SIMPSON'S PARADOX: A DEMOGRAPHIC CASE STUDY ABOUT POVERTY, INEQUALITY AND POPULATION DYNAMICS

Raphael Mendonça Guimarães[1,2], Natália Martins Arruda[3], Flávia Cristina Drumond Andrade[2]

ABSTRACT

Background: Brazil is experiencing a demographic transition, marked by regional inequalities. It is possible to assume that aspects related to poverty, development and inequality may modify the pattern of demographic transition, typifying a phenomenon known as Simpson's Paradox. **Objective:** To analyze the effect of inequality, poverty and social development on the Brazilian population dynamics, checking the occurrence of Simpson's paradox in the demographic transition. **Methods:** We evaluated the correlation between demographic indicators, stratifying Brazilian federation units into groups according to social indicators. **Results:** The transition has been occurring in all FUs, with persistence of the distance between them, even with reduction over the years. Simpson's paradox was observed when the analysis was performed according to the census year, and social indicator, especially for the year 1991. **Conclusion:** The challenge is to understand how the Brazilian demographic dynamics can be analyze, and to understand how these contextual factors change their pattern.

Keywords: Demography, Demographic Transition, Poverty, Iniquity, Development, Brazil.

STATEMENT OF RESEARCH QUESTION

Demographic Transition is a model that describes population change over time[1]. In fact, most regions and countries have experienced unprecedented demographic changes over the past 100 years. However, contemporary societies are now at very different stages of their demographic transitions[2]. With the largest population in Latin America, and the largest territorial extension, Brazil is going through this process. Today, most of the population lives in urban areas, and has been experiencing a faster decline in fertility since the 1970s, a movement after the drop-in mortality that began in the 1930s[3]. The country has a marked regional inequality. Although Brazil has reduced inequalities from the bottom, i.e. removing part of the Brazilian population below the poverty line, the large concentration of income at the top remained stable[4]. It is possible to assume that aspects related to poverty, development and social inequality may represent an effect modifier of these characteristics, typifying a phenomenon known as Simpson's Paradox. Also known as the Yule-Simpson effect, this phenomenon refers to an association or effect found in various subgroups, but is reversed when data from these groups are aggregated[5,6]. Therefore, the objective of the present study is to analyze the effect of poverty and inequality on the Brazilian population dynamics, verifying the occurrence of Simpson's paradox in the demographic transition.

DATA AND RESEARH METHODS

Background

It is an ecological study, whose units of analysis are the federation units (FUs), in a total of 27. The FUs are subnational entities with a certain degree of autonomy (self-government, self-regulation and self-collection) and endowed with their own government and constitution[7]. They are made up of 26 states and one federal district, located in 5 major regions. Like what has happened in many countries on all continents, the constraints of regionalization processes that account for territorial transformations in Brazil follow demographic, economic, political and social criteria. It is important to emphasize, however, that this process is equally marked by the increase of inequalities between social classes and territories. Thus, despite the Brazilian progress in the social and economic areas, these federative units maintain

[1] Oswaldo Cruz Foundation, Ministry of Health, Oswaldo Cruz, Brazil.
[2] University of Illinois at Urbana-Champaign, United States.
[3] Federal Institute od Education, Science and Technology, Campus Campinas, Brazil.
Correspondent Author: Raphael M. Guimarães. Mail to: raphael.guimaraes@fiocruz.br

considerable heterogeneity with respect to indicators of poverty, inequality, income and social development[8].

Data source

We used population data from the Brazilian Demographic Census of 1991, 2000 and 2010, collected by the Brazilian Institute of Geography and Statistics Foundation (IBGE), according to age in Brazil, disaggregated by FU. Additionally, the following social indicators were extracted from the Brazilian Human Development Atlas[9]:

a) Human Development Index (HDI): This is a summary measure of long-term progress in three basic dimensions of human development: income, education and health.

b) Gini Index: is a measure of inequality that evaluates the concentration of income in a population.

c) Proportion of extremely poor: Proportion of individuals with per capita household income equal to or less than ¼ of the minimum wage.

d) Average income per capita: average income of each resident of a given FU. This is the sum of the income of all residents divided by the number of inhabitants.

Population data also allowed the calculation of the following demographic indicators:

i) Crude Birth Rate (TBN): number of live births per thousand inhabitants;

ii) Crude Mortality Rate (TBM): total number of deaths per thousand inhabitants.

