FEDERAL UNIVERSITY OF MINAS GERAIS (UFMG)
FACULTY OF ECONOMIC SCIENCES (FACE)
CENTER FOR DEVELOPMENT AND REGIONAL PLANNING (CEDEPLAR)
GRADUATE PROGRAM IN DEMOGRAPHY

Hisrael Passarelli Araujo

SOCIODEMOGRAPHIC PROFILES AND MORTALITY PATTERNS OF FOREIGN-BORN ADULTS LIVING IN SÃO PAULO, BRAZIL

Belo Horizonte
2022

Hisrael Passarelli Araujo

SOCIODEMOGRAPHIC PROFILES AND MORTALITY PATTERNS OF FOREIGN-BORN ADULTS LIVING IN SÃO PAULO, BRAZIL

A dissertation submitted in partial fulfillment for the master's degree in Demography at the Center for Development and Regional Planning of the Federal University of Minas Gerais

Advisor: Prof. Gilvan Ramalho Guedes

Co-advisor: Prof. Cássio Maldonado Turra

Belo Horizonte
2022

Ficha Catalográfica

A663s 2022

Passarelli-Araujo, Hisrael.
Sociodemographic profiles and mortality patterns of foreign-born adults living in São Paulo, Brazil [manuscrito] / Hisrael Passarelli Araujo. – 2022.
66 f.: il.

Orientador: Gilvan Ramalho Guedes
Coorientador: Cássio Maldonado Turra
Dissertação (mestrado) - Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional.
Inclui bibliografia (f. 52-57) e anexos

1. Imigrantes - São Paulo – Teses. 2. Brasil - Migração - Teses. 3. - Imigrantes – Mortalidade - Teses. I. Guedes, Gilvan Ramalho. II. Turra, Cássio Maldonado. III. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. IV. Título.

CDD: 304.8098161

Elaborada pela Biblioteca da FACE/UFMG – 055/2021
Fabiana Pereira dos Santos CRB6 2530

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL
PROGRAMA DE PÓS-GRADUAÇÃO EM DEMOGRAFIA

ATA DE DEFESA DE DISSERTAÇÃO DE HISRAEL PASSARELLI ARAUJO Nº DE REGISTRO 2020657206

Às nove horas do dia vinte e dois do mês de fevereiro de dois mil e vinte e dois, reuniu-se por videoconferência a Comissão Examinadora de DISSERTAÇÃO, indicada "ad referendum" pelo Colegiado do Curso em 14/02/2022, para julgar, em exame final, o trabalho final intitulado ***"Sociodemographic Profiles and Mortality Patterns of Foreign-born Adults Living in São Paulo, Brazil"***, requisito final para a obtenção do Grau de Mestre em Demografia.

Abrindo a sessão, o Presidente da Comissão, Prof. Gilvan Ramalho Guedes, após dar a conhecer aos(às) presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos(as) examinadores(as), com a respectiva defesa do candidato. Logo após, a Comissão composta pelos(as) professores(as): Gilvan Ramalho Guedes (Orientador) (CEDEPLAR/FACE/UFMG), Cássio Maldonado Turra (Coorientador) (CEDEPLAR/FACE/UFMG), Bernardo Lanza Queiroz (CEDEPLAR/FACE/UFMG) e Rosana Aparecida Baeninger (NEPO/UNICAMP) se reuniu, sem a presença do candidato e do público, para julgamento e expedição do resultado final.

A Comissão **APROVOU** o candidato por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 22 de fevereiro de 2022.

Prof. Gilvan Ramalho Guedes (Orientador) (CEDEPLAR/FACE/UFMG)
Prof. Cássio Maldonado Turra (Coorientador) (CEDEPLAR/FACE/UFMG)
Prof. Bernardo Lanza Queiroz (CEDEPLAR/FACE/UFMG)
Profª. Rosana Aparecida Baeninger (NEPO/UNICAMP)

LAURA LÍDIA RODRIGUEZ WONG
Coordenadora do Programa de Pós-Graduação em Demografia

Documento assinado eletronicamente por **Gilvan Ramalho Guedes**, **Subcoordenador(a)**, em 22/02/2022, às 12:45, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Bernardo Lanza Queiroz**, **Professor do Magistério Superior**, em 22/02/2022, às 14:09, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Cassio Maldonado Turra**, **Professor do Magistério Superior**, em 23/02/2022, às 11:19, conforme horário oficial de Brasília, com fundamento no art. 5º

do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rosana Baeninger**, **Usuário Externo**, em 23/02/2022, às 12:48, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Laura Lidia Rodriguez Wong**, **Professora do Magistério Superior**, em 03/03/2022, às 11:49, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1253445** e o código CRC **71DA38E4**.

Referência: Processo nº 23072.208757/2022-15 SEI nº 1253445

ABSTRACT

As most countries are now at advanced stages of the demographic transition, international migration has regained importance for explaining population dynamics. However, the lack of good data limits our knowledge about the characteristics of the foreign-born population living in many countries. In this context, socioeconomic and health factors associated with migration are critical research topics. This dissertation first examines the diversity of sociodemographic profiles among foreign-born adults living in São Paulo, Brazil. It also investigates whether and how mortality patterns differ between foreign-born and Brazilian-born adults living there. We rely on data from the 2010 Brazilian demographic census and apply the Grade of Membership (GoM) method to define the sociodemographic profiles of foreign-born adults. We found four profiles of foreign-born adults: recent, middle-term, long-term, and old migrants. Profiles differences are marked mainly by age, length of residence, education, and country of birth. Old migrants (Europeans and Japanese) gave place to newer migrants from South America. Next, we examine mortality differences between foreign-born and native-born adults living in São Paulo and evaluate whether these differences vary by age and education. Estimates are based on negative binomial regression models that combine death registers from the Brazilian Mortality Information System from 2009 to 2011 with population figures from the 2010 Brazilian demographic census. We found evidence of mortality disadvantages among foreign-born adults compared to Brazilians living in São Paulo, even after controlling for age, sex, and education. However, the mortality differences by nativity vary by age and education. Migrant mortality disadvantage is highly concentrated among younger adults with low education (individuals who are more likely to come from other South American countries). Our findings contrast with the mortality paradox among migrants in developed countries, especially in the United States.

Keywords: International Migration; Mortality differentials; São Paulo; Brazil.

RESUMO

Como a maioria dos países está agora em estágios avançados da transição demográfica, a migração internacional tem ganhado crescente importância para explicar a dinâmica populacional. No entanto, a falta de dados de boa qualidade tem limitado o nosso conhecimento sobre as características da população migrante que vive em muitos destes países. Nesse contexto, os fatores socioeconômicos e de saúde associados à migração são temas críticos de pesquisa. Esta dissertação examina primeiramente a diversidade de perfis sociodemográficos entre migrantes internacionais adultos que vivem em São Paulo, Brasil. Também investigamos se e como os padrões de mortalidade desses migrantes diferem daqueles observados entre os brasileiros residentes no mesmo Estado. Baseamo-nos em dados do censo demográfico brasileiro de 2010 e aplicamos o método *Grade of Membership* (GoM) para definir os perfis sociodemográficos da população migrante. Encontramos quatro perfis de migrantes: migrantes recentes, de médio prazo, de longo prazo e idosos. As diferenças de perfis são marcadas principalmente pela idade, tempo de residência, escolaridade e país de nascimento. As ondas tradicionais de migração (europeus e japoneses) deram lugar a migrantes mais novos, vindos da América do Sul. Em seguida, examinamos os diferenciais de mortalidade entre migrantes e não-migrantes residentes em São Paulo e avaliamos se essas diferenças variam de acordo com a idade e a escolaridade. As estimativas são baseadas em modelos de regressão binomial negativo que combinam registros de óbitos do Sistema Brasileiro de Informações sobre Mortalidade de 2009 a 2011 com dados populacionais do censo demográfico brasileiro de 2010. Encontramos evidências de desvantagens de mortalidade entre migrantes em relação aos brasileiros residentes em São Paulo, mesmo após o controlarmos por idade, sexo e escolaridade. No entanto, esses diferenciais de mortalidade variam de acordo com a idade e escolaridade. A desvantagem da mortalidade migrante é altamente concentrada entre os adultos mais jovens com baixa escolaridade (indivíduos com maior probabilidade de virem de outros países sul-americanos). Nossos achados contrastam com o paradoxo da mortalidade entre migrantes em países desenvolvidos, especialmente nos Estados Unidos.

Palavras-chave: Migração Internacional; Diferenciais de Mortalidade; São Paulo; Brasil.

AGRADECIMENTOS

Não é uma tarefa fácil agradecer a todas as pessoas que de algum modo estiveram ao meu lado e que me ajudaram a concluir o mestrado em um período tão atípico como o que o temos vivido desde o início da pandemia em 2020. No entanto, preciso fazer jus àqueles que estiveram comigo nesses dois anos repletos de momentos serenos, mas também apreensivos. Inicio os meus agradecimentos ao corpo docente do Cedeplar pelos momentos valiosos de aprendizado. À minha professora e coordenadora do programa de pós-graduação em Demografia Laura Wong, por jamais medir esforços para me ajudar a alcançar um maior crescimento e qualificação profissional durante a minha jornada no mestrado. Ao Sebastião por sempre se prontificar a me ajudar com todas as questões burocráticas da academia de uma maneira atenciosa e com a maior celeridade possível.

Ao meu orientador Gilvan Guedes, obrigado por confiar em mim desde o início. Você me ensinou muito mais do que métodos estatísticos e novas formas de se fazer pesquisa. Mesmo sendo gigante, você sempre fez questão de romper com a relação assimétrica de poder entre aluno *x* orientador que existe dentro do ambiente acadêmico. Você me trouxe para perto de você. Me fez ampliar os meus horizontes e extrair o melhor de mim como pesquisador. Você me ensinou sobre coragem e altruísmo. Me ensinou a importância de ser presente em um contexto de completo distanciamento físico. E você esteve presente a todo momento. Deixo aqui o meu sentimento de gratidão por você ter desempenhado um papel crucial no meu processo de amadurecimento acadêmico nos últimos dois anos. Muito obrigado.

Ao Cássio, meus agradecimentos mais sinceros por você ter aceitado o convite de ser o meu coorientador no mestrado. Muito obrigado pelo estímulo, incentivo e principalmente pela inspiração durante todo o processo. Seus comentários sempre muito atenciosos e extremamente detalhados fizeram toda a diferença na qualidade do trabalho. Obrigado por sempre estar disposto a me ajudar e a compartilhar comigo sua vasta experiência sobre o tema. Me sinto honrado em tê-lo ao meu lado. Também agradeço por você ter me apresentado à Mirian Ribeiro, que gentilmente compartilhou comigo todos os seus conhecimentos sobre como trabalhar com diferenciais de mortalidade por escolaridade. Sem esse suporte, o meu trabalho seria ainda mais árduo. Espero um dia poder retribuir de uma forma tão atenciosa e solícita todo o apoio que recebi de você e da Mirian durante esse período. Muito obrigado por tudo.

Aos meus amigos da Demografia, Thiago, Amanda, Luís, Ana Luiza, Ana Carolina, Abdoul, Flávia, Maria Laura, Larissa Gonçalves e Nathália. Vocês tornaram a minha caminhada no mestrado muito mais leve. Minha profunda gratidão a vocês por toda companhia e parceria. Também agradeço ao meu irmão Hemanoel, que esteve sempre ao meu lado desde o meu primeiro suspiro de vida. São muitos anos de companheirismo, de lutas e de sonhos. Você faz parte dos melhores momentos da minha vida. Mesmo nos dias cinzas, você sempre esteve ao meu lado. Minha gratidão por você é eterna.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Agradeço ao CNPq por tornar a minha pesquisa factível. Como cidadãos, temos o dever de defender o financiamento público e o fortalecimento das agências de fomento pelo importante papel que elas desempenham no avanço da educação, ciência, tecnologia e inovação do país.

A quem não mencionei, mas esteve ao meu lado durante esse período, prometo reconhecer essa proximidade, ajuda e incentivo todos os dias da minha vida.

Content

Introduction

As the demographic transition advances worldwide, the relevance of migration for understanding population dynamics increases. Migration can change demographic, economic, and social structures and create a new cultural diversity in both origin and destination countries (Castles, Haas, & Miller, 2015; Livi-Bacci, 2011; Massey et al., 1998). For that reason, policymakers worldwide need timely, reliable, accessible, and comparable data on migration to manage international flows effectively and protect the rights of migrants (IOM, 2017).

International migration is among the leading policy issues of the 21st century (United Nations, 2020). However, lack of good data defies the agenda on the determinants of migration, especially in traditionally sending communities. For many countries, Brazil included, there is a gap in the understanding of the socioeconomic and health factors associated with the migration process. Lack of migration data often leads to misperceptions about the scale of migration and its effects (Zapata & Guedes, 2017). It can also result in poorly designed policies and interventions, making it harder to identify and assist migrants in vulnerable situations (IOM, 2017). Demographers can contribute to the debate by finding alternative methodologies to overcome the existing limitations and provide the best estimates possible (Nepomuceno, 2017).

Most previous research in developed countries has reported that migrants are healthier and live longer than native-born residents in their destination country (Guillot, Khlat, Elo, Solignac, & Wallace, 2018; Palloni & Arias, 2004; Wallace, Khlat, & Guillot, 2019; Wallace & Kulu, 2014; Young, 1987). This situation in which migrants have lower mortality than non-migrant populations is denominated '*migrant mortality advantage*'. It is one of the most pervasive findings from the social sciences (Guillot et al., 2018), and has garnered increased attention in recent years because of the growing share, diversification, and aging of foreign-born populations in high-income countries (Wallace & Wilson, 2021). However, migrant mortality advantage might not reflect the health outcomes of migrants living in developing countries (Aldridge et al., 2018). The scarcity of studies reflects the lack of data available for this group with adequate quality and representativeness. This is an important gap in the existing literature.

Despite the challenges in analyzing mortality differences between migrants and their host populations in developing countries, there is significant merit in doing so. Firstly, mortality is a robust health indicator and offers insights into the health conditions of migrants. Moreover, comparing migrant mortality with their host population can help discuss the factors that underpin these differences and rethink policies to increase their accessibility to national healthcare and welfare systems in developing countries such as Brazil.

This dissertation deals with migration-related issues, particularly among adult migrants living in São Paulo, Brazil. The core interest is to access the diversity of sociodemographic profiles of adult international migrants residing in the State of São Paulo and examine whether and how their mortality patterns differ from those observed among Brazilian-born adults in the same State.

São Paulo represents an ideal context to conduct this research. The State has high-quality mortality data, has full coverage of adult deaths (Queiroz, Freire, Gonzaga, & Lima, 2017), and is the leading destination of international migrants in Brazil. Currently, the State ranks as the largest financial center in South America and has the largest population in the region, with an estimated 46,649,132 residents in 2021 (IBGE, 2021).

This dissertation is divided into two essays. The first study investigates the diversity of sociodemographic profiles of adult international migrants living in the State of São Paulo in 2010. Based on data from the 2010 Brazilian demographic census, we apply the Grade of Membership (GoM) technique to define the sociodemographic profiles of foreign-born adults. Additionally, we examine the association between countries of origin and duration of stay since arrival in Brazil. We are particularly interested in measuring the differences between recent and long-term waves of migration to the State of São Paulo. For that, we use data from previous censuses to capture the changes in size and composition of migrant figures over the decades.

The second study examines the intersection between migration and mortality. We look at mortality differences between foreign-born and native-born adults living in São Paulo and evaluate whether these differences vary by age and education groups. We combine death registers from the Brazilian Mortality Information System from 2009 to 2011 with population figures from the 2010 Brazilian demographic census to estimate

death rates for migrants and non-migrants by age, sex, and educational attainment through negative binomial regression models.

We organized this dissertation as follows. Chapter 2 presents the paper entitled "***Sociodemographic profiles of foreign-born adults living in São Paulo, Brazil***". The following chapter is titled "***Mortality differences between foreign-born and native-born adults in São Paulo, Brazil***". Each study is organized into six sections: abstract, introduction, materials and methods, results, discussion, and conclusion. At the end of the dissertation, we offer concluding remarks, references, and further information about the research in the form of supplementary material.

Paper 1

Sociodemographic profiles of foreign-born adults living in São Paulo, Brazil

Abstract: International migration is a global phenomenon that has gained prominence in the public debate over the past decades. International organizations have encouraged nations to develop and use regional-specific migration profiles, including disaggregated data on all migration-relevant aspects in a regional context to foster evidence-based migration policies. This paper contributes to the topic by investigating the diversity of sociodemographic characteristics of adult migrants living in São Paulo. We rely on data from the 2010 Brazilian demographic census and apply the Grade of Membership (GoM) method to delineate the sociodemographic profiles of these populations. The results suggest four distinct profiles of foreign-born adults: recent, middle-term, long-term, and old migrants. The most relevant characteristics that distinguish these profiles are age, length of residence, education, and country of birth. Recent migrants are younger adults, mainly from Bolivia, with few years of residence in São Paulo and low socioeconomic status. Middle-term migrants, formed predominantly by white men born in Argentina, Chile, and other countries, are active in the labor market and have high education and income levels. Long-term migrants are married, aged between 60 and 70 years, with intermediate education and wide variation in income level. Since this group includes the nationalities with the largest contingent of migrants, especially the Portuguese, this profile is the most prevalent among adult migrants. Finally, in the four profile - old migrants - the sociodemographic characteristics include a high proportion of retired widows, born predominantly in Japan.

Keywords: International Migration. Grade of Membership. Brazil.

