QUALITY OF INFORMATION IN THE QUESTION AGE IN DEATH RECORDS IN BRAZIL, 1996 - 2015.

Fatima Valeria Lima Jacques[1], Raphael Mendonça Guimarães[2]

ABSTRACT

Objective: To analyze the quality of the age declaration in death registries in Brazil, from 1996 to 2015. **Method:** A simple age analysis was performed in the microdata of deaths in Brazil between 1996 and 2015. The preference for '0' and '5' was evaluated using the Whipple index (IW). Already the preference for all 10 terminal digits was expressed using the Myers (IM) method. **Results:** The quality of the data of age was high in the period (IWtot = 0.55 - 0.83 (male) and 0.71 - 0.93 (female); IM = .388 - .004 (male) and. female)). The quality of the information was more satisfactory among men, there was no significant trend in the improvement, suggesting stability in quality in the 20 years. Preference was given to the terminal digit '0', mainly among women. **Conclusion:** Death data in Brazil, with respect to age, are satisfactory, and can be used in demographic and epidemiological analyzes.

Keywords: Data quality. Whipple Index. Myers Index. Mortality. Information systems.

INTRODUCTION

In most surveys, quantitative data are collected through questionnaires and interviews. When these instruments are self-applied, there is often a systematic underestimation or overestimation of certain clinical information and parameters, which may lead to misinterpretation[1].

Age is a widely used demographic variable for descriptive and statistical analysis of population structure and prediction of population growth, since many demographic and socioeconomic data are attributed to age and gender[2]. In addition, the relevance of information on age in epidemiology and public health is evident, since the age of the individual is considered in several situations, from the development of a risk profile to the diagnosis, management and prognosis of a disease[3].

Incorrect registration of information is a common phenomenon in developing countries and constitutes one of the greatest challenges to demography[4]. The most common irregularities related to age are a preference for certain digits and an increase in occurrence around some attractive ages[5,6]. Age data often exhibit high occurrence at round or attractive ages, such as even numbers and multiples of five[5]. Thus, evaluation of information on age is considered a measure of the quality and consistency of data[6].

When properly measured, the probability of frequency of each terminal digit (0 to 9) is about 10%. There is, however, a greater frequency, consistently, of the final digits 0 or 5[5,7]. This preference for rounding numbers may mask or exaggerate the actual differences between populations and could also explain why the differences

[1] Federal University of Rio de Janeiro, Brazil

[2] Fundação Oswaldo Cruz and Population Studies Center/University of Campinas, Brazil

between the estimates measured and reported by an individual him or herself vary across cultures. To account for this bias and explore all data detection potential, the preference for final digit reports should be evaluated for different sources of information[8].

The accuracy of age data varies from country to country and depends on a number of factors, mainly associated with data collection and tabulation problems, as well as the lack of an efficient flow of information in administrative records and population surveys[9-11]. Regarding demographic censuses and population-based surveys, such as the National Household Sample Survey (PNAD), there is a routine for evaluating the quality of the statements, as well as the magnitude and dimension of possible errors[1]. Administrative records, however, due to their nature, have a distinct logical evaluation of data quality, and this often poses a challenge to obtaining reliable statistics for mortality, which are often the only indicators available to perform analysis of the health status of the population[12]. In this sense, the objective of this study was to analyze the quality of the age declaration in the death registries in Brazil, according to sex, from 1996 to 2015.

METHODS

Study design

This is an ecological study to evaluate the quality of information.

Data source

Microdata were used referring to deaths in Brazil, by state, then compiled into a national general report for 1996 to 2015. Data were extracted from the Mortality Information System provided by the Ministry of Health.

Data analysis

In order to ensure greater precision in the analysis, differences by sex were considered. First, a visual inspection of the age pyramid by simple age was performed, checking for distortions around specific digits (such as 0 and 5, for example), for ages 40, 45, 50 and 60 years. At these ages, when there are problems in data quality, it is possible to observe inflection points in the age distribution for both sexes.

In order to evaluate the quality of death declarations in the death registries, two methods were used during the period analyzed to indicate the quality level of these data. These indicators are extensively described in Handbook II of the United Nations[13].

Digit preference in the declaration of the variable “age”, according to sex, was determined using the methods of Whipple and Myers, and evaluated statistically using an adhesion test based on Pearson’s chi-squared test. These rates are applicable when age is referred to in simple years.

a) Whipple index

The Whipple method was applied to detect preferences for terminal digits 0 and 5 in the population group from 23 to 62 years of age. This age interval was chosen as data for younger and older ages are more inaccurate (ALVES et al, 2016).

In this way, one has to:

$$IW_5 = \frac{P_{25} + P_{35} + \cdots + P_{55}}{\frac{1}{10} * (P_{23} + P_{24} + \cdots + P_{62})} x\ 100$$

The following is considered for classification:

Quality	Whipple's Index
Very Good	99,0 – 104,9
Good	105,0 – 109,9
Regular	110,0 – 124,9
Poor	125,0 – 174,9
Very Poor	175,0 and more

The calculation was carried out for the set of ages ending in 0 or 5; for this, the formula is:

$$IW_{0,5} = \frac{P_{25} + P_{30} + \cdots + P_{60}}{\frac{1}{5} * (P_{23} + P_{24} + \cdots + P_{62})} x\ 100$$