Data analysis

Brazil has incomplete coverage of births and deaths in vital records[10]. For this reason, birth and mortality indicators are often obtained through census data and other standardized retrospective surveys, such as the Demographic Health Survey. Thus, it is necessary to adjust the data using indirect demographic methods. In this research, we used the census data from 1991, 2000 and 2010, and corrected the undercounting of births in the birth registry in Brazil using the adjustment factors provided by the Brazilian Institute of Geography and Statistics, through established data adjustment methods. as the P/F relational method by Brass and Gompertz[11]. To correct mortality data, we used: The Adjusted Synthetic Extinct Generations (SEG-adj) method proposed by Hill et al[12] for adult mortality; and for infant mortality we used the method of Brass and Coale[13] with the variant proposed by Trusell[14].

Social indicators were ranked so that it was possible to assess the heterogeneity of Brazilian units regarding inequality, poverty and social development. Also, we compared the indicators by FU in the census years and then we estimated their variation over the 20 years of the time series studied. To check for the time-effect on the relationship between birth and mortality (which is indicative of demographic transition), we calculated the correlation between these two rates for each census year and compared the direction of the correlation and its magnitude. The correlation between demographic indicators (birth and mortality) and economic indicators (Gini index, average per capita income, proportion of extremely poor and HDI) is estimated for each year, and we applied the Brazilian reference values in 1991. To observe the change in the FU with their data from 1991, 2000 and 2010. In order to smooth out the correlation, when it was not linear, we used locally based running line smoother (*loess*) analyzes to evaluate the relationships between the variables of interest.

From this initial diagnosis, the FUs were divided into two groups, based on the median social indicators, and the demographic indicators were then evaluated based on the stratification of the two groups. Since conceptually the demographic transition is marked by a change in birth and mortality patterns, we summarize the relationship between gross birth rates and mortality. We provided correlation measurements using Spearmann coefficient, and we also considered statistical significance of the calculated measures ($p<0.05$).

FINDINGS

There is a general advance of the Brazilian states in relation to social indicators. The evolution of the HDI shows global growth of human development in the federative units, reducing inequality. Importantly, there is a certain formation of similar HDI spatial clusters. We consider the HDI ranges

proposed by the United Nations[9] (very low up to 0.444; low between 0.500 and 0.599; average between 0.600 and 0.699; high between 0.700 and 0.799; and very high above 0.800). The very low HDI group aggregates almost all Northeast states. The low HDI group, on the other hand, includes the northern and northeastern states not included in the previous group. The average HDI group brings together the Midwestern states and the two Southeastern states with the largest variation in social indicators, either by their own state profile, or by aggregating many municipalities and with many structural and demographic differences. Finally, the group with high HDI concentrates the states of the South and Southeast, showing the disparity between South-Southeast axis and North-Northeast of Brazil. We perceived similar relationship to that observed for the HDI also for the indicators of average per capita income and proportion of extremely poor population. When considering the proportion of the extremely poor population, there was a noticeable reduction in this proportion between 1991 and 2010, and in the opposite direction there was a general increase in average per capita income. Still, inequality between North-Northeast and South-Southeast seems to remain despite the overall improvement.

By analyzing the inequality indicator, we observed that this evolution is still accompanied by some disparity. Gini index has reduced in Brazil in most of the FUs. However, UF with greater inequalities have distanced themselves from those where inequality seems to have decreased or stabilized. Importantly, changes in income indicators do not necessarily follow trends of inequality indicators, neither at the level or pattern. This means that some states have greater wealth, and this does not necessarily mean that they do not have inequalities.

There is a correlation effect between economic and demographic indicators. Overall, this correlation is persistent over time, although it is not necessarily linear across all years of observation. This evidence is most important for average per capita income and the proportion of the population in extreme poverty, possibly because these two indicators do not have normal distribution.

It is important to note that there is an association between mortality and birth. By using the Brazilian rates of 1991 as high and low definition parameters to compare the behavior of these rates in the FUs, we created the image of four quadrants that, clockwise, describe four different scenarios: high birth rate and high mortality rate; low mortality rate and high birth rate; low birth rate and low mortality rate; and low birth rate and high mortality rate. These scenarios describe, by approximation, the phases of the demographic transition. We have observed that FUs has been experiencing the transition over the years. However, inequalities and the distance between these two groups persist over time. There is a spatial grouping of the FUs to the stage at which they are in the demographic transition. The North and Northeast units, compared to the South and Southeast units, are systematically at earlier times of transition, regardless of the year of observation. When observed the correlation between mortality and birth, it is noted that, considering the FUs of the different years of analysis as different places, there is a positive correlation between these two components. However, when the analysis is stratified by census year, the correlation becomes negative for each year. This change, it is believed, is related to a change in the course of the demographic transition in the country. Thus, we note that the year of analysis is an effect modifier, and it highlights Simpson's paradox for this phenomenon.