Introduction

International migration is a global phenomenon that has gained prominence in the public debate and in the governmental agenda of many countries over the past decades. Almost all world regions are affected by the arrival, departure, or transit of populations that have become more mobile and diverse in terms of sociodemographic profiles (United Nations, 2020; Zapata & Guedes, 2017). For that reason, reliable evidence on migrants has been vital for assessing current and future trends in societies shaped by the presence of migrants in their territory. That is the case of the State of São Paulo, Brazil (OIM, 2009).

São Paulo is the leading financial center and the most populous state in Brazil. Therefore, it attracts individuals from different countries to study or work, making migration flows highly heterogeneous (De Maria & Baeninger, 2016). Social scientists have devoted special attention to examining the sociodemographic characteristics of migrants residing in São Paulo. Usually, they look at the countries of origin, socioeconomic status, the reasons for migrating, and types of occupation in the labor market (De Maria & Baeninger, 2016; Domeniconi & Baeninger, 2017; Dutra, Almeida, Tonhati, & Palermo, 2015; Magalhães, Bógus, & Baeninger, 2018; OIM, 2009; Oliveira, 2015).

Understanding the sociodemographic profiles of migrant populations is critical for several reasons. First, migration can change demographic, economic, and social structures and foster cultural diversity, thus affecting health, housing, education, occupation, and transport needs (Castles et al., 2015; Koser, 2010; Livi-Bacci, 2011). Second, evidence on the characteristics of migrants may help promote policies that ensure their access to the labor market and the health and welfare systems. It also dialogues with the first objective of the Global Compact for Migration[1], which encourages States to develop and use evidence-based migration policies from regional-specific migration profiles, including disaggregated data on all migration-relevant aspects in a regional context (United Nations, 2019).

This paper contributes to this policy-based migration agenda by investigating the diversity of sociodemographic characteristics of adult migrants living in the Brazilian State of São Paulo. We rely on data from the 2010 Brazilian demographic census and apply the Grade of Membership (GoM) technique to delineate sociodemographic profiles of this population and its subgroups. Additionally, we analyze how long migrants have resided in São Paulo since their arrival and evaluate what nationalities are part of recent and older waves of migration to São Paulo. For that, we use data from previous censuses to capture the changes in the size and composition of surviving[2] migrant populations over the decades.

[1] The Global Compact is the first inter-governmentally negotiated agreement, prepared under the auspices of the United Nations, covering all dimensions of international migration in a holistic and comprehensive manner. The Global Compact is framed in a way consistent with target 10.7 of the 2030 Agenda for Sustainable Development in which Member States committed to cooperate internationally to facilitate safe, orderly, and regular migration.

[2] Our target population in the demographic censuses are international migrants who survived mortality and re-(out)emigration by the time of the interview.

The paper is structured as follows. The following section presents a brief contextualization of international migration in Brazil, in which São Paulo has always played a prominent role. Next, we describe the data and the main variables used in this study to delineate the sociodemographic profiles of adult migrants living in São Paulo. We also provide more detailed information about the GoM method. In the fourth section, we present the overall changes in size and composition of migrant populations over the past decades and characterize them according to the profiles provided by the GoM method. Finally, we conclude and discuss future research on this topic.

Background

International migration in Brazil

Brazil's history of international migration reveals different stages and migrant profiles over time. As part of the colonization process, many migrants arrived in Brazil, catering to the interests of the Portuguese Crown and assuring territorial occupation (Holanda, 1995; Levy, 1974). The slave trade that followed ensured the highest number of Africans forcibly arriving in Brazil (Graham, 1973).

The literature points to at least three distinct waves of migration to Brazil, mainly motivated by the end of the slave trade (Amaral & Fusco, 2005; Levy, 1974). Each wave comprised different countries and migration profiles. The first one occurred between 1850 and 1903. It coincided with the beginning of the great international movements of Europeans towards the Americas, especially to the United States, Argentina, and Brazil (Graham, 1973). The 1890s marked the apogee of international migrants in Brazil (Levy, 1974). The primary migrants' birthplace countries were Portugal, Italy, and Spain. In addition to replacing the slave labor force, the arrival of migrants during this period was part of a eugenic strategy to miscegenate the native population according to the European racial pattern (Brito, 1995, 2004).

The second wave of migration continued with the expansion of coffee plantations due to the recovery of prices in the international market. It resulted in the absorption of many workers by the agricultural sector, especially in São Paulo (Graham, 1973). During this period, the Japanese arrived in Brazil, subsidized by the Japanese government and directed toward farms located in the interior of São Paulo. In 1920, the foreign-born accounted for more than 5% of the total population (more than one and a half million), clustered in the two main industrial centers - Rio de Janeiro and São Paulo (Hoffmann,

1977). The high concentration of migrants in São Paulo reflected the local subsidies until 1927 and the coffee-growing based economy (Patarra, 2003). When coffee production began to decline, the foreign-born population living in São Paulo formed the bulk of the labor force in the emerging industrial activity (Brito, 1995; Graham, 1973).

The third wave of international migration (1931-1953) was numerically smaller than the previous ones. The economic recession caused by the 1929 world crisis and the decline in coffee prices discouraged migratory flows to Brazil. In addition, the laws resulting from the measures established in the constitutions of 1934 and 1937 protected Brazilian-born workers and made it difficult for foreign-borns to seek employment in Brazil.

The 1964 military coup d'état directly impacted the flow of international immigrants (Amaral & Fusco, 2005). With a strongly patriotic bias, the new ideology defended that Brazil should trust native-born populations to fill the labor market, making the Brazilian economy primarily based on internal migration (Patarra & Fernandes, 2011). As a result, the country witnessed a sharp decline in international migrant flows (Patarra & Fernandes, 2011).

After the 1980s, emigration from Brazil started to grow. Carvalho (1996) estimated a loss of approximately 1.8 million people over a decade, with the United States, Japan, and Paraguay as the leading destinations (Patarra & Fernandes, 2011). However, even considering the highest estimates available, this group represented less than 2% percent of the total population (OIM, 2009).

The passage from the 20th to the 21st century brought new reflections on the study of international migration in the Brazilian context. Although the foreign-born population reduced significantly in recent decades, new migration flows have intensified (Oliveira, 2015). Part of the Brazilian-born population living abroad has returned to the country, and new international migrants have arrived. Two factors have favored this pattern: Brazil's economic stability since the late 1990s and the financial crisis in developed countries since 2008, which devastated the world financial system due to the bursting of the US housing bubble (Oliveira, 2013). Yet, the emigration of Brazilians did not stop; the arrival of new foreign-borns did not exceed the number of new departures, leading to migration rates close to zero in the first decade of the 21st century (Oliveira, 2015).

Brazil's historical and recent migration trends suggest the diversification of migratory flows in terms of motivations, origins, and sociodemographic characteristics (OIM, 2009). Unlike migrations from the late nineteenth century until the 1930s, in which people from the Global North (primarily Europeans) constituted the main migratory flows in the country, migrants from the Global South, including South-South flows, dominated the figures in recent periods (Zapata & Guedes, 2017). In the Brazilian migration history, São Paulo has always been prominent, reinventing and rebuilding itself as the preferred choice of residence for migrant populations.

Data and methods

Data, definitions, and selected variables

This research uses Brazilian census data gathered from IPUMS International. Although some studies have used administrative records to examine recent aspects of international migration in Brazil (Baeninger, 2019), census data have some advantages over other sources. When it comes to rare events like international migration, population censuses are often the only source capable of ensuring complete coverage of the entire population (Carvalho, Campos, Rigotti, & Pinho, 2018). In addition, they allow the characterization of international immigrants based on their age, sex, birthplace country, marital status, occupation, income, and education. It is often difficult to access all these dimensions through other databases.

Although this study includes data from all Brazilian censuses currently available at IPUMS (1960-2010), the emphasis will be on the most recent 2010 census. This census was used to examine the diversity of sociodemographic characteristics of adult migrants living in São Paulo. We restricted the analysis to adults over 30 years of age to ensure that most individuals have already completed their educational attainment.

We defined migrants as individuals born abroad, regardless of their place of birth, including those who acquired Brazilian citizenship or nationality after their arrival. We constructed nine other categorical variables from the census data (Box 1). All variables were observed at an ordinal or nominal qualitative measurement level. Missing data received a specific code for non-information.

Box 1. Selected variables and response categories

Variable	Category
Age group	30-39; 40-49; 50-59; 60-69; 70-79; 80+
Gender	Men; Women
Race/Color	White; Black; Indigenous; Asian; Brown (*pardo*)
Marital Status	Single; Married; Divorced; Widowed
Birthplace country	Portugal; Japan; Bolivia; Italy; Spain; China; Chile; Argentina; South Korea; Other
Length of residence (in years)	0-10; 11-20; 21-30; 31-40; 41-50; 50+
Education	Low (0-3 years); Middle (4-11 years); High (12 years or more)
Employment status	Employed; Unemployed; Inactive
Income (quintiles)	1st quintile; 2nd quintile; 3rd quintile; 4th quintile; 5th quintile;

Note: We included in the model only the ten leading birthplace countries. The remaining countries were included in the "others" category. In the case of the education variable, supplementary material provides a step-by-step algorithm for estimating years of completed education using the R software based on the combination of other variables from the 2010 census microdata.

Analytical strategy to characterize the sociodemographic profile of foreign-born adults

Migrant populations are heterogeneous in cultural identity, ways of living, social conditions, health behavior, and health risks (Spallek, Zeeb, & Razum, 2011). To access the diversity of sociodemographic profiles of migrants, we employed the Grade of Membership (GoM) fuzzy cluster technique. This method estimates the degree of unobserved heterogeneity in a multidimensional dataset. Unlike other multivariate methods, GoM does not require that individuals and objects be organized in well-defined sets (Manton, Woodbury, & Tolley, 1994). It means that GoM can cluster foreign-born adults by shared sociodemographic characteristics (sex, age, marital status, birthplace region, occupation, income, education, and others) while explicitly estimating the heterogeneity within each cluster.

The GoM mathematical model iteratively estimates two sets of parameters, λ_{ik} and g_{ik}. The first parameter describes the probability of an answer at level *l*, of the *j*th question, in the extreme profile *k,* by individual *i*, conditional to the score g_{ik}. The second parameter indicates the membership (*g*) of the individual (*i*) to the profile (*k*). For each individual in the sample, *K* grades of membership (g_{ik}) are estimated concerning the extreme profiles (reference groups). This parameter has no probabilistic interpretation and can be interpreted as an individual attribute measuring the multivariate distance from each

extreme profile (Caetano & Machado, 2009). Because it is an individual-based measure, ranging from 0 to 1, it is the explicit representation of heterogeneity of individuals derived by the model. The model parameters are identified under the following constraints:

$$g_{ik} \geq 0 \qquad for\ each\ i\ and\ j \tag{1}$$

$$\sum_{k=1}^{k} g_{ik} = 1 \qquad for\ each\ i \tag{2}$$

The formulation and estimation of the model parameters require the following assumptions:

a) Random variables (Y_{ijl}) are independent for the different individuals.

b) The g_{ik} (k = 1, 2, ..., k) are realizations of the components of the random vector $\zeta_i = (\zeta_{il}, \ldots, \zeta_{ik})$ with distribution function $H(x) = P(\zeta_i \leq x)$. GoM scores are realizations of random variables when an individual is selected from the population. The sample distribution of realizations (the scores in the sample) provides estimates of the H(x) distribution function.

c) If the degree of membership g_{ik} is known, the responses of individual *i* to the various Y_{ijl} questions are independent for the categories of each variable.

d) The probability of answer *l*, for the *j*th question, by the individual with the extreme *k*th profile is λ_{kjl}. By the model's assumption, at least one individual is a well-defined member of the *k*th profile. This assumption gives the probability of that individual's response to the various levels of each question. We can rewrite this assumption as:

$$\lambda_{kjl} \geq 0 \qquad \text{for each } k, j, \text{ and } l \tag{3}$$

$$\sum_{l=1}^{L} \lambda_{kjl} = 1 \qquad for\ each\ k\ and\ j \tag{4}$$

e) The probability of a level *l* answer, of the *j*th question, by individual *i*, conditioned by the g_{ik} score, will be given by:

$$0 \leq P(Y_{ijl} = 1) = \sum_{k=1}^{k} g_{ik}\, \lambda_{kjl} \leq 1 \tag{5}$$

The probability model for constructing the maximum likelihood estimation procedure relies on these five assumptions. The probability model for a random sample

is the product of the multinomial random variable, with the probability of each cell given by:

$$E(Y_{ijl}) = \sum_{k=1}^{k} g_{ik}\,\lambda_{kjl}\,, \tag{6}$$

where g_{ik} is known and greater than or equal to 0 and less than or equal to 1. Given the independence of individuals and conditional on the λ_{kjl} and g_{ik} values, the maximum likelihood function is defined as:

$$L(y) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{L}\left(\sum_{k=1}^{k} g_{ik}\,\lambda_{kjl}\right)^{y_{ijl}} \tag{7}$$

This function is estimated iteratively based on the numerical method of Gradient descent for parameter estimates convergence (Woodbury, Clive, & Garson, 1978). These iterations continue until L(y) reaches its maximum value. To ensure that our model was representative of the most prevalent profiles in the migrant populations (that is, the converged likelihood sits at the global maximum), we performed five hundred Monte Carlo simulations with an initial λ_{kjl} for each k as randomly drawn from uniform distributions.

The prevalence of extreme profiles in the population was calculated as follows:

$$P_k = \frac{\sum_{i=1}^{1} g_{ik}}{\sum_{i=1}^{1} i} \quad \textit{With k = 1,2...K.} \tag{8}$$

Such a prevalence is a multivariate weighted average because the weight corresponds to the proportion of the population sharing some degree of attributes from profile *k* (Guedes, Siviero, Machado, Pinto, & Rodarte, 2016).

To describe the extreme profiles, we followed the criterion used by Sawyer *et al*. (2002): a category *l*, of a variable *j*, is characteristic of profile *k* if the ratio between the estimated marginal frequency and the observed marginal frequency (Lambda-Marginal Frequency Ratio, LMFR) has a value greater than 1.20 – an estimated probability 20% higher than the occurrence of this category in the population. This means that whenever the estimated probability of a category occurring in a given profile is at least 20% higher

than that observed in the sample, that category is classified as dominant in the profile. Values that fulfill this condition are highlighted in bold (Annex 1).

We defined the number of extreme profiles by assessing their substantive significance. We checked whether the observed changes could be theoretically justified for each new profile created. In the end, we considered the model with four extreme profiles[3]. All analyses were conducted using *gom* library (Andrade & Guedes, 2021) for R (R Core Team, 2021).

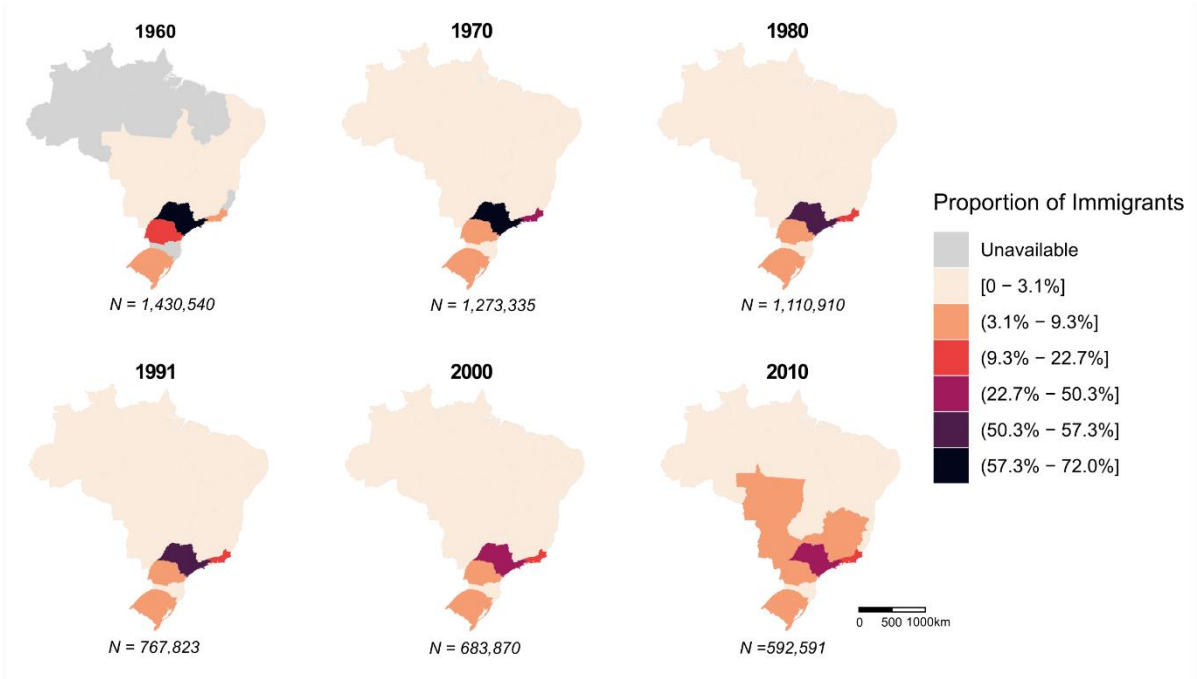
Results

International migration waves to São Paulo

Figure 1 provides a brief overview of the primary waves of migration to São Paulo. In 1960, about 72% of 1.4 million migrants living in Brazil were in São Paulo. From one decade to another, economic changes – particularly the growth of industrial activities and the expansion of the agricultural frontier to Paraná, Goiás, and Mato Grosso (Cano, 1988)– impacted the migratory movements and the spatial distribution of migrants in Brazil (Figure 1). Centripetal forces exerted by São Paulo decreased from the 1980s onwards but did not disappear (Baeninger, 2012). The state was still the largest destination of migrants in 2010, comprising 45% of migrants living in the country.