Additionally, the modified Whipple index (IW) was used, which allows global attraction analysis for all digits[14]. The modification puts in a denominator for the sum of the population by quinquennial groups in which the age with the digit to be analyzed is the midpoint of the group, and is calculated as follows:

$$IW_{m_1} = \frac{5 * (P_{31} + P_{41} + \cdots + P_{61})}{(\ {}_5P_{29} + {}_5P_{39} + \cdots + {}_5P_{59})}$$

Finally, the total modified index is presented as follows:

$$W_{tot} = \sum_{i=0}^{9} |IW_{m_i} - 1|$$

In this case, it is considered that the closer to zero, the better the quality of information.

b) Myers index

The Myers method, in turn, allows determination of the preference for each terminal digit (0 to 9) in the ages ranging from 10 to 99 years. This method assumes that the distribution per terminal digit is homogeneous (ie 10% frequency for each digit). Violation of this principle is measured by the Myers index (IM). Through this index, it is possible to verify the attraction for certain digits, which implies the quality of the information. For implementation of this method, the UN (1955) proposed using the

sum of the population that has the same final digit in the declared age for the group of 10 to 89 years (G1) and for the group of 20 to 99 years (G2). In this method, the population above 100 years is not counted, since it is assumed that it does not significantly affect the results[14]. As the population tends to be smaller with advancement of the final digits (that is, for each succeeding digit, the population is older and smaller than the previous population), these populations are multiplied by coefficients (x) in the range of 1 to 10 for the G1 group, and the complements (10 − x) for the G2 group, according to the following equations:

$$G_1(i) = (i+1) * \sum_{\alpha=10}^{89} P_i \text{, when } i \in \{0, 1, \dots 9\}$$

$$G_2(i) = (9-i) * \sum_{\alpha=20}^{99} P_i \text{, when } i \in \{0, 1, \dots 9\}$$

It is considered that the frequency (*fi*) that each digit i has in total is given by the sum of G1 and G2. It is also assumed that the expected proportion of each digit is equal to 10% of the total (assumption of uniform distribution) [13]. The IM for each final digit of the declared age is calculated as the deviation of *fi* from the expected 10%, while the general index is given by the sum of the indices for each digit, according to the formula:

$$IM_i = |100 \text{ x } f_i - 10|$$

$$IM = \sum_{i=0}^{9} IM_i \text{ , } 0 \leq IM_i \leq 180$$

For the purposes of classification:

Age Heaping	Myers´s Index
Low	Up to 4,9
Medium	from 5,0 to 14,9
High	from 15,0 to 29,9
Very high	from 30,0 to 180,0

Following acquisition of the IW and IM for each year from 1996 to 2015, a time series analysis was performed to evaluate the temporal evolution of the quality of age information in death registries. Initially, the stationarity of the time series, through the Wald–Wolfowitz test, and the trend effect, through the Cox–Stuart test, were verified. Then, after verifying these effects, the mortality trend was analyzed using a polynomial regression technique, with which the trend of the indices in the period was evaluated. The variable considered independent, in order to avoid collinearity, was the year centered by the midpoint of the period (x − 2006). Simple, second- and third-degree linear regression models were tested. The most suitable model was the one that presented the best fit of the coefficient of determination (R^2), statistical significance

(considering the 5% level) and residue analysis. In cases where the models presented similarity, those which presented the simplest model were chosen, by the principle of parsimony[15]. R software version 3.5.1 was used for the study.

RESULTS

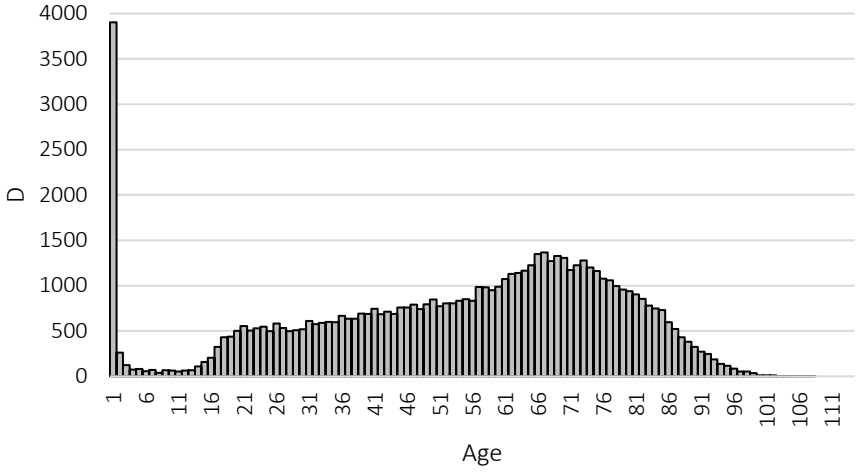
In the 20-year period, Brazil had slightly more than 21 million deaths, distributed among the different age groups. The change in the contribution of the groups was evident: the group of 0 to 9 years had a change in proportional mortality between 1996 and 2015 from 10.31% to 3.67%, and the group of 80 years and older went from 17.88% to 28.53%. Figure 1 shows the changes observed by simple age, stratified by sex. The graphs allow identification of an important distinction between the sexes. Males have a mortality pattern, especially in the young age groups, which decreases from the age of 30, returning to grow in the more advanced bands among the elderly. However, females have an increasing pattern of mortality with advancing age, without this pattern at young ages. It is speculated that this difference is favored by external causes, which are more significant in males. It is still important to highlight the change between 1996 and 2015, corroborating the previous analysis that there is, arguably, a process of compression of mortality in progress.