It is important to recognize a certain pattern in the main indicators that describe the phases of the demographic transition in Brazil. Overall, Brazil changed the relationship between birth and mortality between 1991 and 2010. In 1991 and 2000 the relationship was straightforward, characterizing an earlier stage of the demographic transition. It is worth mentioning that the correlation was stronger in 1991 than in 2000, although in the two years it was not significant. In 2010, the relationship is inverted, becoming inverse (although it has no statistical significance).

Finally, when we analyzed these correlations stratified by the indicators of inequality, development, poverty and income, an atypical phenomenon is observed, confirming the existence of Simpson's Paradox. Some of these correlations, in the dataset, show no statistical significance, and or in a certain direction. When stratified, they behave in the opposite way to the general dataset and become significant in some cases. This situation is particularly noticeable in 1991, where the sense of correlation change in the groups, and it becomes statistically significant. In the year 2000, the change in the direction of correlation remains, but without statistical significance, suggesting that the transition is underway, reducing the distance

between the FUs with respect to the level of birth and mortality. In 2010, subgroup correlations have the same meaning as total correlation, and the only indicator that reflects significant correlation is the income indicator. This phenomenon suggests that income, poverty, development and inequality may behave as modifiers of effect on the relationship between demographic components, with some difference in the force of interaction.

REFERENCES

1. Kalemli-Ozcan S. A stochastic model of mortality, fertility, and human capital investment. Journal of Development Economics. 2003; 70 (1): 103–118.
2. Bongaarts J. Human population growth and the demographic transition. Philos Trans R Soc Lond B Biol Sci. 2009; 364(1532): 2985–2990.
3. Anderson JSN, Schneider S. Brazilian Demographic Transition and the Strategic Role of Youth. *Espace populations sociétés. 2014;* DOI 10.4000/eps.5774.
4. Georges R, Maia K. A distância que nos une: um retrato das desigualdades brasileiras. Oxfam Brasil, 2017. Disponível em https://www.oxfam.org.br/sites/default/files/arquivos/Relatorio_A_distancia_que_nos_une.pdf Acesso em outubro de 2018.
5. Marang-nav de Mheen PJ, Shojania KG. Simpson's paradox: how performance measurement can fail even with perfect risk adjustment. BMJ Qual Saf. 2014; 23: 701–705.
6. Hernán MA, Clayton D, Keiding N.The Simpson's paradox unraveled. International Journal of Epidemiology. 2011; 40:780–785.
7. Lijphart A. Patterns of democracy: government forms and performance in thirty-six countries. New Haven: Yale University Press, 1999.
8. IPEA (Instituto de Pesquisa Econômica Aplicada). Desenvolvimento Regional no Brasil: políticas, estratégias e perspectivas. Rio de Janeiro; 2017.
9. United Nations Development Programme. Human Development Indices and Indicators 2018 Statistical Update. New York. http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf. Acessado em 20 Setembro 2019.
10. Lima EEC, Queiroz BL, Zeman K. Completeness of birth registration in Brazil: an overview of methods and data sources. Genus (online): Journal of Population Studies. 2018; 74: 1-20.
11. Lima EEC, Queiroz BL, Zeman K. Completeness of birth registration in Brazil: an overview of methods and data sources. Genus (online): Journal of Population Studies. 2018; 74: 1-20.
12. Hill K, You DZ, Choi YJ. Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors. Demogr Res. 2009; 21: 235-54.
13. Brass W, Coale AJ. Methods of analysis and estimation," in Brass, W, AJ Coale, P Demeny, DF Heisel, et al. (eds). The Demography of Tropical Africa. Princeton NJ: Princeton University Press, 1968, pp. 88-139.
14. Coale AJ, Trusell TJ. Estimating the time to which Brass estimates apply. In: Preston SH, Palloni A. Five tuning Brass type mortality estimates with data on ages of surviving children. Popul Bull United Nations. 1978; 10:49-54.