[3] Based on a technical criterion, such as the Akaike Information Criteria (AIC), the optimal number of extreme profiles is two (AIC = 4972228). However, the model with two extreme profiles generated an excessive aggregation of sociodemographic characteristics, hiding important differences between adult migrants. The same situation happened with the model with three extreme profiles. On the other hand, a number of profiles greater than four tended to homogenize the results; they started showing a high similarity in terms of sociodemographic characteristics to each other.

Figure 1. Spatial distribution of international migrants in Brazil (1960-2010)



Source: IPUMS International. Minnesota Population Center. Brazilian Demographic censuses (1960-2010).

Between 1960 and 2010, the age distributions of the foreign-born population living in São Paulo and the total population differed (Figure 2). There was a much higher concentration of migrants at older ages, especially women born in European countries (Portugal, Spain, and Italy). These are mostly surviving individuals from the cohorts who moved to Brazil from 1890 to 1950, at working ages, following the selective nature of migration (Chiswick, Lee, & Miller, 2008). Yet, the most recent flows from South American countries increased the relative participation of the foreign-born population at ages younger than 30 in 2010.

Figure 2. Distribution of international migrants and total population by age and gender – State of São Paulo, Brazil (1960-2010)

Source: IPUMS International. Minnesota Population Center. Brazilian Demographic censuses (1960-2010).

Table 1 summarizes the total number of international migrants residing in São Paulo between 1960 and 2010, according to major birthplace regions and the annual rate of population change. Europe is the birthplace region that historically concentrates most migrants since 1960. However, the relative participation of Europe has decreased over the years. In 1960, Europeans represented more than 70% of migrants, whereas in 2010, the number reduced to 45%. This scenario suggests two critical trends in migratory dynamics, especially in the first decades of the 21st century: the diversification of migrant nationalities and the intensification of South-South flows in the state (Baeninger, 2012). Indeed, those born in South America, who previously accounted for just 2% of migrants in 1960, surpassed 23% in 2010. The population of South Americans grew at an annual rate of 3.5% between 2000-2010.

Table 1. Total population of international migrants residing in São Paulo by major birthplace regions and the annual rate of population change (1960-2010)

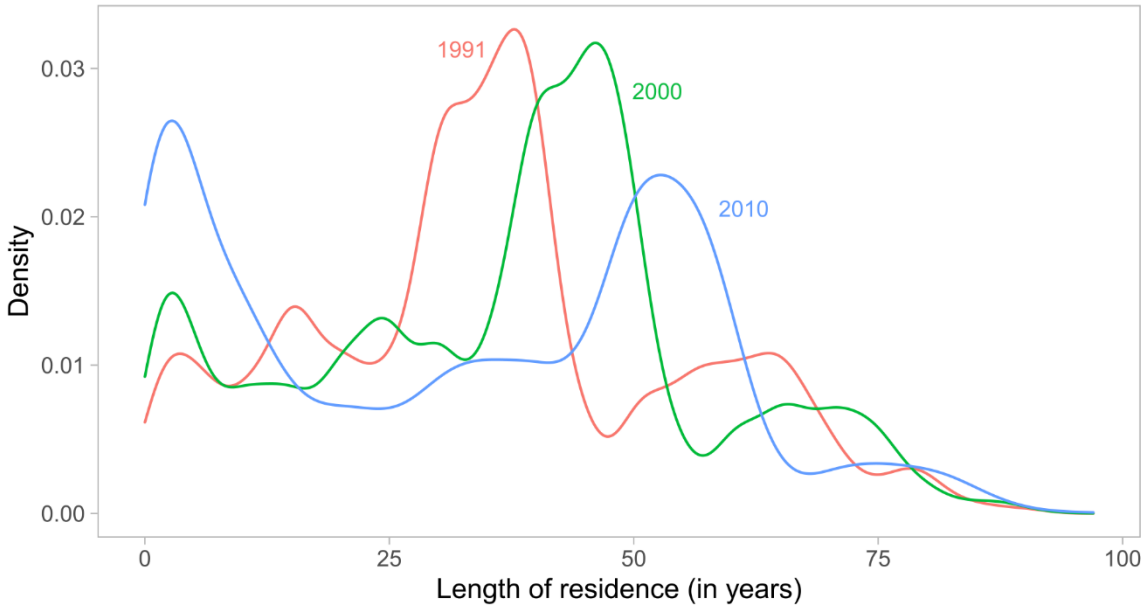
Birthplace region	Year					
	1960	**1970**	**1980**	**1991**	**2000**	**2010**
Africa	6,515	4,980	8,879	6,846	7,222	6,730
Asia	192,010	155,387	142,729	95,232	82,394	64,011
Europe	795,960	541,065	422,538	261,856	203,185	121,151
America	28,505	26,463	48,285	47,551	50,710	71,021
North America	6,055	5,757	5,928	5,222	4,832	6,434
Central America & Caribbean	2,290	811	1,590	1,514	1,787	2,083
South America	20,160	19,895	40,767	40,815	44,091	62,504
Oceania	965	287	339	262	193	140
Other countries	2,265	1,536	4,986	2,536	259	3,736
Total	**1,026,220**	**729,718**	**627,756**	**414,283**	**343,964**	**266,789**

Birthplace region	*Annual rate of population change (%)*				
	1960-1970	**1970-1980**	**1980-1991**	**1991-2000**	**2000-2010**
Africa	-2.7	5.8	-2.4	0.6	-0.7
Asia	-2.1	-0.8	-3.7	-1.6	-2.5
Europe	-3.9	-2.5	-4.3	-2.8	-5.2
America	-0.7	6.0	-0.1	0.7	3.4
North America	-0.5	0.3	-1.2	-0.9	2.9
Central America & Caribbean	-10.4	6.7	-0.4	1.8	1.5
South America	-0.1	7.2	0.0	0.9	3.5
Oceania	-12.1	1.7	-2.3	-3.4	-3.2
Other countries	-3.9	11.8	-6.1	-25.4	26.7
Total	-3.4	-1.5	-3.8	-2.1	-2.5

Source: IPUMS International. Minnesota Population Center. Brazilian Demographic censuses (1960-2010).

Another crucial dimension for analyzing migratory flows is how long migrants have resided in São Paulo after their arrival. It allows capturing the waves of international migration that have taken place in the state in recent and previous decades. Figure 3 shows the length of residence of international migrants living in São Paulo at three different points in time: 1991, 2000, and 2010. There are two well-defined waves of international migration. Alongside the long-term European migrants residing more than 50 years in São Paulo, the core trend is the renewal of arrivals in the first decade of the 21st century, especially those from South American countries (2010 curve).

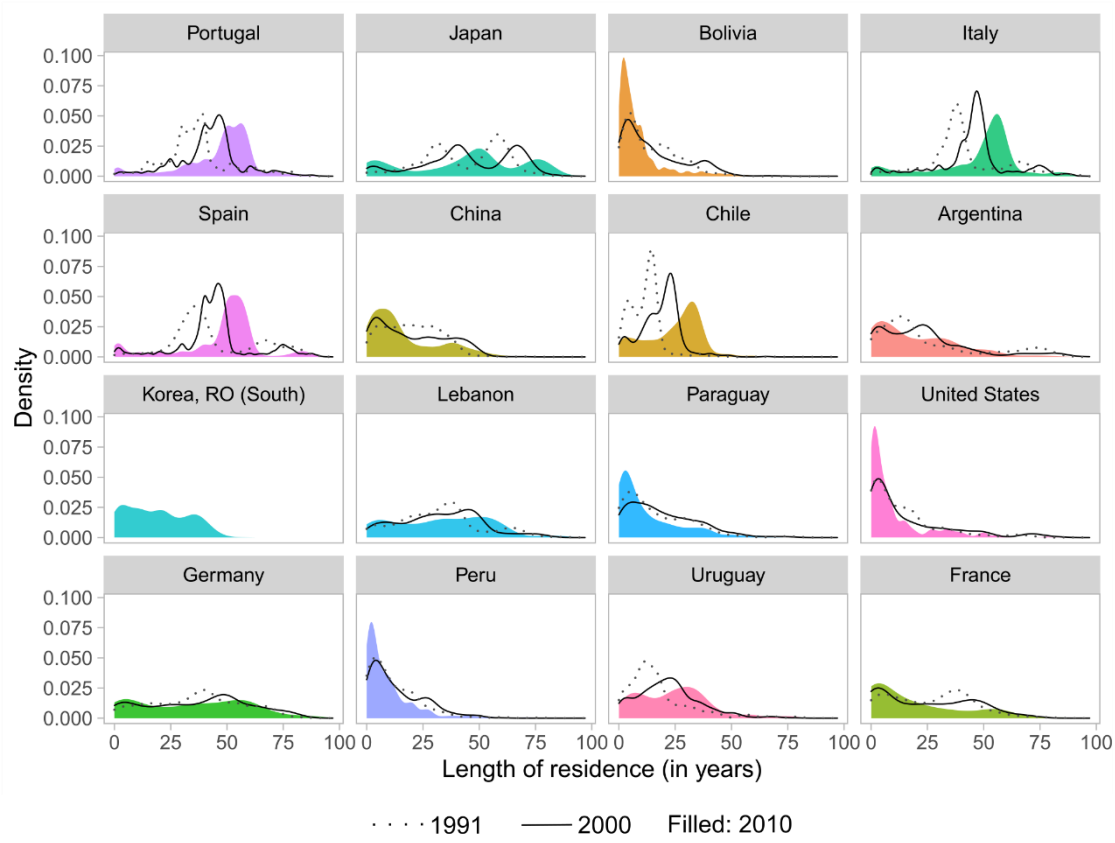
Figure 3. Length of residence of international migrants – State of São Paulo (1991-2010)



Source: IPUMS International. Minnesota Population Center. Brazilian Demographic censuses (1991-2010).

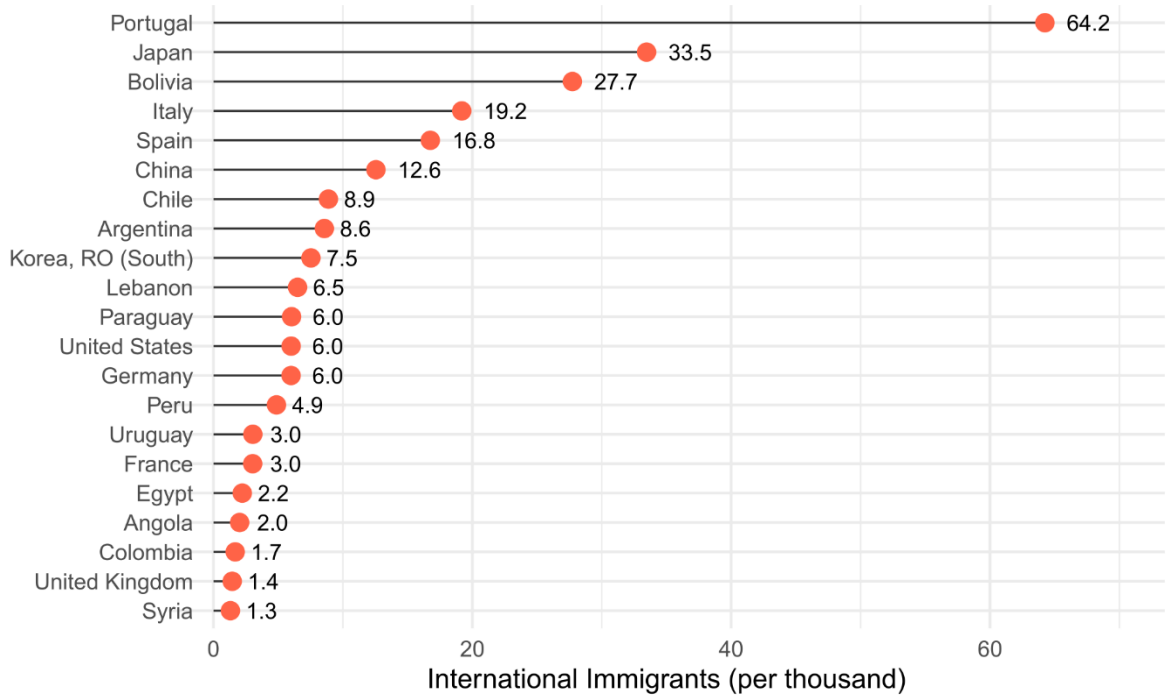
The decomposition of migratory waves by country of birth reveals greater diversity (Figure 4). Recent migration waves are formed by individuals born in South American countries – especially Bolivia, Paraguay, and Peru – and China and South Korea. The composition of current migration flows contrasts with the historical ones formed by Portuguese, Spanish, Italians, and Portugal. However, these population groups are still among the largest foreign-born groups in São Paulo (Figure 5). Individuals born in Chile and Uruguay challenge the assumption that migrants are divided between recent and long-term migrants. These groups break with the traditional dichotomy, suggesting the existence of intermediate profiles of migration in São Paulo (Figure 4).

Figure 4. Length of residence of international migrants by main countries of birth - State of São Paulo, Brazil (1991-2010)



Notes: Birthplace countries ranked from highest to lowest number of immigrants in 2010. There is no available information about the population of South Korea in 1991 and 2000 because demographic censuses have joined North Korea and South Korea into a single category. Source: IPUMS International. Minnesota Population Center. Brazilian Demographic censuses (1991-2010).

Figure 5. Main countries of birth of international migrants living in the state of São Paulo, Brazil (2010)



Source: IPUMS International. Minnesota Population Center. Brazilian Demographic census (2010).

Sociodemographic profiles of adult migrants living in São Paulo in 2010

Results from the GoM method suggest the existence of four distinct profiles of international migrants residing in São Paulo. These profiles differ in several dimensions, including income, education, and work status[4]. We describe below the main predominant sociodemographic characteristics and the prevalence of each extreme profile among the population under analysis (Figure 6).

Recent migrants. The first extreme profile comprises 3.29% of the 219,326 adult migrants living in São Paulo in 2010. They are younger adults, single or divorced, with up to 30 years of residence in São Paulo. They are predominantly Indigenous, Asian, Black, or *Pardos,* with primary/secondary education. The main birthplace countries are Bolivia, China, and South Korea. Most of them are employed and unemployed with income levels

[4] Annex 1 presents the selected variables and the marginal distributions of absolute and relative frequencies for each category analyzed. It also shows the values of λ_{kjl} in each extreme profile. When compared with the corresponding marginal frequency, they indicate the dominance in a given profile.

at the 1st and 3rd quintiles. Profile 1 records the lowest weighted prevalence (15.8 %) of membership among adult migrants.

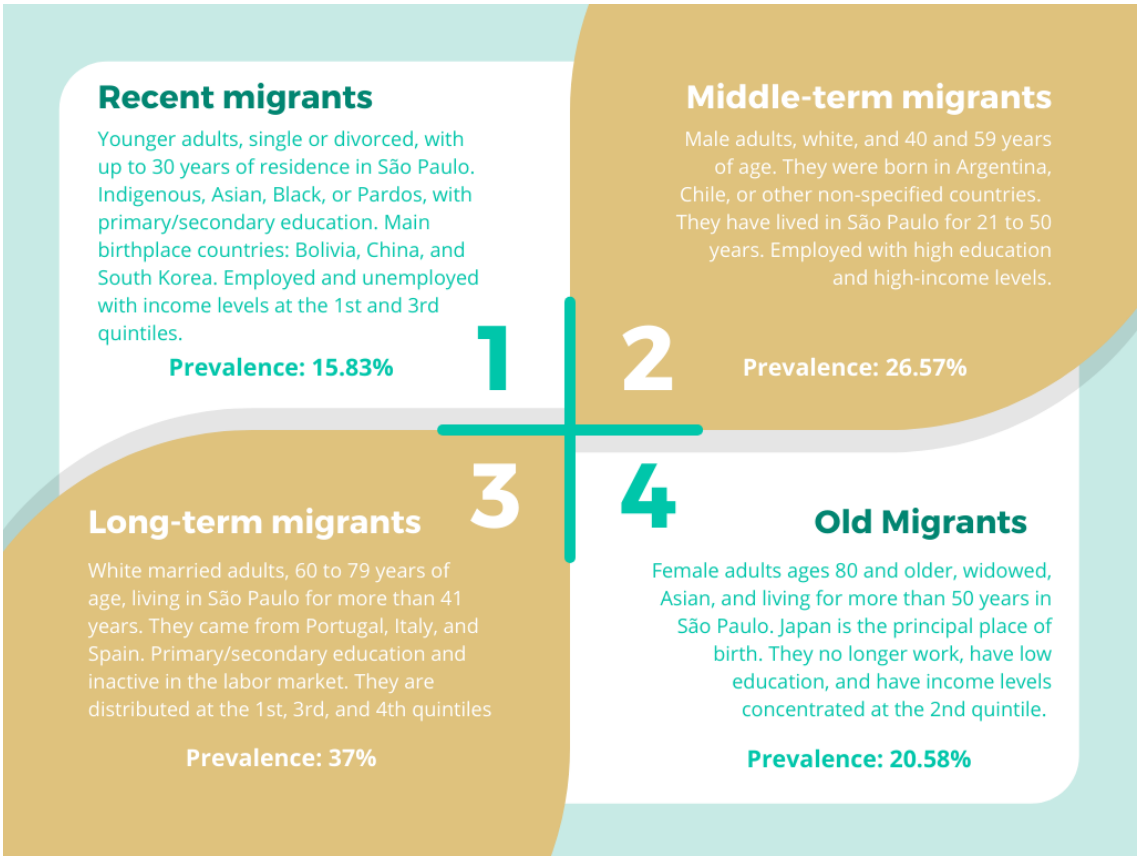
Middle-term migrants. The second extreme profile comprises another 3% of adult migrants. They are individuals born in Argentina, Chile, and other non-specified countries[5] who are primarily white, male, between 40 and 59 years of age. They have lived in São Paulo for 21 to 50 years. They are employed and have high education and high-income levels. The prevalence of these characteristics among the population under study is 26.56%.