It is possible to say that there is a certain stability in the quality of the information about the variable age. This is reflected in the two indices evaluated (IW and IM). When evaluating digit deviation, through the IM (Figure 2), it can be seen that, from 1996, the information already presented good quality, presenting moderate variability according to the preferred digit, although, considering the unit of measurement (%), the deviation values are small. This suggests that this difference is due to the random fluctuation of the data. Greater variability is observed for the digit 0, and less variability for the digit 1. There is still a difference between the sexes, with some advantage in the quality for males.

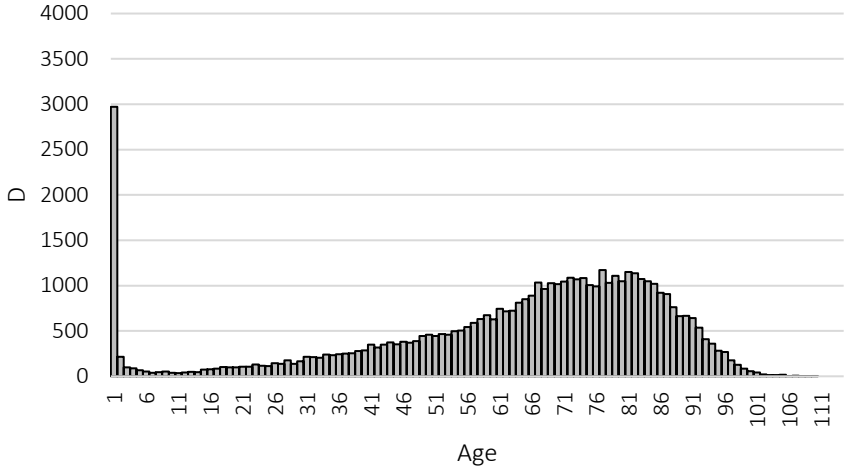
Figure 3 allows evaluation of the attraction for certain digits from the IW and its variations. In general, considering the trend of higher data stacking between digits 0 and 5, one can say that the data quality is satisfactory. This is also reflected when evaluating the total index. However, when stratifying the digits 0 and 5, it is perceived that the quality tends to reduce when analyzing the terminal digit 0, evidencing some attraction. In fact, as shown in Figure 4, which provides the specific modified total indices per digit for each of the sexes, in the years 1996 and 2015, we see some difference for the digits 0, 2 and 3, for both sexes, regardless of the year of observation.

Finally, when analyzing the trend of the two indices (Figure 5), the hypothesis of random fluctuation is corroborated, since analysis of the time series shows that there is no significant trend in the period. It should be noted, however, that this fluctuation is more evident for females, where the variability seems to be greater.

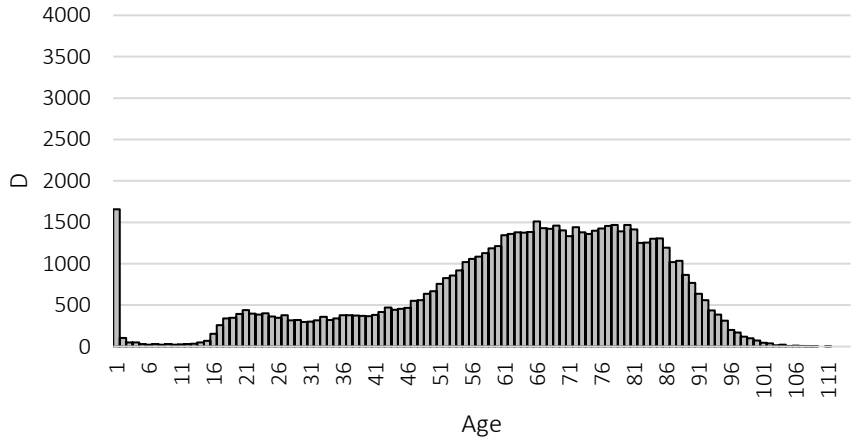
Figure 1: Age Structure of death records according to sex. Brazil, 1996 e 2015.



