field surveys by public health experts (eg active search).

The comparison between the estimates of completeness of death counts, life expectancy and adult mortality represents a useful strategy to evaluate the potential of each method. We find substantial sub-national differences between estimates of completeness of death counts coverage, adult mortality and life expectancy between sources analyzed in this paper. The differences exist because of various differences in data and modelling assumptions used by each different agency and author. It is possible that those differences are reflecting limitations

of data and methods, but the differences should reduce as data quality improves. In addition to that, Improved transparency on methods and data used will help to improve understanding about the drivers of the differences. We expect that the methods can be widely and easily applied, with concerns about their assumptions and limitations, with the purpose of providing reliable mortality statistics for the planning of public policies.

References

- Agostinho, C S; Queiroz, B L. Estimativas da mortalidade adulta para o Brasil no período 1980/2000: uma abordagem metodológica comparativa. In: **Encontro Nacional de Estudos Populacionais**, 16, 2008, Caxambu. Anais. Belo Horizonte: ABEP, 2008. 20p.
- Agostinho, C S; Queiroz, B L. Estimativas de mortalidade adulta para as Unidades da Federação, Brasil, 2000. In: **Encontro Nacional de Estudos Populacionais**, 16, 2010, Caxambu. Anais. Belo Horizonte: ABEP, 2010. 20p.
- Albuquerque , F.R.P.C; Senna, J.R.X. Tábuas de Mortalidade por sexo e grupos de idade – Grandes Regiões e Unidades da Federação - 1980, 1991 e 2000. Textos para discussão. Diretoria de Pesquisas; Rio de Janeiro : IBGE, Coordenação de População e Indicadores Sociais, 2005.
- Alexander, M., Zagheni, E., & Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography*, 54(6), 2025-2041.
- Banister, J.; Hill, K. Mortality in China 1964-2000. **Population Studies**, 2004, v. 58, n. 1, p.55-75
- Bennett N.G. and Horiuchi S. Estimating the Completeness of Death Registration in a Closed Population. **Population Index** 1981, 47(2):207-21.
- Bhat, Mari. General Growth Balance method: a reformulation for populations open to migration. **Population Studies**, 2002, 56, 23-34
- Brass W. **Methods for Estimating Fertility and Mortality from Limited and Defective Data**. Chapel Hill, NC: Carolina Population Center, University of North Carolina, 1975.

Dorrington R.E. General Growth Balance. In Moultrie TA, RE Dorrington, AG Hill, KH Hill, IM Timæus and B Zaba (eds), **Tools for Demographic Estimation**. <http://demographicestimation.iussp.org/content/general-growth-balance>, 2011. Acessado em 25/07/2012

Dorrington R.E. Synthetic extinct generations. In Moultrie TA, RE Dorrington, AG Hill, KH Hill, IM Timæus and B Zaba (eds), **Tools for Demographic Estimation**. <http://demographicestimation.iussp.org/content/synthetic-extinct-generations>, 2011. Acessado 25/07/2012

Dorrington, R e Timæus, I. Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors, Revisited. Paper presented at **the 2008 Annual Meeting of the Population Association of America** – Nova Orleans, Estados Unidos., 2008

Dicker, Daniel, et al. "Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017." *The Lancet* 392.10159 (2018): 1684-1735.

Lima, EEC.; Queiroz, BL. (2011). A evolução do sub-registro de mortes e causas de óbitos mal definidas em Minas Gerais: diferenciais regionais. *Revista Bras. Est. Pop.*, Rio de Janeiro, v. 28, n. 2, p. 303-320, jul./dez.

Lima, EEC.; Queiroz, BL. (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*, v. 30, n. 8, p. 1721-1730.

Lima, EEC.; Queiroz, BL; Sawyer, DO. (2014). Método de estimação de grau de cobertura em pequenas áreas: uma aplicação nas microrregiões mineiras. Rio de Janeiro: *Cad. Saúde Coletiva*, v.22, n. 4, p. 409-18.

Glei, D., Barbieri, M., & Santamaría-Ulloa, C. (2019). Costa Rican mortality 1950–2013: An evaluation of data quality and trends compared with other countries. *Demographic Research*, 40, 835-864.

Gonzaga, M. R., & Schmertmann, C. P. (2016). Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. *Revista Brasileira de Estudos de População*, 33(3), 629-652.

Hill K. "Estimating census and death registration completeness." **Asian and Pacific Census Forum**, 1(3): 8-13,23-24, 1987.

Hill K. Analytical methods to evaluate the completeness and quality of death registration: Current state of knowledge. Population Division Technical Paper, no. 2017/02. New York: United Nations Population Division; 2017.