Long-term migrants. The third extreme profile includes 5.06% of the adult migrants, comprising white married adults, ages 60 to 79, living in São Paulo for more than 41 years. There is no predominance of men or women in this profile. They came from Portugal, Italy, and Spain, have primary/secondary education (4-11 years), and are no longer active in the labor market. They are distributed at the 1st, 3rd, and 4th quintiles. The results suggest that long-term migrants have the highest prevalence among adult migrants living in São Paulo. About 37 in 100 migrants have the characteristics of this profile.

Old migrants. The fourth extreme profile is very different from the first one. Migrants from this profile are female adults ages 80 and older, widowed, Asian, and living for more than 50 years in São Paulo. Japan is the principal place of birth. They no longer work, have low education, and have income levels concentrated at the 2nd quintile. Pure type 4 represents only 1.3% of adult migrants. The prevalence of this profile in the population is about 20%.

[5] Birthplace countries specified in the model: Portugal, Japan, Bolivia, Italy, Spain, China, Chile, Argentina, and South Korea.

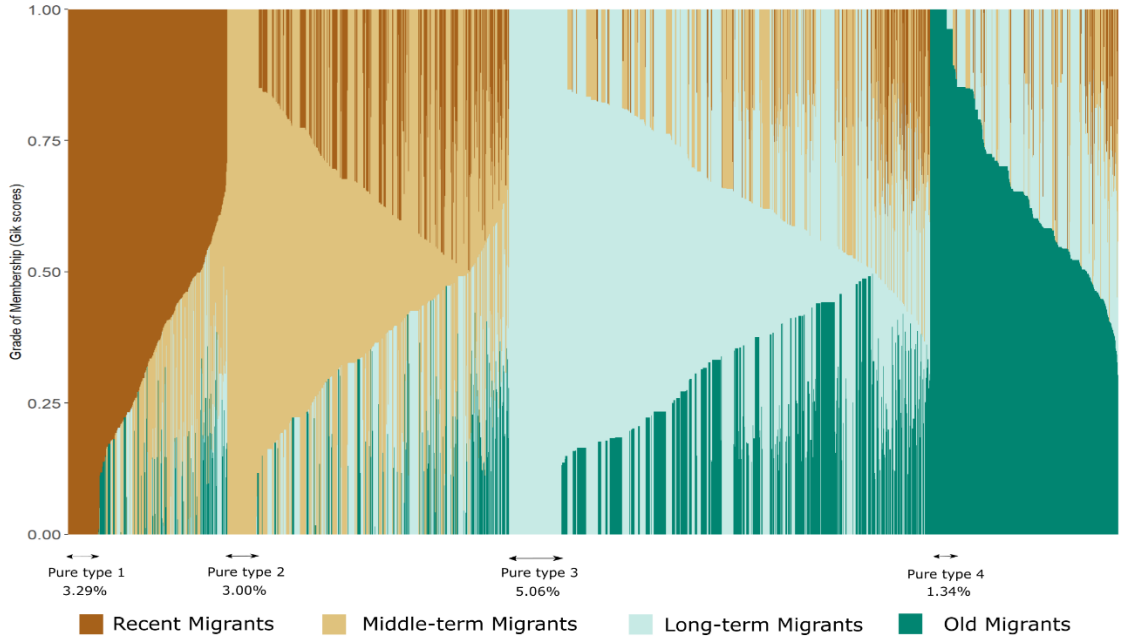
Figure 6. Sociodemographic profiles of international migrants, according to predominant characteristics – State of São Paulo (2010)



Source: IPUMS International. Minnesota Population Center. Brazilian Demographic census (2010). Own elaboration.

Together, the four extreme profiles included only 12.7% of migrant adults living in São Paulo in 2010. Therefore, another 87.3% differed somewhat from them, revealing the heterogeneity of migrants in terms of sociodemographic characteristics (Figure 7). By definition, the more an individual differs from the pure type profile in terms of individual attributes, the lower its importance for their categorization (Guedes et al., 2016). Although it is challenging to deal with such heterogeneity, the results suggest relevant characteristics in each profile that can give rise to new profiles capable of merging different sociodemographic characteristics from these extreme profiles, as illustrated in Figure 7. This information is relevant as it requires different strategies and policies to respond to health, housing, and occupational needs. It can also help gain further insights into the cultural and demographic diversity of the migrant population residing in São Paulo in 2010.

Figure 7. Grade of membership (g_{ik}) to the extreme profile k of adult migrants– State of São Paulo, Brazil (2010)



Source: IPUMS International. Minnesota Population Center. Brazilian Demographic census (2010).

Discussion and Conclusion

This paper investigated the diversity of sociodemographic characteristics of adult migrants living in São Paulo. We relied on data from the 2010 Brazilian demographic census and applied the Grade of Membership (GoM) technique to delineate the sociodemographic profiles of adult migrants. We also analyzed how long migrants have resided in São Paulo after their arrival, their nationalities, and the changes in size and composition of migrant populations over the decades.

Adult migrants living in São Paulo are highly heterogeneous. However, some common sociodemographic characteristics allowed us to identify four distinct migration profiles: recent, middle-term, long-term, and old migrants. The most relevant characteristics that distinguish these profiles are age, length of residence, education, and country of birth. We summarize below the main findings of this research.

The migrant population has a significantly aged composition than that of Brazilians. However, this is not true for all foreign-born subgroups. Recent migrants, for

example, are different. In addition to having a short period of residence in the state and a lower average age than other migrants, this group also has a lower socioeconomic status. Although South Koreans and Chinese are also included in this group because they have been residing in the state for a short time and are younger, low socioeconomic status is not a characteristic that manifests among them with such intensity, as observed among Bolivians. Recent literature already presents some studies that point to the precarious work relationships of Bolivians who currently reside in the capital of São Paulo, whose demographic characteristics dialogue with the findings of our work (Magalhães et al., 2018; Oliveira, 2015).

The middle-term and long-term migrant profiles present similarities but differ mainly in age, length of residence, and socioeconomic status. Middle-term migrants, formed predominantly by white men born in Argentina, Chile, and other countries not specified in the model, are active in the labor market and have high education and income levels. On the other side, the long-term migrants are married, aged between 60 and 70 years, with intermediate education and wide variation in income level. Since this group includes the nationalities with the largest contingent of migrants, especially the Portuguese, this profile is the most prevalent among adult migrants. Finally, in the four profile - old migrants - the sociodemographic characteristics include a high proportion of retired widows, born predominantly in Japan.

Overall, this paper contributes to the literature by analyzing the profile of migrants in São Paulo, which concentrates the largest foreign-born population in Brazil. It opens new possibilities for future research with alternative and more recent databases that include other dimensions not explored in this paper, such as health status, housing conditions, and spatial patterns of these populations across the municipalities of São Paulo.

The evidence provided about the predominant sociodemographic characteristics in each profile of adult migrants leaves some questions to answer in future research. Although we have accessed the diversity among different groups of migrants, it is vital to understand how these characteristics differ from those observed among native-born living in São Paulo. Comparing migrants with Brazilian-born adults can help discuss the impact of migration on the demographic dynamics, examine the factors that underpin

these differences, and devise possible strategies that ensure migrants have the same opportunities to access public services for migrants and non-migrants.

From another perspective, it is important to assess how the profile of these populations differs from those who remained in their country of origin. A significant challenge that emerges from this last question is the lack of available databases with adequate quality to carry out comparative research of this nature. However, the lack of data or the existence of inaccurate data has never precluded demographers from finding alternative methodologies to overcome the existing limitations and to provide the best estimates possible (Nepomuceno, 2017).

Through reliable evidence of the sociodemographic profiles of migrants, researchers will be able to advance the understanding of the diversity that permeates this population. It can also enhance policy coherence, help evidence-based policymaking on migration, and improve the mainstreaming of migration into development plans in the most populous state in South America.

Paper 2

Mortality differences between foreign-born and native-born adults in São Paulo, Brazil

Abstract: Previous research showed that migrants have lower mortality rates than natives in several high-income countries. However, evidence on the migrant-native mortality gap in developing countries remains scarce. This paper examines mortality differences between foreign-born and native-born adults living in the Brazilian State of São Paulo and evaluates whether these differences are associated with specific age and education categories. We combined death registers from the Brazilian Mortality Information System from 2009 to 2011 with population figures from the 2010 Brazilian demographic census. We adjusted death rates for migrants and non-migrants by age, sex, and educational attainment through negative binomial regression models. The main findings confirm the hypothesis that mortality levels of foreign-born differ from those observed among Brazilian-borns living in São Paulo when controlling for age, sex, and education. However, unlike the commonly observed pattern in developed countries, migrants have a mortality disadvantage compared to non-migrants. Migrant mortality disadvantage is highly concentrated among younger foreign-born adults with low education. Therefore, there is insufficient evidence to support the existence of a mortality paradox in São Paulo.

Keywords: International Migration. Health Inequalities. Migrant Mortality. Brazil.

Introduction

As countries advance the demographic transition, migration becomes critical for understanding population dynamics. However, migrants differ from the population in the destination countries in many attributes, including fertility and mortality levels. These differences can take many forms, depending on the early-life conditions, health behaviors, sociodemographic characteristics, life-course stages, and reasons for migrating (Hermalin, Ofstedal, Sun, & Liu, 2009).

The literature on mortality differences between foreign-born and native-born adults has gained relevance in the last decades, especially in high-income countries, with some consistent findings. For instance, most studies have shown that migrants have lower mortality natives in the United States (Guillot et al., 2018; Palloni & Arias, 2004; Turra &

Goldman, 2007), Australia (Young, 1987), Belgium (Anson, 2004; DeBoosere & Gadeyne, 2005), Sweden (Wallace & Wilson, 2021), France (Wallace et al., 2019), Denmark (Norredam, Olsbjerg, Petersen, Juel, & Krasnik, 2012), England (Wallace & Darlington-Pollock, 2020), and Spain (Gimeno-Feliu et al., 2019). Some other studies, in contrast, have found no difference (Hedlund, Pehrsson, Lange, & Hammar, 2008) or even higher levels of migrant mortality compared with natives (Rodriguez et al., 2017; Wild, Fischbacher, Brock, Griffiths, & Bhopal, 2007). For that reason, much of the recent debate on mortality patterns of international migrants revolves around whether the migrant mortality advantage is pervasive or not (Shor & Roelfs, 2021). These issues are even more urgent in middle and low-income countries, which still lack reliable data to examine how migrants differ from the native-born population in terms of health and mortality. It is a critical gap in the existing literature (Abubakar et al., 2018; Aldridge et al., 2018).

Despite the challenges in analyzing mortality differences among foreign-born and native-born adults in developing countries, there is significant merit in doing so. First, mortality is a robust indicator of health and provides essential information about the population's survival trajectory. Second, documenting the differences in mortality between migrants and non-migrants is the first step in understanding the factors that underpin these differences, which are crucial for developing public policies that facilitate the accessibility of the migrant population to the health and welfare systems.

This paper examines mortality differences between foreign-born and native-born adults living in the Brazilian State of São Paulo and evaluates whether these differences are associated with specific age and education categories. São Paulo represents an ideal context to conduct this research. The State has high-quality mortality data, with full coverage of adult deaths (Queiroz, Freire, Gonzaga, & Lima, 2017). It is also the leading destination of international migrants in Brazil, accounting for 45% of all foreign-born individuals in 2010. Currently, the State ranks as the largest financial center in South America and has the largest population in the region, with an estimated 46,649,132 residents in 2021 (IBGE, 2021).

The following sections describe the theoretical background, the methods used to test the hypotheses, the main findings, and potential explanations for the mortality inequalities between foreign-born and native-born adults in São Paulo, Brazil.

Theoretical Framework

The literature on migration suggests a substantial heterogeneity in the migration-mortality association. In many cases, migrants appear to have better health than expected compared to the native-born population (Shor & Roelfs, 2021). This result is paradoxical, given the relatively lower socioeconomic status of migrants. This paradox has generated many possible explanations (DeBoosere & Gadeyne, 2005). Below we outline the theoretical relevance of three common causes (selection processes, cultural effects, and data artifacts) and the rationale for their inclusion in our analyses. Even if they draw upon the experience of high-income countries, these explanations can provide further insights into the discussion of mortality differences in Brazil.

Selection Processes

Selection process hypotheses consider the selection of migrants as the primary source of diverging mortality rates (DeBoosere & Gadeyne, 2005; Kohls, 2010; Riosmena, Wong, & Palloni, 2013). Two hypotheses are worth mentioning: the healthy-migrant effect and the salmon-bias effect.

According to the healthy-migrant hypothesis, selection occurs in the population of origin. The selection process may be strong enough so that migrants end up being also more robust, on average, than the population living in the origin country (DeBoosere & Gadeyne, 2005; Guillot et al., 2018). Positive health selection will be most prominent for individuals who migrate for study or work (young adult ages) and less relevant for children and older adults near retirement ages, for whom individual characteristics play a less critical role (Chiswick et al., 2008; Guillot et al., 2018).

In the case of the salmon-bias effect, selection occurs when migrants with lower health status return to their country of origin (DeBoosere & Gadeyne, 2005). In a broader sense, it includes all migrants, regardless of age, who return to their original countries because they are less adaptive to endure harsh working and living conditions and are more likely to experience higher mortality (DeBoosere & Gadeyne, 2005). This hypothesis has been developed concerning the return of older migrants from the U.S. to Latin America and other origin areas (Abraído-Lanza, Dohrenwend, Ng-Mak, & Turner, 1999). As a result of this "unhealthy return-migration," the foreign-born population living in the host

country may show better health, on average, than would have been observed in the absence of out-migration (Guillot et al., 2018).

Overall, selection effects are theorized to be strongest among recent migrants and wear off with time spent in the host country (Harding, 2003). This selection process may be accelerated by exposure to adverse social conditions and/or progressive acculturation to prevailing beliefs, attitudes, and behaviors of the host society, which causes a shift in the disease patterns of migrants towards that of the host population (Wallace et al., 2019).

Cultural Effects

The central premise of cultural effects is that migrant mortality is a function of social and cultural characteristics that differentiate foreign-born from native-born populations rather than selection (Abraído-Lanza et al., 1999; Palloni & Arias, 2004). Culture can be outlined as a linked group of customs, practices, and beliefs jointly held by individuals, social networks, and groups (Abubakar et al., 2018). These factors help define who the migrants are and where they stand in relation to those who remained in their place of origin or those born in the country of destination.

Migrant populations are heterogeneous concerning cultural identity, ways of living, social situation, health behavior, and health risks (Spallek et al., 2011). Culture affects mortality outcomes by influencing individual health and lifestyle behaviors, family structure, and social networks (Palloni & Arias, 2004). Additionally, the explanation of cultural effect suggests that migrants may have better health outcomes than non-migrants because they tend to benefit from denser social support networks, including stronger family ties (Guillot et al., 2018; Palloni & Arias, 2004). Migrants' health advantage is also assumed to result from cultural buffering (i.e., retention of migrant cultures). In contrast, health deterioration with increased years of residence results from cultural assimilation in the host country (Vang, Sigouin, Flenon, & Gagnon, 2017).

Data Artifacts

Data artifacts encompass a broad range of potential error sources that disproportionately affect migrant mortality measures (Palloni & Arias, 2004; Riosmena et al., 2013). There are three equally salient data problems: (i) misreporting of age in death registers or population estimates; (ii) under-registration of deaths, (iv) population

coverage issues; (iii) mismatches of records (Guillot et al., 2018; Palloni & Arias, 2004). These problems are relevant for mortality estimates calculated by matching death and census records and where the individuals' migratory status is defined according to country of birth.

Misreporting of age in death registers or population estimates affects mortality rates specifically for the foreign-born populations. Migrants from developing countries often lack proper documentation about their actual date of birth, contributing to age misreporting both on census records and death certificates (Guillot et al., 2018). Preston, Elo & Stewart (1999) analyzed the effects of age misreporting on mortality estimates at older ages and showed that such errors tend to produce death rates that are too low, with increased biases at advanced ages. Following this logic, Guillot *et al*. (2018) proposed that age misreporting could decrease the relative mortality risk of migrants with age, particularly at older ages – this pattern may be pronounced for migrant groups originating from lower-income countries, who are more subject to age misreporting.

When analyzing mortality differences between foreign-born and native-born populations, the under-registration of deaths and population coverage issues are essential to consider. Migrants are more likely to be undercounted, especially if they live alone or travel with their household members (Guillot et al., 2018). Also, the proportion of undocumented migrants (who are typically undercounted in censuses) is expected to decrease with age (Guillot et al., 2018). However, it is difficult to assess the net effect of these errors without more information on the processes discussed above.

The third source of error (mismatches of records) that could lead to downward/upward biases in migrant mortality rates is pertinent for studies that calculate mortality rates by dividing death and census records. While this mismatch may not generate critical errors for the native-born population, it is potentially problematic for the foreign-born population (Guillot et al., 2018). Although the mismatches of records, which is a critical problem when using ethnicity to determine the origin of migrants (Rosenberg et al., 1999), it is not so relevant when migrants are defined based on the country of birth since it is less likely subject to response bias.