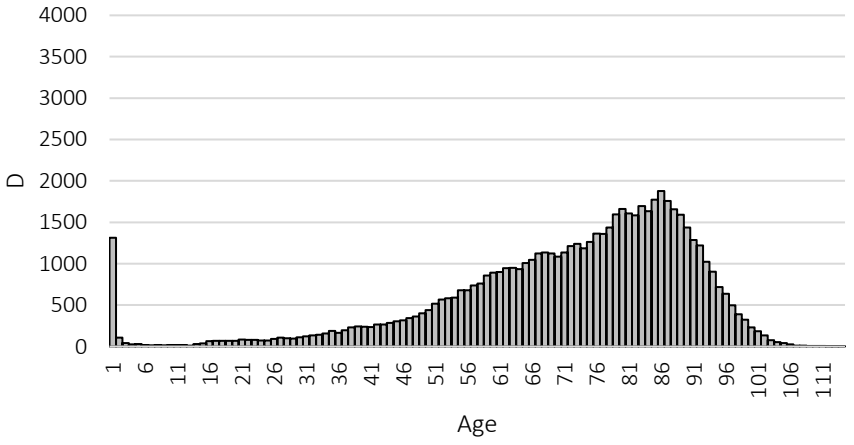
1996 - Male



1996 - Female



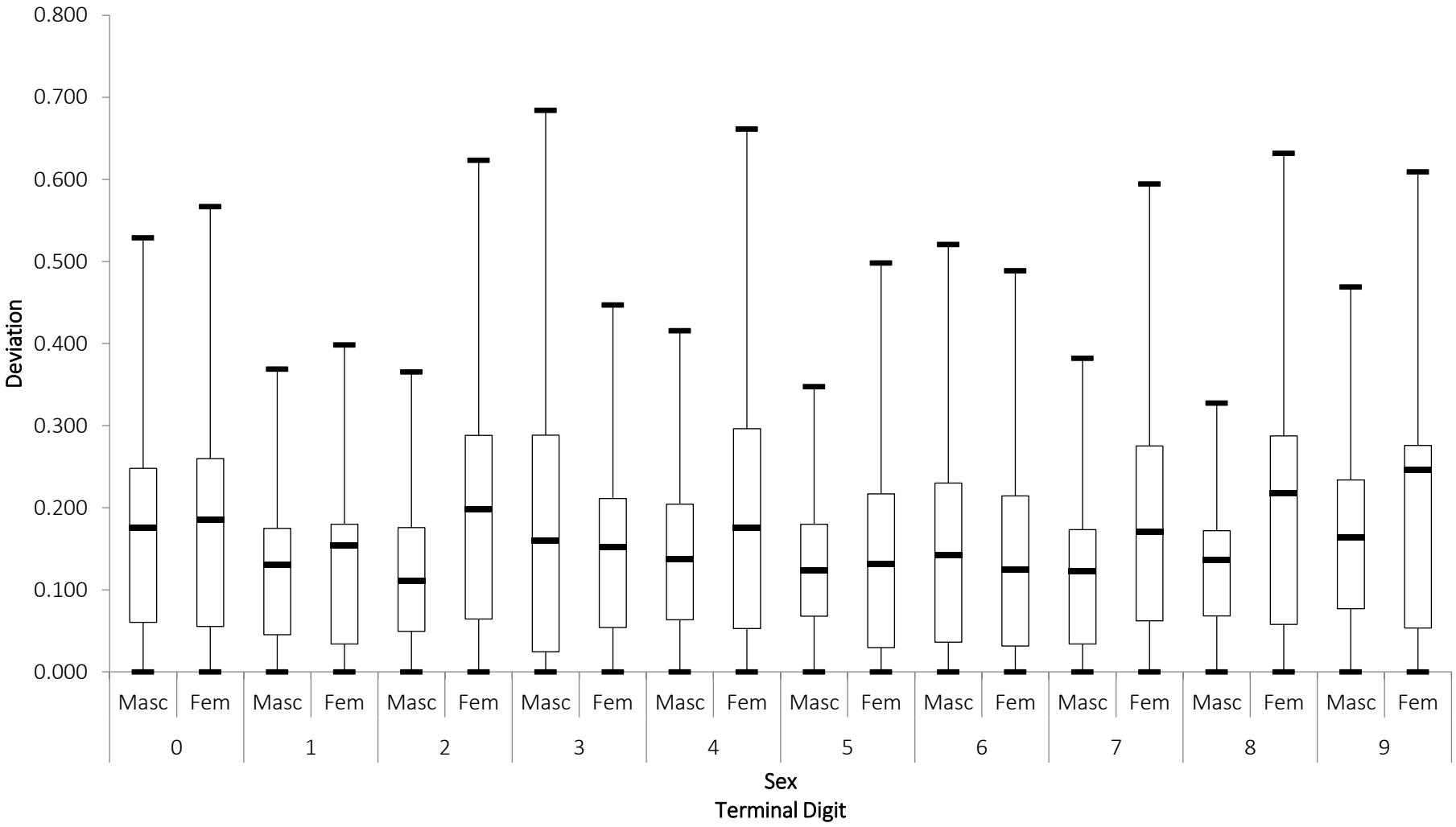
2015 - Male



2015 - Female

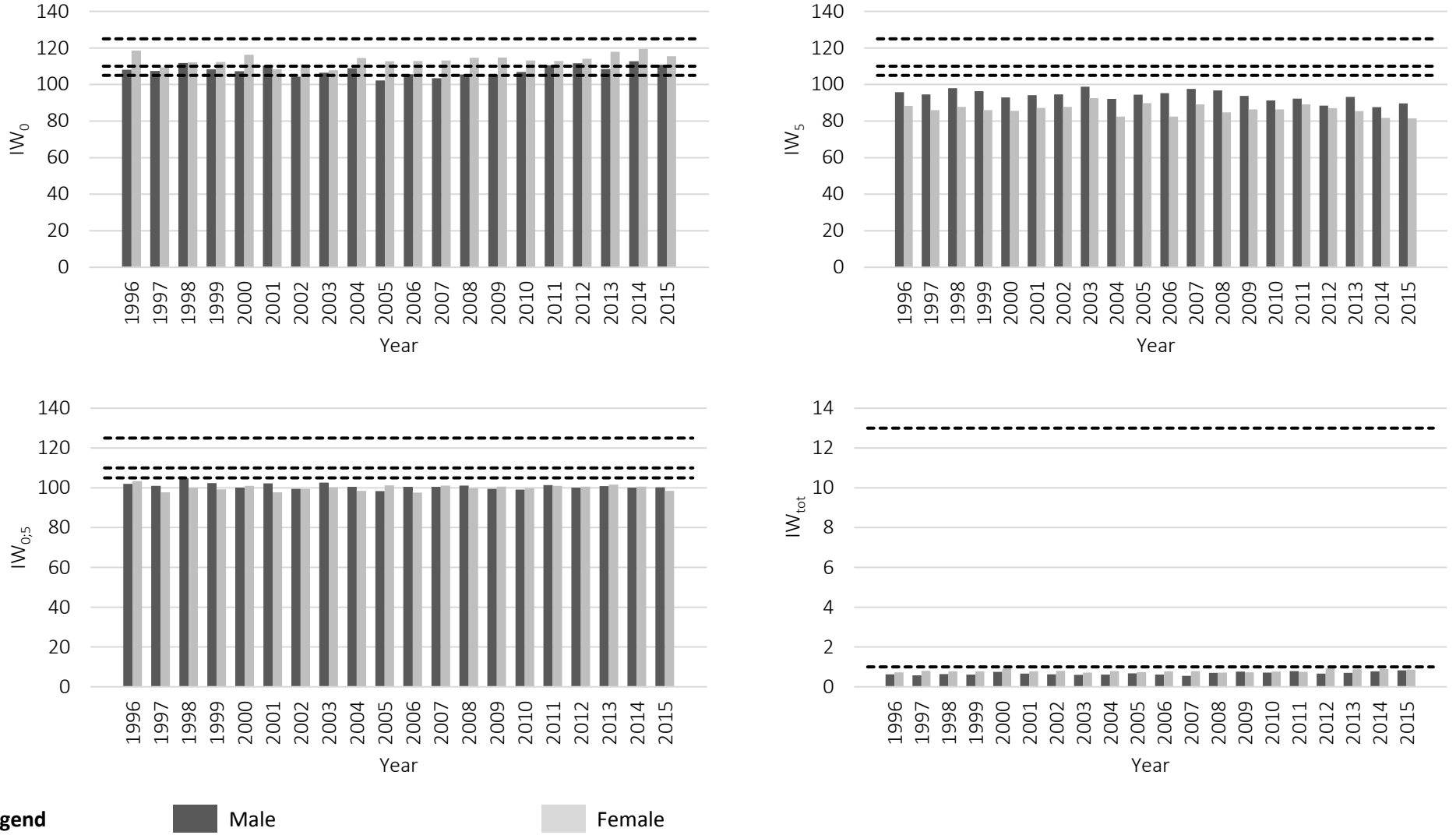