Hill, K. Metodos para estimar la mortalidad adulta en los paises en desarrollo: una revision comparativa. **Notas de Poblacion**, no. 76 pps. 81-111, 2003

Hill, K.; Choi Y.; Timaeus, I. Unconventional approaches to mortality estimation. **Demographic Research**, v.13, p.281-300, 2005

Hill, K.; Queiroz, B. Adjusting General Growth Method to Migration. **Revista Brasileira de Estudos de População**, vol. 27, n.1, 2010.

Hill, K.; You, D.; Choi, Y. Death Distribution Methods for Estimating Adult Mortality: sensitivity analysis with simulated data errors. **Demographic Research**, vol. 21 (9), pps 235-254 , 2009.

Institute for Health Metrics and Evaluation (IHME). Mortality Visualization . Seattle, WA: IHME, University of Washington, 2016. Available from <http://vizhub.healthdata.org/mortality>.

Instituto Brasileiro de Geografia e Estatística, ed. 2013. Tábuas abreviadas de mortalidade por sexo e idade: Brasil, grandes regiões e unidades da federação, 2010. Estudos e pesquisas. Informação demográfica e socioeconômica. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE.

Luy, M. A Classification of the Nature of Mortality Data Underlying the Estimates for the 2004 and 2006 United Nations' World Population Prospects. **Comparative Population Studies**, vol. 35, no. 2, 2010.

Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? Improved methods for evaluating completeness. **PLoS Medicine**. April 13; 7(4), 2010

Naghavi, Mohsen, et al. "Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016." *The Lancet* 390.10100 (2017): 1151-1210.

Paes, N. A. Avaliação da cobertura dos registros de óbitos dos Estados brasileiros em 2000. **Revista de Saúde Pública**, v.39, n.6, p.882-890, 2005.

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

Preston, S., Coale, A. J., Trussell, J. & Weinstein, M. Estimating the Completeness of Reporting of Adult Deaths in Populations That Are Approximately Stable. *Popul. Index* **46**, 179 (1980).

Preston, S., Heuveline, P. E Guillot, M.. **Demography: Measuring and Modeling Population Processes**. Oxford: Blackwell, 2000.

Preston, S.H. (1983). An integrated system for demographic estimation from two age distributions. *Demography*, vol.20, No.2, pp. 213-226.

Queiroz B, Freire F, Gonzaga M, Lima E. Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010. *Revista Brasileira de Epidemiologia* (Brazilian Journal of Epidemiology). 2017; 20 Suppl. 1: 21–33.

Queiroz, B. L. "Estimating Maternal Mortality Differentials Using Census Data: experience in Honduras". **Journal of Population Research**. Vol. 28, No. 1, 2011.

Queiroz, B. L. & Sawyer, D. O. T. What can the mortality data from the 2010 Census tell us? *Rev. Bras. Estud. Popul.* 29, 225–238 (2012).

Queiroz, B. L., Freire, F. H. M. D. A., Gonzaga, M. R., & Lima, E. E. C. D. (2017). Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010. *Revista Brasileira de Epidemiologia*, 20, 21-33.

Queiroz, B. L., Lima, E. C., Freire, F. H., & Gonzaga, M. R. (2013). Adult mortality estimates for small areas in Brazil, 1980–2010: a methodological approach. *The Lancet*, 381, S120.

Schmertmann, Carl P.; Gonzaga, Marcos R. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. **Demography**, v. 55, n. 4, p. 1363-1388, 2018.

Souza, M. F. M., Malta, D. C., França E. B., Barreto, M. L. Changes in health and disease in Brazil and its States in the 30 years since the Unified Healthcare System (SUS) was created. *Ciênc. saúde coletiva* 2018, vol.23, n.6, p.1737-1750.

Timaeus, I. M. Measurement of adult mortality in less developed countries: A comparative review. **Population Index**, v.57, n.4, p.552-568, 1991

Timaeus, Ian, M.; Chackiel, Juan; Ruzicka, Lado. **Adult mortality in Latin America**. Clarendon Press, Oxford, 1996.

Trussell, T. James. A re-estimation of the multiplying factors of the Brass technique for determining survivorship rates. *Population Studies*, v. 19, n.3, p. 97-107, 1975. UNITED NATIONS. Mortality Research: Reading in Population Research Methodology, volume 2. Mortality Research. United Nations. Population Fund.

United Nations – Population Division. **Methods for Estimating Adult Mortality**. New York, 2002.

United Nations, United Nations. **Manual X: Indirect Techniques for Demographic Estimation**. New York: United Nations, 1997.

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Carter, A., Casey, D. C., ... & Dandona, L. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific

mortality for 249 causes of death, 1980â 2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053), 1459-1544.