Overall, selection effects, culture, and data artifacts are some of the possible explanations for mortality differences between migrants and non-migrants, especially the

migrant mortality advantage found in developed countries. However, these explanations can also help discuss why migrants experience worse health and mortality outcomes than non-migrants in developing countries such as Brazil. In the discussion section, we return to these points presented here in light of the results from this investigation.

Data and Methods

Data, definitions, and selected covariates

We defined migrants as individuals born abroad, regardless of country of birth. This group also includes individuals who acquired Brazilian citizenship or nationality after arrival. Deaths come from the Mortality Information System (SIM/Datasus) for 2009-2011, and the population comes from the 2010 Brazilian demographic census, disaggregated by age, sex, nativity, and education. We restricted our analysis to adults over 30 years of age to ensure that most individuals have already completed their education and exclude younger adults for whom there is a small number of deaths. To get more robust mortality estimates by education, we grouped the adult population into five age groups: 30-49, 50-59, 60-69, 70-79, and 80 years or older. The use of larger age groups, while reducing variability, also helps attenuate age misreporting.

To make the education variable compatible both in the 2010 census and in the SIM dataset, we grouped migrant and non-migrant populations into three distinct educational categories: low education (0 to 7 years of study); middle education (8 to 11 years of study); and high education (12 years or more). Ribeiro *et al*. (2021) applied the same categorization. Supplementary material provides a step-by-step algorithm in R used to transform the categorized variable of education into years of study in the census microdata.

The death records for individuals 30 years and older had 33% of missing data on education. We distributed the missing cases by a random hot deck imputation procedure, with age, sex, and nationality as the pairing observables. We found an anomaly of missing cases in the death records for the *nativity* variable. By crossing two distinct databases, we found that the anomaly is due to a filling error in the variable. We corrected it and redistributed the remaining missing cases according to age and sex. Annex 4 provides

detailed information on how we solved it. After completing the steps listed above to organize and correct the data, we divided deaths and person-years of exposure to estimate age-specific death rates (M_i) by sex, nativity, and educational attainment.

Negative binomial regression models for estimating mortality rates

We employed negative binomial regression models to calculate mortality rates for foreign-born and native-born populations. The underlying assumption of count models for estimating case fatality rates is the existence of an incidence rate at which the event occurs, which can be multiplied by an exposure time to obtain the number of events that occurred. Therefore, the model includes an additional component, called *offset*, responsible for controlling the number of deaths by the number of confirmed cases in each selected category. The term offset is included in the logarithmic scale to match with the linking function employed. Thus, the model is given by:

$$logE(Y) = 1\log(N) + \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

The response variable is the expected number of deaths, and X is a matrix of covariates (nativity, age group, sex, and education). Log(N) is a covariate with a fixed coefficient equal to one, representing the total population in each category. The advantage of modeling the number of deaths using a negative binomial function is that it introduces an additional parameter that gives the model more flexibility to deal with overdispersion (Roback & Legler, 2021)

We estimated seven different models' specification to test the potential association between migration status and mortality. The first model included only nativity. We gradually added the remaining covariates in the following six models – age, sex, education, and the interaction terms. The low number of deaths precludes us from adding all two and three-way interaction terms. Therefore, one methodological alternative to better measure the role of education in mortality differences is running separate models for each education level, while keeping the regression models parsimonious. Each model controled for age, sex, nativity, and one two-way interaction term for age and nativity. As a final step, we estimated separate models for each education category, controlling for age, sex, nativity, and an interaction term for age and nativity. We selected the model with

the best adjustment based on Akaike's Information Criterion (AIC). All analyses were performed using the R software (R Core Team, 2021).

Results

Table 1 shows the number of deaths and the total population exposed to the risk of death for each covariate. Migrant deaths represent 3.37% of the average 242,153 deaths in São Paulo between 2009 and 2011. They are highly concentrated among older adults. This result is consistent with the age composition of the migrants living in São Paulo. Also, about 70.78% of migrants have low education (0 – 7 years).

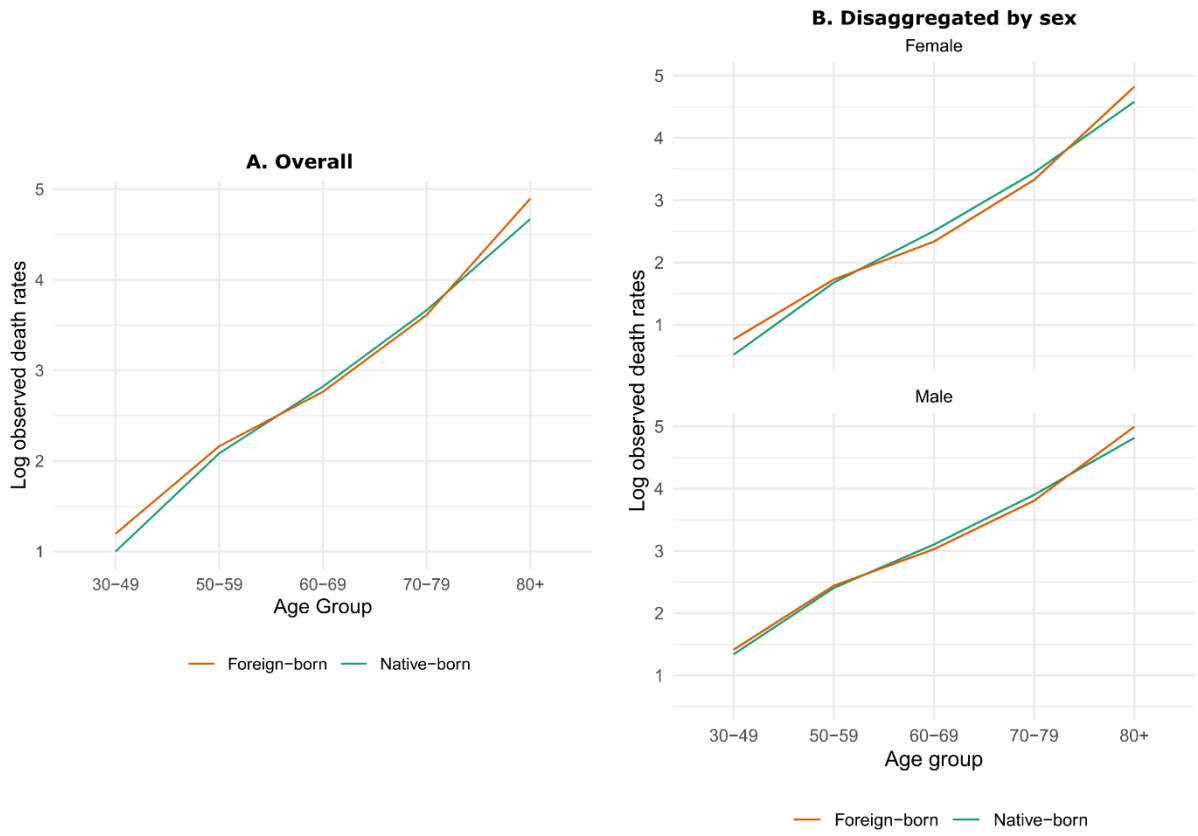
Table 1. Number of deaths and adult population by selected covariates – State of São Paulo, Brazil (2010)

Variable	Foreign-born			Native-born		
	Deaths	Population	CDR	Deaths	Population	CDR
Age group						
30-49	165	49,870	3.34	33,594	12,360,678	2.70
50-59	308	35,438	8.69	35,191	4,376,949	8.04
60-69	756	47,670	15.86	43,622	2,597,472	16.79
70-79	1,725	46,625	37.00	54,923	1,407,854	39.01
80+	5,217	39,043	133.62	66,652	624,702	106.69
Sex						
Female	3,877	104,648	37.05	105,284	11,333,396	9.29
Male	4,294	113,998	37.67	128,698	10,034,259	12.83
Educational attainment						
Low	5,784	81,371	71.08	179,761	9,640,950	18.65
Middle	1,439	71,271	20.19	35,612	7,823,433	4.55
High	948	66,004	14.36	18,609	3,903,272	4.77
Overall	**8,171**	**218,646**	**37.37**	**233,982**	**21,367,655**	**10.95**

Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

The age-specific mortality rates for migrants and non-migrants can help reveal existing differences by nativity. According to Figure 1, there is a migrant mortality disadvantage among adults in their 30s. This pattern reduces with age and even reverses to a slight mortality advantage in the 60-79 age group. However, among the oldest adults (80 years or older), the rates suggest once more a mortality disadvantage compared to the native-born living in São Paulo. The mortality rates by place of birth may not be statistically different, though.

Figure 1. Age-specific mortality rates (per thousand) by nativity and sex - State of São Paulo, Brazil (2010)



Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

Among foreign-borns, Europe (especially Portugal, Italy, Spain, and Germany) is the leading region of birth, accounting for 67.7% of deaths and more than 52.8% of the migrant population living in São Paulo in 2010 (Table 2). The Asian groups (notably Japanese) take the second position. Europeans and Asians combined account for 92.46% of total deaths. The results in Table 2 underscore the small number of deaths for specific birthplace regions (Americas, Africa, and Oceania) that limited our ability to consider further desegregations or interactions between this variable with age, sex, education in regression models.

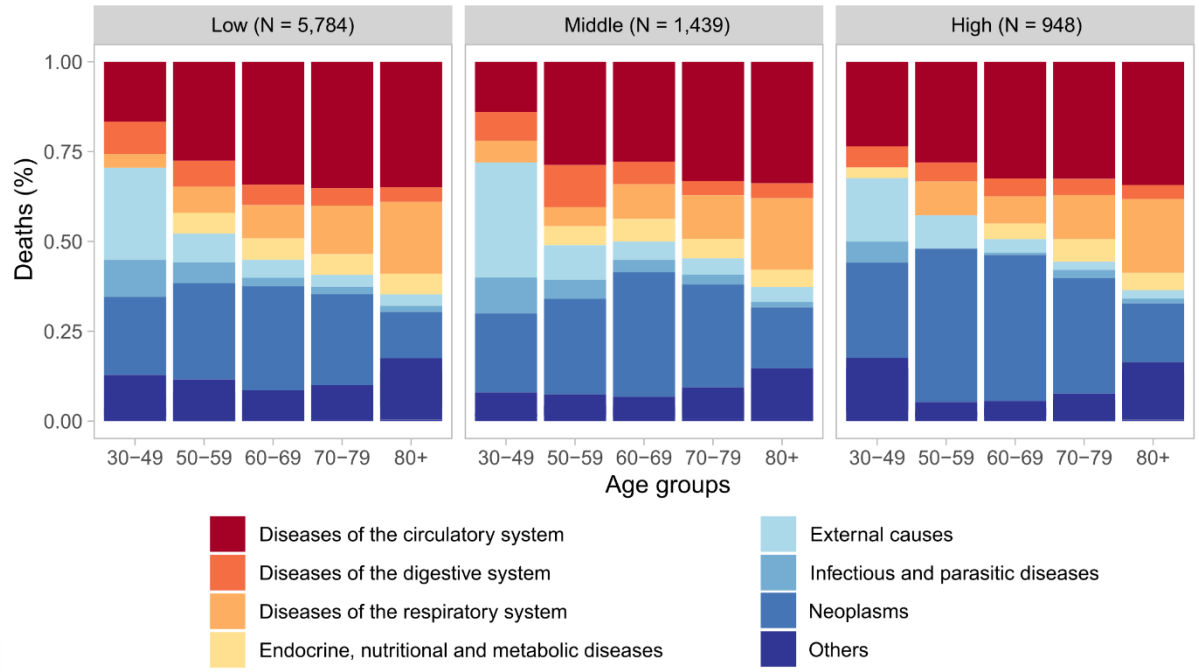
Table 2. Number of deaths and adult population by birthplace region – State of São Paulo, Brazil (2010)

Region of birth	Deaths		Population		CDR
	N	%	N	%	
Europe	5,536	67.75	115,797	52.80	47.81
Asia	2,019	24.71	52,934	24.13	38.14
America	447	5.47	41,612	18.97	10.74
South America	392	4.80	37,434	17.07	10.47
North America	30	0.37	2,684	1.22	11.18
Central America & Caribbean	25	0.31	1,494	0.68	16.73
Africa	142	1.74	6,088	2.78	23.32
Oceania	13	0.16	103	0.05	126.21
Unspecified	14	0.17	2,794	1.27	5.01
Total	**8,171**	**100**	**219,328**	**100**	**37.25**

Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

Figure 2 presents the leading causes of death for migrants by educational level. Mortality from external causes is more prominent among younger adult migrants with low and middle education. For the highly educated, the leading cause of death is neoplasms. These differences are consistent with the sociodemographic profiles of the migrant population in these age groups. Young adult migrants are primarily from South American countries, especially from Bolivia. The greater vulnerability of Bolivians to mortality from external causes in the city of São Paulo, which concentrates the largest proportion of migrants in the state, was also observed by Silveira (2016).

Figure 2. Leading causes of death of foreign-born adults by educational attainment - State of São Paulo, Brazil (2010)



Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

Table 3 presents the results from the regression models. We adopted the analytical strategy of gradually adding each covariate, starting from a simple model with only nativity and ending with a model that includes interactions between both age and nativity and education and nativity. In the first model, we considered only nativity as a covariate. Compared to the native-born, mortality for migrants are 19.6% higher (IRR = 1.196, C.I 95% [0.659; 2.171], p = 0.55). Model 2 includes age, a covariate related to mortality levels, and nativity status. As expected, all coefficients are significant at α=0.01, indicating that mortality rates differ across the age range; the older the individuals, the higher the mortality rate. Mortality levels remain higher for migrants than non-migrants.

In the third model, we included sex as covariate. As expected, there are significant differences in mortality rates between men and women (IRR =1.850, C.I 95% [1.458; 2.348], p < 0.001). Mortality differences by nativity are statistically significant, but only at the α =0.1 level. We also have evidence (likelihood ratio test, p <0.01; AIC = 885.4) that the inclusion of sex as a covariate provides a significant improvement in the previous model.

The fourth model includes education. Mortality rates differ across educational categories. Low education is associated with the highest mortality, confirming results from earlier research (Ribeiro et al., 2021). Also, the mortality disadvantage among migrants became highly significant when controlling for age, sex, and education. The likelihood ratio test (p-value < 0.01) and the AIC (814.79) provide evidence that the inclusion of education offered significant improvement in the previous model.

Model 5 includes an interaction term between age and nativity. Mortality differences by nativity are associated with specific age categories. The mortality disadvantage occurs in the 30-59 age groups and among individuals 80 years and older. However, as abovementioned, the difference reverses into a mortality advantage in the 60-79 age groups: mortality of migrants is 4% to 9% lower than those for Brazilians (Table 3).

The last two models examine whether mortality differences between foreign-born and native-born are also associated with specific education categories. Therefore, they include the interaction between nativity and education, in addition to the previous covariates. It suggests that, on average, the mortality of foreign-born individuals with middle education is not statistically different from that of native-born with low education. The advantage only shows up among migrants with higher education levels.