Source: DATASUS, 2018.

Figure 2: Myers's Index age heaping according to sex and digit. Brazil, 1996 – 2015.



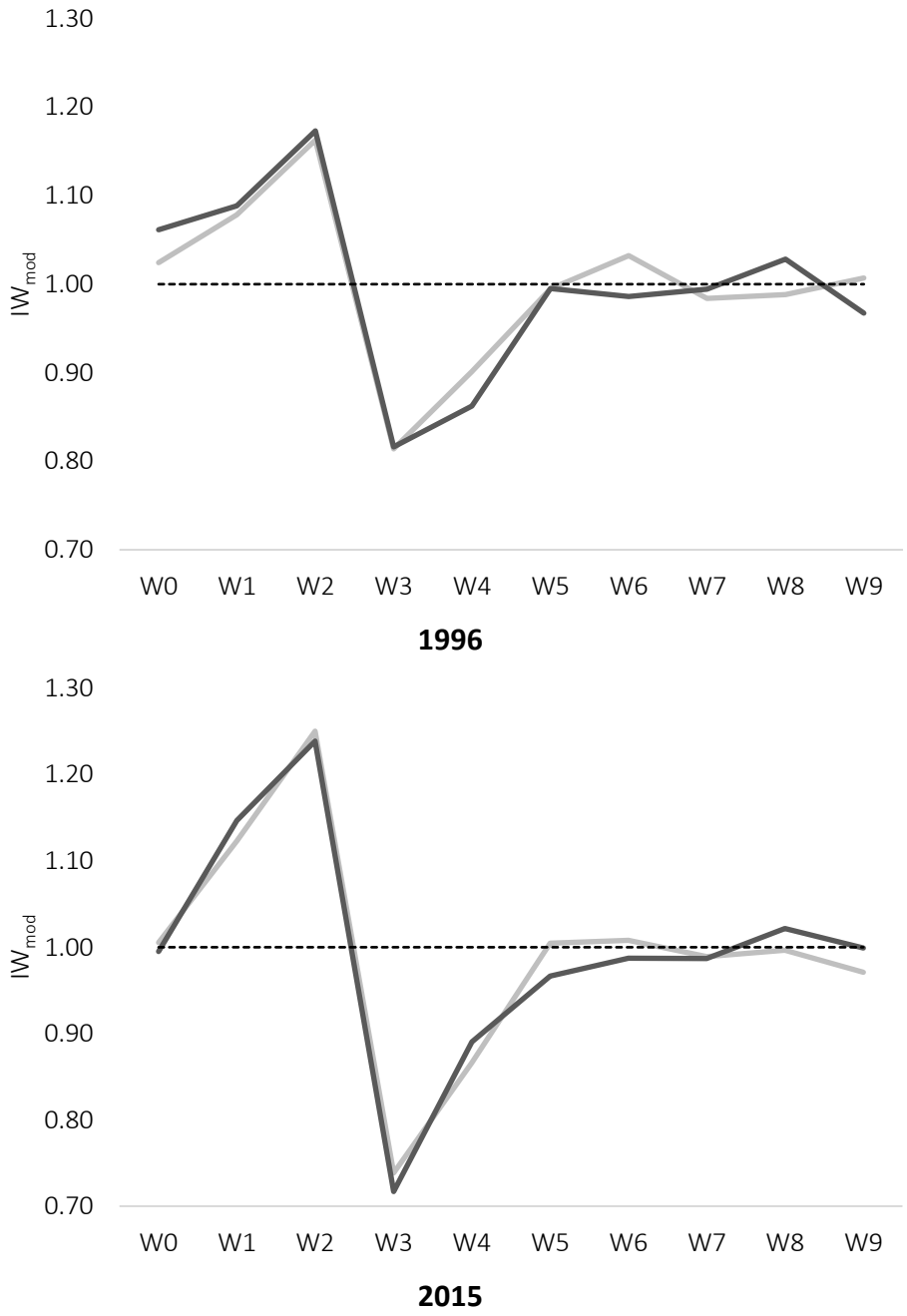
Fonte: DATASUS, 2018.