Table 3. Incidence Rate Ratios (IRR) from the negative binomial regression models by selected covariates

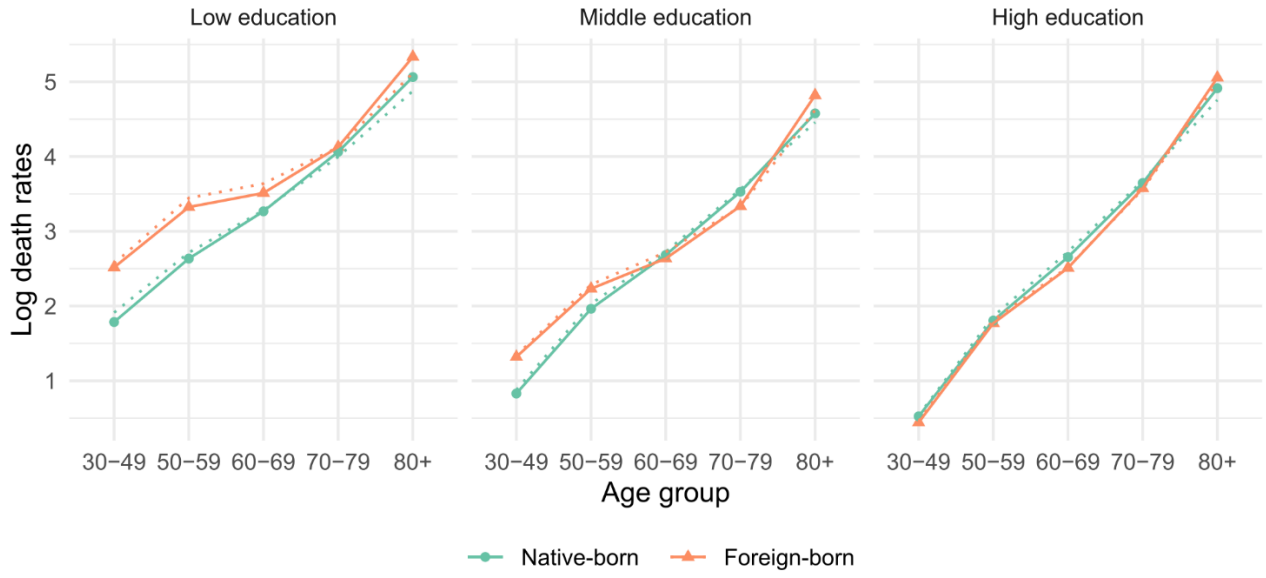
	Dependent variable: deaths						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Constant	0.031***	0.003***	0.002***	0.003***	0.003***	0.003***	0.003***
	[0.021; 0.049]	[0.002; 0.004]	[0.002; 0.003]	[0.002; 0.004]	[0.002; 0.003]	[0.002; 0.003]	[0.002; 0.003]
Nativity							
Native-born (ref)	1	1	1	1	1	1	1
Foreign-born	1.196	1.276*	1.261*	1.187***	1.577***	1.442***	1.826***
	[0.659; 2.171]	[0.962; 1.691]	[0.995; 1.600]	[1.043 1.350]	(1.175; 2.113)	(1.163; 1.789)	(1.325; 2.511)
Age group							
30-49 (ref)		1	1	1	1	1	1
50-59		2.614***	2.631***	2.814***	2.917***	2.799***	2.887***
		[1.662; 4.110]	[1.790; 3.866]	[2.274; 3.482]	[2.283; 3.727]	[2.274; 3.444]	[2.274; 3.666]
60-69		4.434***	4.519***	5.090***	6.084***	5.070***	5.994***
		[2.827; 6.950]	[3.086; 6.613]	[4.131; 6.269]	[4.760; 7.776]	[4.137; 6.212]	[4.719; 7.614]
70-79		9.833***	10.263***	11.871***	14.760***	11.842***	14.485***
		[6.272; 15.411]	[7.011; 15.017]	[9.638; 14.618]	[11.543; 18.873]	[9.664; 14.508]	[11.396; 18.413]
80+		31.940***	34.817***	41.741***	45.817***	41.771***	44.708***
		[20.393; 50.015]	[23.781; 50.969]	[33.857; 51.457]	[35.796; 58.644]	[34.054; 51.234]	[35.122; 56.910]
Sex							
Female (ref)			1	1	1	1	1
Male			1.850***	1.851***	1.843***	1.850***	1.845***
			[1.458; 2.348]	(1.629; 2.103)	(1.639; 2.072)	(1.633; 2.096)	(1.645; 2.069)
Educational attainment							
Low (ref)				1	1	1	1
Middle				0.436***	0.437***	0.487***	0.483***
				(0.373; 0.510)	(0.379; 0.505)	(0.398; 0.596)	(0.401; 0.581)
High				0.464***	0.466***	0.544***	0.535***
				(0.396; 0.543)	(0.404; 0.539)	(0.444; 0.666)	(0.444; 0.645)
Interaction term between nativity and age group							
Native-born*30-49 (ref)					1		1
Foreign-born*50-59					0.890		0.902
					[0.597; 1.327]		[0.610; 1.334]
Foreign-born*60-69					0.644**		0.658**
					[0.436; 0.952]		[0.449; 0.966]

Foreign-born*70-79					0.586***		0.606***
					[0.398; 0.865]		[0.414; 0.889]
Foreign-born*80+					0.773		0.815
					[0.527; 1.137]		[0.558; 1.192]
Interaction term between nativity and educational attainment							
Native-born*Low (ref)						1	1
Foreign-born*Middle						0.787	0.806
						[0.580; 1.069]	[0.609; 1.068]
Foreign-born*High						0.702**	0.734**
						[0.517; 0.953]	[0.552; 0.975]
Observations	60	60	60	60	60	60	60
Log likelihood	-499.039	-445.3349	-434.7217	-397.3983	-392.6545	-394.7501	-390.2719
Residual Deviance	71.87343	64.20341	64.3433	65.0709	65.2543	62.77488	63.33184
Res. Deg. of Freedom	58	54	53	51	47	49	45
Akaike Inf. Crit.	1004.0781	904.6698	885.4435	814.7967	813.3091	813.5003	812.5438

Note:*p<0.1; **p<0.05; ***p<0.01

We estimated separated regression models for each education group as an additional step. Each model controls for the nativity status, age, sex, and an interactive term between nativity and age. Figure 3 illustrates the age-specific mortality rates by nativity and educational level. The solid lines represent predicted mortality rates, while dotted lines correspond to observed mortality rates. The largest mortality difference is observed among younger and less educated adult migrants. For individuals with high education, the mortality rates for Brazilians and migrants are similar at all ages.

Figure 3. Predicted age-specific mortality rates by nativity and educational level - State of São Paulo, Brazil (2010)



Note: Solid lines represent predicted mortality rates while dotted lines correspond to observed mortality rates.
Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

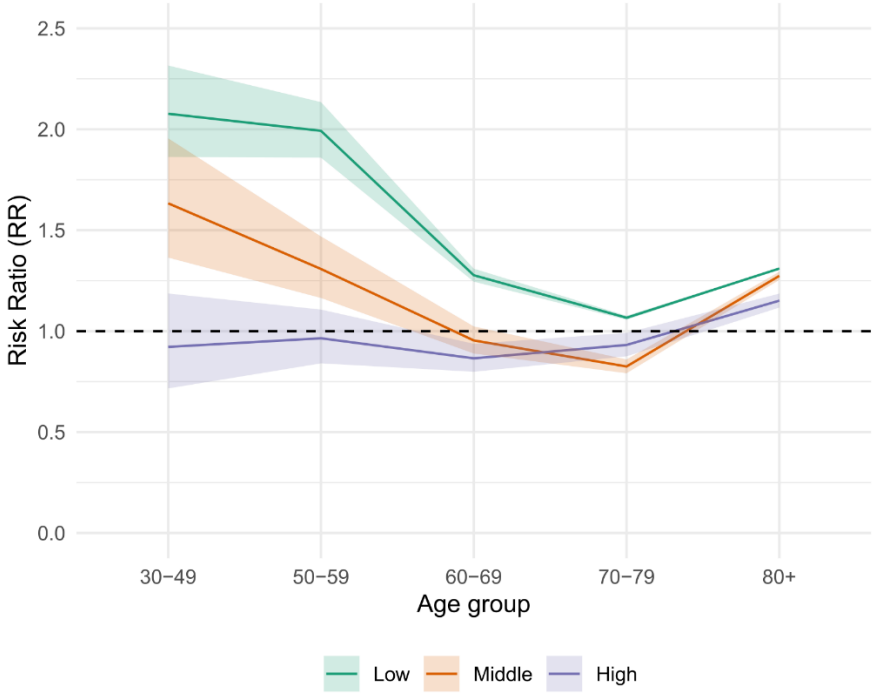
Table 4. Predicted death rates (per 1,000) for foreign-born and native-born populations by age, sex, and educational level – State of São Paulo, Brazil (2010)

Age Group	Foreign-born						Native-born					
	Males			Females			Males			Females		
	Death Rate	CI 95%		Death Rate	CI 95%		Death Rate	CI 95%		Death Rate	CI 95%	
Low Education												
30-49	12.39	9.38	16.36	7.08	5.33	9.41	5.97	5.04	7.07	3.41	2.88	4.04
50-59	27.77	21.89	35.23	15.88	12.48	20.20	13.94	11.77	16.51	7.97	6.73	9.44
60-69	33.51	27.60	40.69	19.16	15.75	23.30	26.24	22.16	31.07	15.00	12.67	17.76
70-79	62.01	51.90	74.08	35.44	29.64	42.38	58.15	49.12	68.85	33.24	28.08	39.35
80+	207.26	174.60	246.02	118.48	99.81	140.63	158.19	133.62	187.28	90.43	76.38	107.05
Middle Education												
30-49	3.74	2.78	5.05	1.86	1.37	2.53	2.29	2.04	2.58	1.14	1.01	1.29
50-59	9.32	7.36	11.79	4.64	3.65	5.89	7.12	6.32	8.03	3.55	3.15	4.00
60-69	13.96	11.55	16.88	6.95	5.72	8.44	14.64	12.99	16.50	7.29	6.46	8.22
70-79	28.12	23.94	33.04	14.00	11.85	16.54	34.10	30.25	38.43	16.97	15.05	19.14
80+	123.87	108.17	141.85	61.67	53.76	70.74	97.23	86.26	109.60	48.40	42.94	54.56
High Education												
30-49	1.56	1.08	2.24	0.93	0.65	1.35	1.69	1.51	1.89	1.01	0.91	1.13
50-59	5.87	4.58	7.52	3.52	2.74	4.53	6.09	5.45	6.80	3.65	3.27	4.08
60-69	12.31	10.17	14.89	7.38	6.06	8.99	14.22	12.73	15.89	8.53	7.63	9.54
70-79	35.70	30.03	42.44	21.41	17.86	25.67	38.32	34.29	42.82	22.98	20.55	25.71
80+	156.80	136.29	180.39	94.04	81.36	108.71	136.23	122.00	152.12	81.71	73.11	91.32

Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

Figure 4 presents the age-specific mortality ratios (foreign-born vs. native-born), with 95% confidence intervals based on the models by education. Ratios above one indicate excess of mortality for the foreign-born, while ratios below one represent a mortality advantage for them. The results suggest that the mortality disadvantage is primarily concentrated among adults with low education (all ages). It also exists among the middle-income group, but only in the younger age groups (30-59 years). Among the highly educated, mortality ratios are close to one. Some mortality advantage for the foreign-born was found for the middle-education group at ages 70 to 79 and among the higher-education group at ages 60-69.

Figure 4. Age-specific mortality ratios (foreign-born vs native-born) by educational level – State of São Paulo, Brazil (2010)



Note: 95% Confidence Intervals. Source: Mortality Information System (SIM). Brazilian Demographic census 2010.

Discussion and Conclusion

This paper examined mortality differences between foreign-born and native-born adults living in the Brazilian State of São Paulo. We looked at whether these differences are associated with specific age and education categories. The main findings of this study

confirm the central hypothesis that the mortality levels of the migrant population differ from those for Brazilians when controlled for age, sex, and education. We found higher levels of adult mortality between migrants when compared to Brazilian-borns. The mortality disadvantage among migrants persists after controlling for age, sex, and educational attainment. However, the mortality disadvantage depends on age, and it is concentrated in the 30-59 and 80 years and older age groups. Among 60 to 79 years old individuals, migrant mortality is 4% to 9% lower than Brazilians'. This age group comprises individuals born in European countries, such as Portugal, Italy, and Spain.

Migrant mortality disadvantage is highly concentrated among foreign-born adults with lower educational attainment. Individuals with higher levels of education present mortality rates very close to those of Brazilians. In some age groups, there is even a mortality advantage. This situation differs from a commonly observed pattern in developed countries. For instance, Turra and Goldman (2007) showed the mortality advantage among Hispanics in the United States is concentrated at lower levels of socioeconomic status. Unlike the United States, migrant mortality in São Paulo does not suggest the existence of an epidemiological paradox.

The mortality disadvantage among migrants in Brazil might be linked to a combination of underlying mechanisms that may operate in different directions. First, international immigration in contemporary Brazil occurred mainly in the first half of the XX century and comprised mostly Europeans (Italians, Germans, and Portuguese) and the Japanese. South Americans dominate the more recent wave of international migrants, especially Bolivians. Some specific studies carried out in the city of São Paulo have shown that younger Bolivians are more vulnerable in terms of mortality than Brazilians (Silveira, 2018). They also have a higher proportion of deaths from external causes than other nationalities. For this reason, the current flows of foreign-born with lower socioeconomic status may help explain some of the mortality disadvantages observed among younger adults.

The migration waves may impact our estimates in two different ways. First, it selects the migrants with better health from the first half of the XX century into the pool of the 2010 death records, reducing the gap for those 50 years and older. Second, the degree to which the health of migrants differs from their native counterparts can explain

part of the disadvantage. The healthy migrant hypothesis, one of the explanations for the mortality advantage of migrants in high-income countries, is more likely to operate in long-distance migration. As migrants from South American countries dominate current flows, it is unclear to what extent the healthy migrant hypothesis could be happening in São Paulo, given that the results presented here do not suggest any mortality advantage for these migrants over Brazilians.

Migrants are a selective group. Country of origin and duration of residence in the destination are important factors to consider. In this study, due to the low number of deaths by place of birth and the lack of a variable relating to the length of residence, we could not control the estimates for these two variables. However, as more than 90% of migrant mortality is concentrated among the traditional nationalities of migrants in São Paulo (European and Japanese), the composition effect of the length of residence and birthplace country is less aggravating and does not discredit the estimates.

We acknowledge that certain data artifacts may influence our findings. We acknowledge that certain data artifacts may influence our findings. In particular, there may be some subestimation of undocumented migrants with low education and their preference for not responding to the census. However, our simulations and consistency checks showed that migrant mortality disadvantage only would disappear in a hypothetical scenario of a migrant under-enumeration on the order of 108%, which is completely unrealistic for the quality standards of population coverage of the Brazilian census in 2010 (IBGE, 2016). This situation reinforces the reliability of our findings and opens new possibilities for investigating the migrant-native mortality gap in future works, considering other disaggregations, such as country of birth and causes of death.

It is noteworthy that the form of data collection and the need to impute missing information in the variable relating to birth and education in SIM makes methodological issues remain central in the analysis of mortality differences, generating uncertainties about the proper patterns of mortality in the population in question. This information is essential for the variable birthplace country, which presents an anomaly of missing cases in the selected period and depends on the assumption that the percentage of missing cases in the State is equal to that observed in the city of São Paulo.

Analysis of all-cause mortality may also hide essential differences between migrants and non-migrants. A recent cohort study in Denmark, for instance, found no evidence supporting the healthy migrant hypothesis when looking for cause-specific mortality (Norredam et al., 2014). In our case, the migration mortality disadvantage in the 30 to 49 age range is likely to be affected by higher risks of death from external causes among migrants. A cause-specific analysis could yield very different results.

Overall, this paper provides evidence of mortality disparities between migrants and non-migrants in middle-income countries. Results from this paper are of interest not just to Brazil but to a much broader audience that seeks examples of how the health and mortality of migrants differ from native-borns. Through reliable evidence of mortality, policymakers will be able to ensure that the most vulnerable individuals in terms of health have adequate access to health services in the State with the highest concentration of migrants in Brazil.

Concluding Remarks

This dissertation dealt with migration-related issues for adults living in São Paulo, Brazil. The core interest of this research was to access the diversity of sociodemographic profiles of migrants living in the main migrant hub in Brazil and examine how mortality patterns of this population differed from those observed for Brazilian-borns residing in São Paulo in 2010.

The first paper investigated the diversity of sociodemographic characteristics of adult migrants living in São Paulo. We relied on data from the 2010 Brazilian demographic census and applied the Grade of Membership (GoM) technique to delineate the sociodemographic profiles of adult migrants. We also analyzed how long migrants have resided in São Paulo after their arrival, their nationality, and changes in the size and composition of migrant populations over the decades. We found four different profiles of foreign-born adults: recent, middle-term, long-term, and old migrants. These four profiles showed how heterogeneous migrant populations are in terms of sociodemographic characteristics. Such diversity pervades the dimensions of age, gender, socioeconomic status, and birthplace region

The second paper examined mortality differences between foreign-born and native-born adults living in the Brazilian State of São Paulo. We looked at whether these differences are associated with specific age and education categories. The main findings confirm the hypothesis that mortality levels of foreign-born differ from those observed among Brazilian-borns living in São Paulo when controlling for age, sex, and education. Unlike the commonly observed pattern in developed countries, migrants have a mortality disadvantage compared to non-migrants. Therefore, there is insufficient evidence to support the existence of an epidemiological paradox in São Paulo. The results also revealed that mortality differences are associated with specific age categories and are concentrated among younger foreign-born adults with low educational attainment.

This study represents an effort to expand the international debate on the mortality differentials of migrants and non-migrants with evidence from regions in developing countries such as Brazil. It also opens new opportunities to further investigations to understand of the socioeconomic and health factors associated with the migration process. An essential point that deserves more attention is to analyze how working

conditions affect the health of migrant populations living in São Paulo, especially among recent foreign-born adults. It is also an important gap in the current research. Second, the 2010s witnessed important social and structural changes that may have had direct and indirect impacts on the sociodemographic profile of the migrant population. The combination of the results of this study, with more recent data from the Brazilian demographic census of 2022 could provide important information on sociodemographic profiles and mortality differentials of foreign-born and native-born adults in future works.

In conclusion, this dissertation provides evidence on the sociodemographic profiles and mortality patterns of foreign-born adults, especially the socioeconomic and health factors associated with migration. Evidence from this work can be used to enhance policy coherence, help evidence-based policymaking on migration, and improve the mainstreaming of migration into development plans in the most populous State in South America.

References

Abraído-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health*, *89*(10), 1543–1548. https://doi.org/10.2105/AJPH.89.10.1543

Abubakar, I., Aldridge, R. W., Devakumar, D., Orcutt, M., Burns, R., Barreto, M. L., … Zhou, S. (2018). The UCL–Lancet Commission on Migration and Health: the health of a world on the move. *The Lancet*, *392*(10164), 2606–2654. https://doi.org/10.1016/S0140-6736(18)32114-7

Aldridge, R. W., Nellums, L. B., Bartlett, S., Barr, A. L., Patel, P., Burns, R., … Abubakar, I. (2018). Global patterns of mortality in international migrants: a systematic review and meta-analysis. *The Lancet*, *392*(10164), 2553–2566. https://doi.org/10.1016/S0140-6736(18)32781-8

Amaral, E. F., & Fusco, W. (2005). Shaping Brazil: The Role of International Migration. *Migration Policy Institute*, (June).

Andrade, J., & Guedes, G. (2021). *gom: Grade of Membership Mixture Models*. Belo Horizonte, Brazil.

Anson, J. (2004). The migrant mortality advantage: A 70 month follow-up of the Brussels population. *European Journal of Population*, *20*(3), 191–218. https://doi.org/10.1007/s10680-004-0883-1

Baeninger, R. (2012). *Faces e fases da migração em São Paulo* (1st ed.). São Paulo: Observatório das Migrações em São Paulo.