Figure 3: Whipple's Index age heaping according to sex and digit. Brazil, 1996 – 2015.



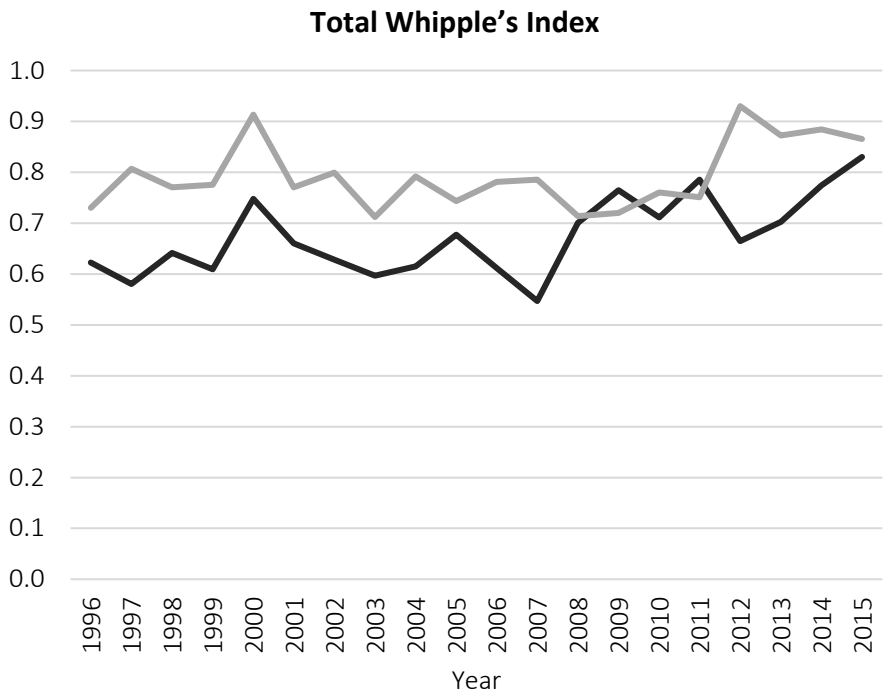
Source: DATASUS, 2018.

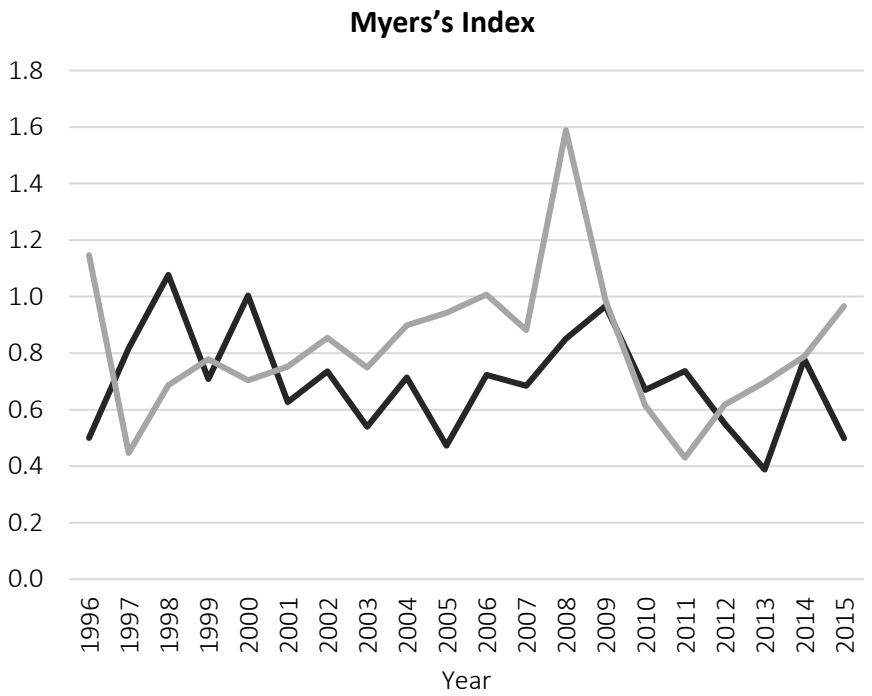
Figure 4: Whipple's Index Age heaping description according to sex and digit. Brazil, 1996 – 2015.



Source: DATASUS, 2018.

Figure 5: Time trend of age heaping indicators. Brazil, 1996 – 2015.