Brito, F. (1995). Os povos em movimento: as migrações internacionais no desenvolvimento do capitalismo. In N. L. Patarra (Ed.), *Emigração e Imigração Internacionais no Brasil contemporâneo* (2ª). São Paulo: FNUAP.

Brito, F. (2004). Crescimento demográfico e migrações na transição para o trabalho assalariado no Brasil. *Revista Brasileira de Estudos de População*, *21*(1), 5–20.

Caetano, A. J., & Machado, C. J. (2009). Consistência e identificabilidade no modelo Grade of Membership: Uma nota metodológica. *Revista Brasileira de Estudos de Populacao*, *26*(1), 145–149. https://doi.org/10.1590/S0102-30982009000100012

Cano, W. (1988). *A interiorização do desenvolvimento econômico no estado de São Paulo (1920-1980)* (Coleção Ec, Vol. 1). São Paulo: Fundação Seade.

Carvalho, J. A. M. de. (1996). O saldo dos fluxos migratórios internacionais no Brasil na década de 80 - uma tentativa de estimação. *Revista Brasileira de Estudos Populacionais*, *13*(1), 12.

Carvalho, J. A. M. de, Campos, J., Rigotti, J. I. R., & Pinho, B. A. T. D. de. (2018). Migrações internacionais no Brasil no período 2005-2010, com ênfase nos efeitos diretos e indiretos da imigração de retorno dos brasileiros. *Revista Brasileira de Estudos de População*, *35*(3), 1–26. https://doi.org/10.20947/s0102-3098a0057

Castles, S., Haas, H. de, & Miller, M. J. (2015). The Age of Migration: International Population Movements in the Modern World. In S. Castles (Ed.), *Palgrave Macmillan* (5ª). United Kingdom: Palgrave Macmillan.

Chiswick, B. R., Lee, Y. Li., & Miller, P. W. (2008). Immigrant selection systems and immigrant health. *Contemporary Economic Policy*, *26*(4), 555–578. https://doi.org/10.1111/j.1465-7287.2008.00099.x

De Maria, P. F., & Baeninger, R. (2016). Imigração em São Paulo: perfis segundo pobreza e ocupação no século XXI. *Geografias*, *1*, 141–160. https://doi.org/10.5151/socsci-ix-enm-st3-2

DeBoosere, P., & Gadeyne, S. (2005). Adult migrant mortality advantage in Belgium: Evidence using census and register data. *Population*, *60*(5–6), 655–698. https://doi.org/10.3917/pope.505.0655

Domeniconi, J. de O. S., & Baeninger, R. (2017). A dinâmica da migração internacional qualificada para o estado de São Paulo no século XXI: os espaços da migração dos "trabalhadores do conhecimento." *Cadernos Metrópole*, *19*(40), 749–775. https://doi.org/10.1590/2236-9996.2017-4003

Dutra, D., Almeida, S. de, Tonhati, T., & Palermo, G. (2015). Os estrangeiros no mercado de trabalho formal brasileiro: Perfil geral na série 2011, 2012 e 2013. In L. Cavalcanti, A. T. R. Oliveira, & T. Tonhati (Eds.), *A inserção dos imigrantes no mercado de trabalho brasileiro* (1st ed., pp. 74–135). Brasília: Observatório das Migrações Internacionais.

Gimeno-Feliu, L. A., Calderón-Larrañaga, A., Díaz, E., Laguna-Berna, C., Poblador-Plou, B., Coscollar-Santaliestra, C., & Prados-Torres, A. (2019). The definition of immigrant status matters: Impact of nationality, country of origin, and length of stay in host country on mortality estimates. *BMC Public Health*, *19*(1), 1–8. https://doi.org/10.1186/s12889-019-6555-1

Graham, D. H. (1973). Migração estrangeira e a questão da oferta de mão de obra no crescimento econômico brasileiro - 1880-1930. *Estudos Econômicos*, *3*(1), 7–64.

Guedes, G. R., Siviero, P. C. L., Machado, C. J., Pinto, J., & Rodarte, M. M. S. (2016). *Grade of Membership: conceitos básicos e aplicação empírica usando o programa GoM para Windows, Linux, Stata e R* (1st ed.). Belo Horizonte: Editora UFMG.

Guillot, M., Khlat, M., Elo, I., Solignac, M., & Wallace, M. (2018). Understanding age variations in the migrant mortality advantage: An international comparative perspective. *PLoS ONE*, *13*(6), 1–17. https://doi.org/10.1371/journal.pone.0199669

Harding, S. (2003). Mortality of Migrants from the Indian Subcontinent to England and Wales: Effect of Duration of Residence. *Epidemiology*, *14*(3), 287–292. https://doi.org/10.1097/01.EDE.0000050369.88416.82

Hedlund, E., Pehrsson, K., Lange, A., & Hammar, N. (2008). Country of birth and survival after a first myocardial infarction in Stockholm, Sweden. *European Journal of Epidemiology*, *23*(5), 341–347. https://doi.org/10.1007/s10654-008-9240-8

Hermalin, A. I., Ofstedal, M. B., Sun, C., & Liu, I.-W. (2009). Nativity Differentials in Older Age Mortality in Taiwan: Do They Exist and Why? *Ren Kou Xue Kan / Kuo Li Tai-Wan Ta Hsueh*, *39*(1), 1–58. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf

Hoffmann, H. (1977). Desemprego e subemprego no Brasil. In *Séries: Ensaios* (2nd ed., Vol. 24). São Paulo: Ática.

Holanda, S. B. de. (1995). *Raízes do Brasil* (26th ed.). São Paulo: Companhia das Letras.

IBGE. (2016). *Metodologia do censo demográfico 2010* (2nd ed.). Brasília: Instituto Brasileiro de Geografia e Estatística - IBGE.

International Organization for Migration (IOM). (2017). International Migration Statistics. *Global Compact Thematic Paper*, (September 2016), 6.

Kohls, M. (2010). Selection, social status or data artefact - What determines the mortality of migrants in Germany? *Demographic Aspects of Migration*, 153–177. https://doi.org/10.1007/978-3-531-92563-9_6

Koser, K. (2010). Introduction: International Migration and Global Governance. *Global Governance: A Review of Multilateralism and International Organizations*, *16*(3), 301–315.

Levy, M. S. (1974). O papel da migração internacional na evolução da população brasileira (1872 a 1972). *Revista de Saude Pública*, *8*, 49–90. https://doi.org/10.1590/S0034-

89101974000500003

Livi-Bacci, M. (2011). International migration: a global force, ungoverned. *Demografia Em Debate*, *4*, 51–58.

Magalhães, L. F. A., Bógus, L. M. M., & Baeninger, R. (2018). Migrantes haitianos e bolivianos na cidade de São Paulo: transformações econômicas e territorialidades migrantes. *REMHU, Rev. Interdiscip. Mobil. Hum*, 75–94.

Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). Statistical applications using fuzzy sets. In *Computer Science* (1st ed., Vol. 51). https://doi.org/10.2307/2532968

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1998). Worlds in Motion: understanding International Migration at the End of the Millennium. In *International Studies in Demography* (1st ed.). United Kingdom: Claredon Press.

Nepomuceno, M. R. (2017). *Overcoming the Limitations of Demographic Data: Papers on Mortality, the Extreme Aged and Education*. Federal University of Minas Gerais.

Norredam, M., Agyemang, C., Hoejbjerg Hansen, O. K., Petersen, J. H., Byberg, S., Krasnik, A., & Kunst, A. E. (2014). Duration of residence and disease occurrence among refugees and family reunited immigrants: Test of the "healthy migrant effect" hypothesis. *Tropical Medicine and International Health*, *19*(8), 958–967. https://doi.org/10.1111/tmi.12340

Norredam, M., Olsbjerg, M., Petersen, J. H., Juel, K., & Krasnik, A. (2012). Inequalities in mortality among refugees and immigrants compared to native Danes - A historical prospective cohort study. *BMC Public Health*, *12*(1). https://doi.org/10.1186/1471-2458-12-757

OIM. (2009). *Perfil Migratório do Brasil* (1st ed.). São Paulo: Organização Internacional para as Migrações (OIM).

Oliveira, A. T. (2013). Um panorama da migração internacional a partir do Censo Demográfico de 2010. *REMHU : Revista Interdisciplinar Da Mobilidade Humana*, *21*(40), 195–210. https://doi.org/10.1590/S1980-85852013000100012

Oliveira, A. T. (2015). O perfil geral dos imigrantes no Brasil a partir dos censos demográficos 2000 e 2010. In L. Cavalcanti, A. T. Oliveira, & T. Tonhati (Eds.), *A inserção dos imigrantes no mercado de trabalho brasileiro* (Especial, pp. 48–73). Brasília: Observatório das Migrações Internacionais.

Palloni, A., & Arias, E. (2004). Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography*, *41*(3), 385–415. https://doi.org/10.1353/dem.2004.0024

Patarra, N. L. (2003). Movimentos migratórios no Brasil: tempos e espaços. *Escola Nacional de Ciências Estatísticas*, (7), 44.

Patarra, N. L., & Fernandes, D. (2011). Brasil: país de imigração? *Revista Internacional Em Língua Portuguesa*, *24*, 65–97.

Preston, S. H., Elo, I. T., & Stewart, Q. (1999). Effects of Age Misreporting on Mortality Estimates at Older Ages. *Population Studies*, *53*(2), 165–177.

Queiroz, B. L., Freire, F. H. M. de A., Gonzaga, M. R., & Lima, E. E. C. de. (2017). Estimativas do grau de cobertura e da mortalidade adulta (45q15) para as unidades da federação no Brasil entre 1980 e 2010. *Revista Brasileira de Epidemiologia*, *20*, 21–33. https://doi.org/10.1590/1980-5497201700050003

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Ribeiro, M. M., Turra, C. M., & Pinto, C. C. de X. (2021). Mortalidade adulta por nível de escolaridade em São Paulo: análise comparativa a partir de diferentes estratégias metodológicas. *Revista Brasileira de Estudos de População*, *38*(21), 1–28. https://doi.org/10.20947/s0102-3098a0139

Riosmena, F., Wong, R., & Palloni, A. (2013). Migration Selection, Protection, and Acculturation in Health: A Binational Perspective on Older Adults. *Demography*, *50*(3), 1039–1064. https://doi.org/10.1007/s13524-012-0178-9

Roback, P., & Legler, J. (2021). Poisson Regression. In *Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R*. New York: Chapman and Hall/CRC Press.

Rodriguez, F., Hastings, K. G., Hu, J., Lopez, L., Cullen, M., Harrington, R. A., & Palaniappan, L. P. (2017). Nativity status and cardiovascular disease mortality among Hispanic adults. *Journal of the American Heart Association*, *6*(12), 1–9. https://doi.org/10.1161/JAHA.117.007207

Sawyer, D. O., Leite, I. da C., & Alexandrino, R. (2002). Perfis de utilização de serviços de saúde no Brasil. *Ciência & Saúde Coletiva*, *7*(4), 757–776. https://doi.org/10.1590/s1413-81232002000400012

Shor, E., & Roelfs, D. (2021). A Global Meta-analysis of the Immigrant Mortality Advantage. *International Migration Review*, *XX*(X), 1–30. https://doi.org/10.1177/0197918321996347

Silveira, R. C. (2018). Óbitos de imigrantes residentes no município de São Paulo nos anos de 2006 a 2015. In R. Baeninger (Ed.), *Migrações Sul-Sul* (2nd ed., pp. 558–583). Campinas: Núcleo de Estudos de População “Elza Berquó” – Nepo/Unicamp.

Spallek, J., Zeeb, H., & Razum, O. (2011). What do we have to know from migrants’ past exposures to understand their health status? A life course approach. *Emerging Themes in Epidemiology*, *8*, 1–8. https://doi.org/10.1186/1742-7622-8-6

Turra, C. M., & Goldman, N. (2007). Socioeconomic Differences in Mortality Among U. S. Adults: Insights Into the Hispanic Paradox. *The Journals of Gerontology*, *62*(3), 184–192. https://doi.org/10.1093/geronb/62.3.S184

United Nations. (2019). Resolution adopted by the General Assembly for Global Compact for Safe, Orderly and Regular Migration. *A/Res/73/195*, Vol. 22354, pp. 1–36.

United Nations. (2020). International Migration 2020: Highlights. In Department of Economic and Social Affairs. Population Division (Ed.), *International Migration 2020: Highlights* (1st ed.). https://doi.org/10.18356/9789210052689

Vang, Z. M., Sigouin, J., Flenon, A., & Gagnon, A. (2017). Are immigrants healthier than native-born Canadians? A systematic review of the healthy immigrant effect in Canada. *Ethnicity and Health*, *22*(3), 209–241. https://doi.org/10.1080/13557858.2016.1246518

Vieira, M. G. F. (2016). *Desigualdades intraurbanas na Área Metropolitana de Brasília: uma análise da mortalidade segundo níveis de educação*. Universidade de Brasília.

Wallace, M., & Darlington-Pollock, F. (2020). Poor health, low mortality? Paradox found among immigrants in England and Wales. *Population, Space and Place*, (June), 1–15. https://doi.org/10.1002/psp.2360

Wallace, M., Khlat, M., & Guillot, M. (2019). Mortality advantage among migrants according to duration of stay in France, 2004-2014. *BMC Public Health*, *19*(1), 1–9. https://doi.org/10.1186/s12889-019-6652-1

Wallace, M., & Kulu, H. (2014). Low immigrant mortality in England and Wales: A data artefact? *Social Science and Medicine*, *120*, 100–109. https://doi.org/10.1016/j.socscimed.2014.08.032

Wallace, M., & Wilson, B. (2021). Age variations and population over-coverage: Is low mortality among migrants merely a data artefact? *Population Studies*, *0*(0), 1–18. https://doi.org/10.1080/00324728.2021.1877331

Wild, S. H., Fischbacher, C., Brock, A., Griffiths, C., & Bhopal, R. (2007). Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. *Journal of Public Health*, *29*(2), 191–198. https://doi.org/10.1093/pubmed/fdm010

Woodbury, M. A., Clive, J., & Garson, A. (1978). Mathematical typology: A grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*, *11*(3), 277–298. https://doi.org/10.1016/0010-4809(78)90012-5

Young, C. M. (1987). Migration and Mortality: the experience of birthplace groups in Australia. *International Migration Review*, *21*(3), 531–554.

Zapata, G. P., & Guedes, G. (2017). Nota Editoral Temática: Refúgio e modalidades de deslocamentos populacionais no século XXI: Tendências, conflitos e políticas. *Revista Brasileira de Estudos de Populacao*, *34*(1), 5–13. https://doi.org/10.20947/s0102-3098a0018

Annexes

Annex 1. Descriptive statistics of adult migrants' sociodemographic characteristics, estimated probabilities, and Lambda-Marginal Frequency Ratio – State of São Paulo (2010)

Variable	Observed Frequency		Estimated probabilities (λ)				Lambda-Marginal Frequency Ratio (LMFR)			
	N	%	K1	K2	K3	K4	K1/%	K2/%	K3/%	K4/%
Age										
30-39	25,418	11.59	0.61	0.00	0.00	0.00	**5.30**	0.00	0.00	0.00
40-49	24,640	11.24	0.39	0.16	0.00	0.00	**3.44**	**1.41**	0.00	0.00
50-59	35,513	16.19	0.00	0.63	0.00	0.00	0.00	**3.89**	0.00	0.00
60-69	47,861	21.82	0.00	0.21	0.43	0.00	0.00	0.97	**1.98**	0.00
70-79	46,721	21.30	0.00	0.00	0.57	0.00	0.00	0.00	**2.66**	0.00
80+	39,172	17.86	0.00	0.00	0.00	1.00	0.00	0.00	0.00	**5.60**
Gender										
Male	114,342	52.13	0.53	0.79	0.58	0.00	1.02	**1.51**	1.11	0.01
Female	104,984	47.87	0.47	0.21	0.42	1.00	0.98	0.45	0.89	**2.08**
Birthplace country										
Portugal	62,736	28.60	0.00	0.00	0.65	0.13	0.00	0.00	**2.26**	0.46
Japan	28,239	12.88	0.00	0.00	0.00	0.79	0.00	0.00	0.00	**6.11**
Bolivia	11,749	5.36	0.34	0.00	0.00	0.00	**6.34**	0.00	0.00	0.00
Italy	18,718	8.54	0.00	0.00	0.23	0.00	0.00	0.00	**2.64**	0.00
Spain	16,096	7.34	0.00	0.04	0.13	0.08	0.00	0.50	**1.74**	1.12
China	9,251	4.22	0.26	0.00	0.00	0.00	**6.26**	0.00	0.00	0.00
Chile	7,995	3.65	0.00	0.13	0.00	0.00	0.00	**3.63**	0.00	0.00
Argentina	6,870	3.13	0.00	0.11	0.00	0.00	0.00	**3.63**	0.00	0.00
Korea, RO (South)	6,288	2.87	0.18	0.00	0.00	0.00	**6.16**	0.00	0.00	0.00
Other	51,384	23.43	0.22	0.72	0.00	0.00	0.94	**3.06**	0.00	0.00
Education										
Low	53,801	24.53	0.00	0.00	0.00	1.00	0.00	0.00	0.00	**4.08**
Middle	99,523	45.38	1.00	0.00	1.00	0.00	**2.20**	0.00	**2.20**	0.00
High	66,002	30.09	0.00	1.00	0.00	0.00	0.00	**3.32**	0.00	0.00
Race/Color										
White	159,180	72.58	0.00	1.00	1.00	0.38	0.00	**1.38**	**1.38**	0.52
Black	2,670	1.22	0.07	0.00	0.00	0.00	**6.05**	0.00	0.00	0.00
Indigenous	1,283	0.59	0.04	0.00	0.00	0.00	**6.02**	0.00	0.00	0.00
Asian	42,449	19.35	0.49	0.00	0.00	0.62	**2.56**	0.00	0.00	**3.23**
Brown (*pardo*)	13,743	6.27	0.40	0.00	0.00	0.00	**6.32**	0.00	0.00	0.00
Marital Status										
Single	13,037	5.94	0.33	0.00	0.00	0.00	**5.51**	0.00	0.00	0.00
Married	144,217	65.76	0.55	0.84	0.92	0.00	0.84	**1.27**	**1.40**	0.00
Divorced	20,660	9.42	0.12	0.16	0.08	0.00	**1.30**	**1.75**	0.81	0.00
Widowed	41,412	18.88	0.00	0.00	0.00	1.00	0.00	0.00	0.00	**5.30**
Length of residence (years)										
0-10	26,552	12.11	0.55	0.00	0.00	0.00	**4.51**	0.00	0.00	0.00
11-20	14,183	6.47	0.32	0.00	0.00	0.00	**5.01**	0.00	0.00	0.00
21-30	15,197	6.93	0.13	0.18	0.00	0.00	**1.89**	**2.62**	0.00	0.00
31-40	25,777	11.75	0.00	0.59	0.00	0.00	0.00	**5.02**	0.00	0.00
41-50	33,662	15.35	0.00	0.23	0.26	0.00	0.00	**1.49**	**1.70**	0.00
50+	103,956	47.40	0.00	0.00	0.74	1.00	0.00	0.00	**1.56**	**2.11**
Employment status										
Employed	96,012	43.78	0.89	1.00	0.00	0.00	**2.04**	**2.28**	0.00	0.00