	Equation	R^2	p value	trend		Equation	R^2	p value	trend
Male	y= 0.007x + 0.63	0.39	0.16	non-significate	Male	y= - 0.01x + 0.70	0.09	0.22	non-significate
Female	y= 0.003x + 0.79	0.11	0.30	non-significate	Female	y= 0.01x + 0.83	0.01	0.68	non-significate

Legend Male Female

Source: DATASUS, 2018.

DISCUSSION

The present study used two different techniques to evaluate the quality of the variable age recorded on death certificates. It is worth remembering that inadequacies in the age register are frequent in certain data sources, being more common in developing countries, making it impossible to use the data instantaneously without using any correction[16].

The methodology of data collection used in this study was similar to the method used in almost all studies and domiciliary surveys. Considering the two indices studied, it is inferred that the age data collected in the death registers can be considered of good quality. There was a slight attraction for the terminal digits 0 and 5, indicating a preference of the respondents for the story of ages. In this way, the age distribution suggests an acceptable level of reliability, without the need for adjustments or smoothing. Therefore, such mortality data allow the construction of life tables, an important instrument for assessing mortality and life expectancy, for demography. A previous study[17] has already suggested this level of quality for Brazil, corroborating the present longitudinal analysis, which continues to present such a pattern.

The values obtained for the indices IW and IM indicate a discrete preference for zero digits in the SIM database. However, it is not large enough to call into question the quality of records. One of the possible explanations for the good quality of the information concerning age is the fact that fulfillment of the declaration of death, as well as the declaration of live births, is carried out by means of presentation of an identification document, or made directly from the medical record of the patient[18,19], rather than by verbal statements from family members, a common practice in household surveys[20,21].

Randall and Coast[22] proposed modifying the WI (fully modified WI) to evaluate the quality of data in the elderly by analyzing household surveys from African countries. Their conclusion was that the quality of information on age is poor for most countries in south Africa. The authors also concluded that in the surveys of some countries, there is a considerable omission of the proportion of older women. For Latin America, however, Romero and Freitez[23] showed an improvement in the quality of age declaration, when comparing the census of the 1990s with the census of the 2000s. Andrade et a[24], also analyzing Latin America, found that in a context of rejuvenated age structure, the use of WI or MI is independent. However, in countries with an older age structure, there is a greater weight of data quality for the elderly, and the choice of indicator should be carefully considered.

Studies using IW and IM have shown that digital preference occurs more frequently for females[25,26], and that there is preference for the digits 0 and 5[27,28].

The record quality of the variable age is important not only because age distribution is an invariable part of a research report, but also because the introduction of age-related biases in studies can lead to wrong inferences. Although there is a careful routine to obtain the data, some problems may arise; for example, there are so-called

"systematic errors" [29]. It is known that the approximation of age consciousness is manifested in the phenomenon of age accumulation, for self-referenced age or proxy data[19,30]. When the information is provided by proxy, this problem becomes even more evident. In turn, the impact of such an incorrect statement may lead to misclassification and incorrect assessment of demographic rates, and thus interfere with the planning of effective interventions[29].

The accuracy of age data in health records can ultimately be assessed by means of demographic quality control indices. In this way, innovative data collection methods, as well as statistical techniques for error minimization, should be used to guarantee the accuracy of old data.

FINAL CONSIDERATIONS

Evaluation of the quality of mortality data is important, because for several health situations involving a series of injuries, death data are the only information available. The results point to the good quality of the data, although there is a difference between the sexes. This indicates the need to assess specific situations for data quality, such as the cause of death (assuming there are competing risks and different patterns of mortality for each cause, for example), since the use of demographic tools seems appropriate for health information systems. Additionally, the need for application on bases other than just mortality is pointed out, such as information on births, outpatient procedures and hospital admissions.

REFERENCES

1. HAKKERT, R. Fontes de Dados demográficos. - Belo Horizonte, Textos Didáticos – ABEP 1996. Disponível em http://www.abep.nepo.unicamp.br/docs/outraspub/textosdidaticos/tdv03.pdf Acesso em abril de 2018.
2. JANNUZZI, P.M. Indicadores sociais no Brasil: conceitos, fontes de dados e aplicações para formulação e avaliação de políticas públicas, elaboração de estudos socioeconômicos. Rio de Janeiro: Alínea Editora, 2 ed, 2014.
3. CALAZANS, A.T.S. Qualidade da informação: conceitos e aplicações. Transinformação, vol. 20, n. 1, p.29-45, 2008.
4. DEL POPOLO, F. Los problemas en la declaración de la edad de la población adulta mayor en los censos. CEPAL - SERIE 8. Población y desarrollo. – Chile: CEPAL, 2000.
5. LUY, M. Estimating Mortality Differences in Developed Countries From Survey Information on Maternal and Paternal Orphanhood. Demography, vol.49, n.12, p. 607-27, 2012.
6. BELLO, Y. Error Detection in Outpatients' Age Data Using Demographic Techniques. Int. J. Pure Appl. Sci. Technol., vol.10, n.1, p. 27-36, 2012.

7. BURCH, T.K. Error in Demographic and Other Quantitative Data and Analyses. Document de travail: Vol. 3: Iss. 3, Article 1, 2015. Disponível em: http://ir.lib.uwo.ca/pclc/vol3/iss3/1. Acesso em maio de 2018.
8. LYONS-AMOS, M., STONES, T. Trends in Demographic and Health Survey data quality: an analysis of age heaping over time in 34 countries in Sub Saharan Africa between 1987 and 2015. BMC Res Notes, vol.10, p. 760, 2017.
9. AGRAWAL, G., KHANDUJA, P. Influence of Literacy on India's Tendency for Age Misreporting: Evidence from Census 2011. Journal of Population and Social Studies, vol. 23, n.1, p.47-66, 2015.
10. SINGH, M. Understanding digit preferences in India using modified whipple index: an analysis of 640 districts of India. International Journal of Current Research., vol.9, n.1, p. 45144-52, 2017.
11. ICF International. Demographic and Health Surveys Methodology - Questionnaires: Household, Woman's, and Man's. MEASURE DHS Phase III. Calverton, Maryland, USA: ICF International. Available at http://dhsprogram.com/pubs/pdf/DHSQ6/DHS6_Questionnaires_5Nov2012_DHSQ6.pdf. Acesso em maio de 2018.
12. LIMA, E.E.C, QUEIRÓZ, B.L. 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, vol.30, n.8, p.1721-1730, 2014.
13. ALVES ,L.A., ANDRADE, P.G., DE MARIA, P.F., PEREIRA, A.C.R., MARINS, R.L., BRUSSE, G.P.L., CAMARGO, K.C.M. Uma proposta de utilização do software R para a construção de algoritmos de avaliação da qualidade da declaração da idade. Textos NEPO, n. 73, - Campinas, SP: Núcleo de Estudos de População "Elza Berquó", Unicamp, 2016.
14. ANDRADE, P.G., PEREIRA, A.C.R., CAMARGO, K.C.M., BRUSSE, G.P.L, GUIMARÃES RM. Calidad de la declaración de la edad de las personas mayores en países de América Latina y el Caribe: análisis de los censos demográficos de las décadas de 1960 a 2010. Notas de Población, vol.44, n.105, p.53-84, 2017.
15. LATORRE, M.R.D.O., CARDOSO, M.R.A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. Revista Brasileira de Epidemiologia, vol.4, n.3, p.145-152, 2001.
16. DENIC, S., KHATIB, F., SAADI, H. Quality of age data in patients from developing countries. J Public Health, vol.26, n.2, p. 168–71, 2004.
17. PAES, N.A., ALBUQUERQUE, M.E. Evaluation of population data quality and coverage of registration of deaths for the Brazilian regions. Rev Saude Publica, vol.33, n.1, p. 33–43, 1999.
18. KANSO, S., ROMERO, D.E., LEITE, I.C., MORAES, E.N. Diferenciais geográficos, socioeconômicos e demográficos da qualidade da informação da causa básica de morte dos idosos no Brasil. Cad Saúde Pública, vol.27, n.7, 1323–39, 2011.

19. ROMERO, D.E., CUNHA, C.B. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002. Cad Saude Publica. Vol.23, n.3, p.701–14, 2007.
20. BORKOTOKY, K., UNISA, S. Indicators to examine quality of large scale survey data: An example through District Level Household and Facility Survey. PLoS One, vol.9, n.3, p. 1-11, 2014.
21. PIMIENTA, R., BOLAÑOS, M. La declaración de la edad: un análisis comparativo de su calidad en los censos generales de población y vivienda. Documentos de Investigación, N° 33, México: EL Colegio Mexiquense, 1999.
22. RANDALL, S., COAST, E. The quality of demographic data on older Africans. Demographic Research, vol.34, n.1, p.143–174, 2016.
23. ROMERO, D., FREITEZ, A. Problemas de calidad de la declaracíon de edad de la poplacíon adulta mayor en los censos de America Latina de la ronda del 2000. Córdoba: III Congreso de la Asociación Latinoamericana de Población (ALAP), 2008.
24. ANDRADE, P.G., BRUSSE, G.P.L, CAMARGO KCM, PEREIRA ACR, DE MARIA PF. Evolução da qualidade da declaração da idade na América Latina e Caribe: uma proposta de escolha de métodos a partir da estrutura etária. Foz do Iguaçu: VII Congresso da Associação Latino-Americana de População (ALAP) e o XX Encontro Nacional de Estudos Populacionais (ABEP), 2016.
25. BAILEY, M., MAKANNAH, T.J. Patterns of digit preference and avoidance in the age statistics of some recent African censuses: 1970-1986. J Off Stat., vol.9, n.3, p. 705–15, 1993.
26. YAZDANPARAST, A., POURHOSEINGHOLI, M.A., ABADI, A. Digit preference in Iranian age data. Ital J Public Health. 2012;9(1):64–70.
27. BARUA, R.K. Detection of Digit Preference and Age Misreporting by using Demographic Techniques. East West University, Dhaka; 2015.
28. DAHIRU, T., DIKKO, H.G. Digit preference in Nigerian censuses data of 1991 and 2006. Epidemiol Biostat Public Heal. 2013;10(2):6–10.
29. GROVES, R.M., FOWLER, F.J., COUPER, M.P., LEPKOWSKI, J.M., SINGER, E., TOURANGEAU, R. Survey Methodology (2nd ed.). New Jersey: John Wiley and Sons, 2009.
30. BRESTOFF, J.R., VAN DEN BROECK, J. Reporting data quality. In: Van den BROEK, J, BRESTOFF, J.R. (org). Epidemiology: Principles and Practical Guidelines. Dordrecht: Springer; 2013. p. 557–70.