Variable		Observed Frequency		Estimated probabilities (λ)				Lambda-Marginal Frequency Ratio (LMFR)			
		N	%	K1	K2	K3	K4	K1/%	K2/%	K3/%	K4/%
	Unemployed	3,838	1.75	0.11	0.00	0.00	0.00	**6.03**	0.00	0.00	0.00
	Inactive	119,476	54.47	0.00	0.00	1.00	1.00	0.00	0.00	**1.84**	**1.84**
Income											
	1st quintile	44,178	20.14	0.30	0.00	0.44	0.00	**1.49**	0.00	**2.18**	0.00
	2st quintile	40,105	18.29	0.14	0.00	0.00	0.81	0.75	0.00	0.00	**4.44**
	3st quintile	47,436	21.63	0.41	0.00	0.32	0.19	**1.90**	0.00	**1.50**	0.87
	4st quintile	39,422	17.97	0.15	0.25	0.24	0.00	0.84	**1.39**	**1.32**	0.00
	5st quintile	48,186	21.97	0.00	0.75	0.00	0.00	0.00	**3.41**	0.00	0.00

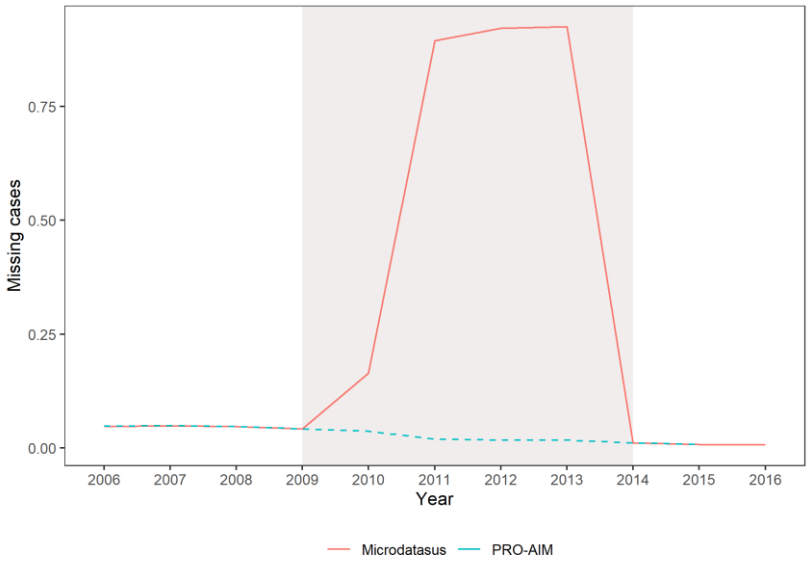
Annex 2

Correcting missing cases in nativity variable

The mortality database needs special attention. An anomaly was identified in the variable nativity between 2009 and 2014 in the SIM/Datasus microdata. The reason for the staggering total of missing cases is still unknown. Figure 6 illustrates this situation. I identified this anomaly by crossing the SIM data with the death database provided by the Mortality Information Improvement Program (PRO-AIM) of the city of São Paulo. PRO-AIM currently manages deaths information in the municipality, carrying the position of municipal manager of SIM/Datasus.

The strategy to analyze the behavior of missing cases in the nationality variable was comparing the total number of deaths between the two different databases between 2006 and 2016. As the PRO-AIM database is exclusively for deaths in the city of São Paulo, I also filtered this municipality in the SIM database for comparison. These values are expected to converge. This situation occurs before 2009 and after 2014. Among these years, the number of missing cases reached its maximum in 2013 with 93% missing cases. This situation, if not resolved, would make the analysis completely unfeasible.

Figure A2. Comparison of the percentage of missing cases of the variable birthplace between the SIM/Datasus and PRO-AIM database



Source: Mortality Information System (SIM/Datasus). Mortality Information Improvement Program (PRO-AIM). Health Department of the city of São Paulo (2006-2015).

It is possible to access and compare the total number of deaths only for the municipality of São Paulo from these two databases. The only viable way to overcome the anomaly of filling in the state database was to assume that the total number of missing cases in the municipality is the same for the State. This assumption draws upon two main arguments. The first one is that missing cases refers mainly to the deaths of Brazilians, which are known through the comparison between the two databases (Table 2). The second reason is that international migrants are highly concentrated in the municipality of São Paulo. In summary, in this work, I will admit that the pattern of missing cases observed in the São Paulo database is the same for the datasus database.

Table A2. Comparison of total deaths in the city of São Paulo by country of birth between two different databases

Country of birth	Deaths				Difference
	DATASUS		PRO-AIM (SMS)		
	N	%	N	%	
Brasil	38,987	55.5	62,854	89.5	- 23,867
NA	**26,201**	**37.3**	**2,295**	**3.3**	**23,906**
Portugal	1,590	2.3	1,589	2.3	1
Other countries	1,345	1.9	1,346	1.9	- 1
Japan	719	1.0	718	1.0	1
Italy	541	0.8	541	0.8	-
Spain	386	0.5	385	0.5	1
Germany	110	0.2	110	0.2	-
Bolivia	80	0.1	80	0.1	-
Argentina	78	0.1	78	0.1	-
China	78	0.1	78	0.1	-
Lebanon	72	0.1	72	0.1	-
Chile	33	0.0	33	0.0	-
Uruguay	23	0.0	23	0.0	-
Paraguay	18	0.0	18	0.0	-
Peru	13	0.0	13	0.0	-
Total	**70,274**	**100.0**	**70,233**	**100.0**	**41**

Source: Mortality Information System (SIM/Datasus). Mortality Information Improvement Program (PRO-AIM). Health Department of the city of São Paulo (2006-2015).

After solving the anomaly in filling the country of birth variable, the 3.3% deaths of unknown nationality were distributed among the other nationalities. In the specific case of other variables such as education, for example, to deal with the percentage of missing information in the SIM, we performed the same imputation procedures used by Ribeiro, Turra and Pinto (2021). This procedure considers age and sex, which indicates that deaths were randomly redistributed by these attributes.

Supplementary material

SM.1 *Creating a standard variable for education from the Brazilian Demographic Census (2010) and SIM/Datasus databases*

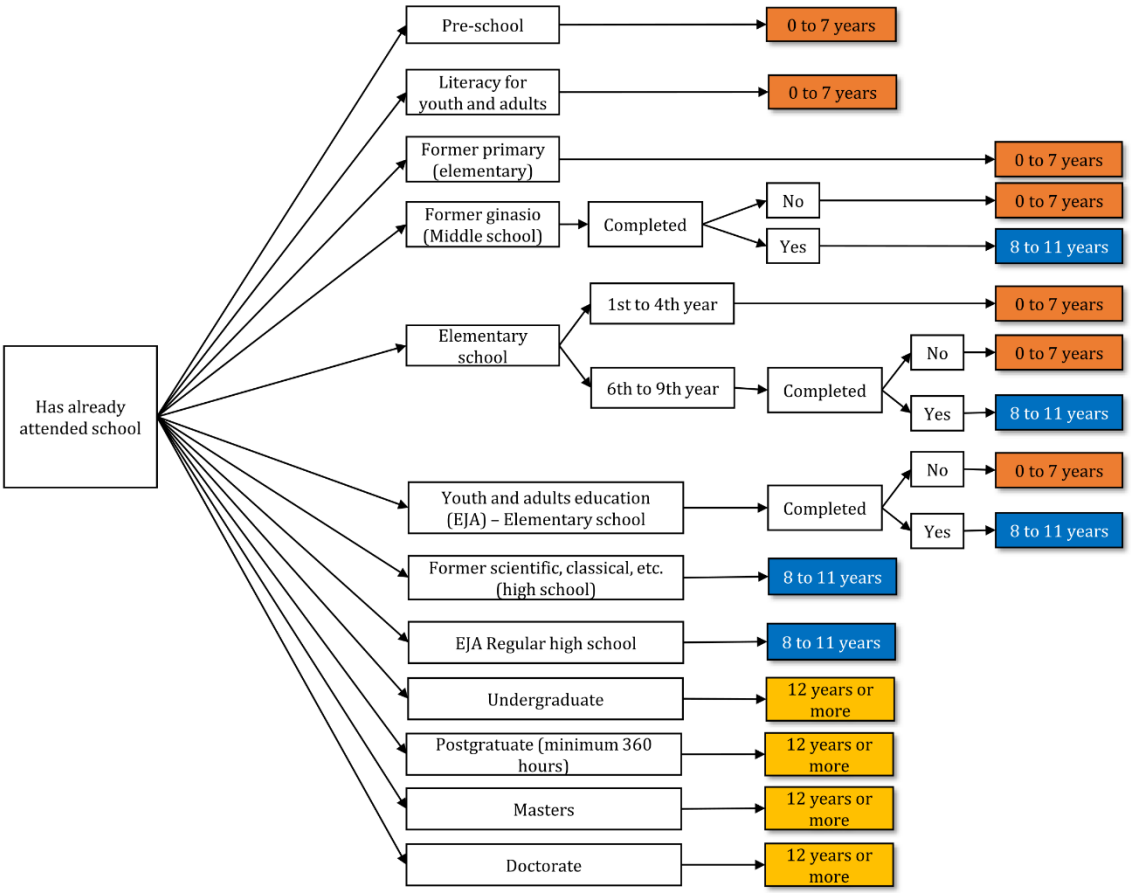
The second study relied on mortality data from Mortality Information System (SIM/Datasus) for 2009-2011 and population figures from the 2010 Brazilian demographic census. Since these two databases do not have a common variable for education, we built a standard variable of completed years of study with the categories 0 to 7 years, 8 to 11 years, and 12 years or more.

For SIM, it was sufficient to combine the first three categories (None, 1 to 3 years, and 4 to 7 years) of the variable "Years of study completed" to form the group from 0 to 7 years. In the case of 2010 Brazilian Census, we needed to create another variable based on seven education-related variables present in the Sample Questionnaire (Table SM.1). This procedure is necessary to obtain a similar metric that makes it possible to match deaths counts and population figures for mortality rates.

Figure 1 illustrates the flowchart containing the level of education attributed to individuals who attend school. To determine the ramifications, we considered that people who had incomplete elementary education or less are in the group with 0 to 7 complete years of study (*low education*). Individuals in high school were assigned to the 8 to 11 years category (middle education). Individuals with higher education onwards fit in with 12 years or more of study.

Finally, we offer a R script containing the step-by-step step algorithm to create a new variable for education based on the combination of other variables present in the Sample Questionnaire in 2010.

Figure SM.1 - Flowchart containing the level of education attributed to individuals who attend school in Brazil



Source: Adapted from Vieira (2016).

Table SM.1 - Selected education-related variables used to create a standard variable for education from the Brazilian Demographic Census (2010)

SOURCE		VARIABLE DESCRIPTION
IPUMS	**IBGE**	
BR2010A_SCHOOL	V0628	**Attending school or daycare** 1. Yes, public 2. Yes, private 3. No, not attending 4. No, never attended
BR2010A_EDLEVEL1	V0629	**Level of school attending** 1. Nursery 2. Pre-school (nursery and kindergarten) 3. Literacy class 4. Literacy for youth and adults 5. Regular elementary school 6. Youth and adults education or supplementary elementary school 7. Regular high school 8. Youth and adults education or supplementary high school 9. Undergraduate 10. Postgraduate (minimum 360 hours)

SOURCE		VARIABLE DESCRIPTION
IPUMS	**IBGE**	
		11. Masters 12. Doctorate
BR2010A_GRADE1	V0630	**Elementary grade/year attending** 1. First year 2. First grade / second year 3. Second grade / third year 4. Third grade / fourth year 5. Fourth grade / fifth year 6. Fifth grade / sixth year 7. Sixth grade / seventh year 8. Seventh grade / eighth year 9. Eighth grade / ninth year 10. Not in a graded course
BR2010A_GRADE2	V0631	**High school grade attending** 1. First grade 2. Second grade 3. Third grade 4. Fourth grade 5. Not in a graded course
BR2010A_COLLEGE	V0632	**Completion of additional undergraduate degree** 1. Yes 2. No
BR2010A_EDATTAIN	V0633	**Highest level of education attended** 1. Daycare, pre-school (nursery and kindergarten), literacy class 2. Literacy for youth and adults 3. Former primary (elementary) 4. Former ginasio (middle school) 5. Elementary school (1st to 3rd grade / 1st to 4th year) 6. Elementary school (4th grade / 5th grade) 7. Elementary school (5th to 8th grade / 6th to 9th year) 8. Supplementary elementary school degree 9. Former scientific, classical, etc. (high school) 10. Regular high school 11. Undergraduate 12. Postgraduate (minimum 360 hours) 13. Masters 14. Doctorate
BR2010A_EDCOMPL	V0634	**Have you completed the highest level of education attended** 1. Yes 2. No

Source: IPUMS International. Minnesota Population Center. Brazilian Demographic census (2010).

R Script - Step-by-step step algorithm to create a new variable with years of study based on the combination of other variables from the 2010 census microdata.

```
Database %<>%
  dplyr::mutate(
    educ_1 = ifelse(
      test = v0628 %in% c(1, 2),

      yes = dplyr::case_when(
      v0629 %in% c(1,2,3,4,6) ~ 0,
      v0629 == 5 & v0630 %in% c(1,2,10) ~ 0,
      v0629 == 5 & v0630 == 3 ~ 1,
      v0629 == 5 & v0630 == 4 ~ 2,
      v0629 == 5 & v0630 == 5 ~ 3,
      v0629 == 5 & v0630 == 6 ~ 4,
      v0629 == 5 & v0630 == 7 ~ 5,
      v0629 == 5 & v0630 == 8 ~ 6,
      v0629 == 5 & v0630 == 9 ~ 7,
      v0629 == 8 ~ 8,

      v0629 == 7 & v0631 == 1 ~ 8,
      v0629 == 7 & v0631 == 2 ~ 9,
      v0629 == 7 & v0631 == 3 ~ 10,
      v0629 == 7 & v0631 == 4 ~ 11,
      v0629 == 7 & v0631 == 5 ~ 99,
      v0629 == 9 & v0632 == 2 ~ 12,
      v0629 == 9 & v0632 == 1 ~ 15,
      v0629 == 10 ~ 15,
      v0629 ==11 ~ 17,
      v0629 == 12 ~ 21),

      no = NA),

    educ_2 = ifelse(
      test = v0628 == 3,

      yes = dplyr::case_when(
        v0634 ==1 & v0633 %in% c(1,2) ~0,
        v0634 ==1 & v0633 == 3 ~ 4,
        v0634 ==1 & v0633 %in% c(4,7,8) ~ 8,
        v0634 ==1 & v0633 %in% c(9,10) ~ 11,
        v0634 ==1 & v0633 %in% c(11,12) ~ 15,
        v0634 ==1 & v0633 == 13 ~ 17,
        v0634 ==1 & v0633 == 14 ~ 21,

        v0634 ==2 & v0633 %in% c(1,2,8) ~ 0,
        v0634 ==2 & v0633 == 3 ~ 1,
        v0634 ==2 & v0633 %in% c(4,7) ~ 5,
        v0634 ==2 & v0633 %in% c(9, 10) ~ 9,
        v0634 ==2 & v0633 == 11 ~ 12,
        v0634 ==2 & v0633 %in% c(12,13) ~ 15,
        v0634 ==2 & v0633 == 14 ~ 17,

        v0628 == 3 & v0633 == 5 ~ 1,
        v0628 == 3 & v0633 == 6 ~ 3),
      no = NA),

    educ = ifelse(
      test = !is.na(educ_1),
```

```
      yes = educ_1,
      no = ifelse(
        test = !is.na(educ_2),
        yes = educ_2,
        no=NA)),

    educ = ifelse(test = v0628==4,
                  yes = 0,
                  no = educ))
```