UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Ciências Econômicas
Programa de Pós-Graduação em Demografia

Maria Laura Lopes Miranda

FORECASTING LIFE EXPECTANCY IN SÃO PAULO CITY, BRAZIL, AMIDST THE COVID-19 PANDEMIC

Belo Horizonte
2023

Maria Laura Lopes Miranda

FORECASTING LIFE EXPECTANCY IN SÃO PAULO CITY, BRAZIL, AMIDST THE COVID-19 PANDEMIC

Dissertação apresentada ao curso de Mestrado em Demografia do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do Título de Mestre em Demografia.

Orientador: Prof. Cássio Maldonado Turra
Co-orientador: Dr. Ugofilippo Basellini

Belo Horizonte
2023

Ficha catalográfica

Miranda, Maria Laura Lopes.
M672f 2023
Forecasting life expectancy in São Paulo city, Brazil, amidst the COVID-19 pandemic [manuscrito] / Maria Laura Lopes Miranda. – 2023.
59 f.: il. e tabs.

Orientador: Cássio Maldonado Turra.
Coorientador: Ugofilippo Basellini.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Centro de Desenvolvimento e Planejamento Regional.
Inclui bibliografia (f. 53-58) e anexos.

1. Mortalidade – Teses. 2. COVID-19 (Doença) – Teses. 3. São Paulo (SP) – Teses. 4. Demografia – Teses. I. Turra, Cássio Maldonado. II. Basellini, Ugofilippo. III. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. IV. Título.
CDD: 304.64

Elaborado por Leonardo Vasconcelos Renault CRB-6/2211
Biblioteca da FACE/UFMG. – LVR/125/2023

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 MARIA LAURA LOPES MIRANDA - NÚMERO DE REGISTRO 2021663250.

Às nove horas do dia vinte e oito do mês de setembro de dois mil e vinte e três, reuniu-se, por videoconferência, a Comissão Examinadora de DISSERTAÇÃO, indicada *ad referendum* pelo Colegiado do Curso em 19/09/2023, para julgar, em exame final, o trabalho final intitulado “Forecasting Life Expectancy in São Paulo City, Brazil, amidst the COVID-19 Pandemic”, requisito final para a obtenção do Grau de Mestre em Demografia*,* área de concentração em Demografia. Abrindo a sessão, o Presidente da Comissão, Prof. Cássio Maldonado Turra, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após, a Comissão composta pelos professores Cássio Maldonado Turra, Ugofilippo Basellini, Bernardo Lanza Queiroz e Enrique Acosta se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. A Comissão **APROVOU** a candidata por unanimidade. O resultado final foi comunicado publicamente à candidata 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, 28 de setembro de 2023.

Prof. Cássio Maldonado Turra (Orientador) (CEDEPLAR/FACE/UFMG)

Prof. Ugofilippo Basellini (Coorientador) (Max Planck Institute for Demographic Research - Alemanha)

Prof. Bernardo Lanza Queiroz (CEDEPLAR/FACE/UFMG)

Prof. Enrique Acosta (Centre d'Estudis Demogràfics/CED - Barcelona)

PROFA. PAULA DE MIRANDA RIBEIRO

Coordenadora do Programa de Pós-Graduação em Demografia

Documento assinado eletronicamente por **Bernardo Lanza Queiroz**, **Professor do Magistério Superior**, em 28/09/2023, às 16:25, 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 29/09/2023, às 09:43, 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 **Ugofilippo Basellini**, **Usuário Externo**, em 02/10/2023, às 13:54, 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 **Jaime Enrique Acosta Gutierrez**, **Usuário Externo**, em 02/10/2023, às 15:49, 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 **Paula de Miranda Ribeiro**, **Coordenador(a) de curso de pós-graduação**, em 03/10/2023, às 13:34, 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 **2640608** e o código CRC **F38AC7E9**.

Referência: Processo nº 23072.257452/2023-18 SEI nº 2640608

AGRADECIMENTOS

Concluir um trabalho acadêmico como este não é uma tarefa fácil. As dificuldades, intensificadas pela realização concomitante de um segundo curso de mestrado em outro país, não poderiam ter sido superadas sem o apoio e suporte de todos que estiveram ao meu lado.

Agradeço ao meu orientador Professor Cássio Maldonado Turra que, além de ter sido um exemplo de pesquisador e demógrafo desde a minha graduação, aceitou a orientação mesmo sabendo dos desafios que teria em orientar uma aluna que buscava trilhar sua trajetória profissional no exterior. Muito obrigada por não só confiar em mim durante a realização desta dissertação, mas por me incentivar a buscar outras oportunidades, sempre aconselhando e dando suporte em tudo o que estivesse ao seu alcance. Sem o seu apoio eu não teria chegado tão longe.

Ao meu coorientador Ugofilippo Basellini, gostaria de agradecer por ter aceito a coorientação e por todo o amparo não só na realização deste trabalho, mas em todas as atividades que participei em Rostock. Seu conhecimento e ensinamentos me trouxeram inspiração e sua atenção, dedicação e acessibilidade me trouxeram calma e foco nos momentos mais necessários. Muito obrigada por todo o aprendizado e por ter tornado esse último ano tão fluido.

Gostaria também de agradecer aos demais professores do departamento de Demografia do CEDEPLAR, em especial, às Professoras Paula de Miranda-Ribeiro e Simone Wajnman que me orientaram ao longo da minha caminhada. Simone foi a minha porta de entrada à Demografia e, desde o início da minha trajetória acadêmica, ainda durante a minha graduação em estatística, se tornou minha inspiração. Muito obrigada por ter me feito apaixonar pela Demografia e por ter me incentivado a aplicar a este programa de mestrado e a tantas outras oportunidades. Seu acolhimento e sensibilidade fazem eu me sentir querida, me trazem conforto.

Aos meus amigos, gostaria de agradecer em especial à Aline e à Tamara por todo o carinho, amor, presença e parceria. Só vocês sabem o quanto reclamei nesses últimos anos. Mas além da liberdade para fazer desabafos, agradeço por permitirem ser o meu refúgio, meu afago. Vocês me trazem cor, energia. Não conseguiria imaginar a vida sem vocês.

Gostaria também de agradecer à Karys, Hisrael, Amanda, Roberta, Thayla, aos demais colegas da coorte 2021 e de outras coortes da Demografia do CEDEPLAR, aos colegas do Summer Incubator e aos colegas do EDSD. O companheirismo de vocês foi fundamental ao longo do meu mestrado.

À minha família, gostaria de agradecer em especial aos meus pais, Andrea e Lázaro, que me proporcionaram todas as condições necessárias para que eu pudesse buscar um curso de pós graduação. Muito obrigada por todo o suporte e compreensão ao longo desses anos. Sem vocês nada disso seria possível.

Por fim, agradeço à CAPES pelo financiamento dos meus estudos.

RESUMO

A pandemia da COVID-19 desencadeou um aumento sem precedentes nos níveis de mortalidade, interrompendo tendências históricas e dificultando as projeções de expectativa de vida em diversos locais. São Paulo, a primeira cidade no Brazil a registrar casos e mortes por COVID-19, registrou entre 2019 e 2021 quedas de mais de quatro anos na expectativa de vida ao nascer para homens e de mais de três anos para mulheres. São Paulo esteve na vanguarda da transição demográfica do país e apresentou queda não linear da mortalidade ao longo do século XX. A trajetória histórica da mortalidade de São Paulo e, mais recentemente, os efeitos disruptivos da COVID-19, apresentaram desafios à projeção de mortalidade na cidade. Neste estudo, usamos um conjunto de dados único iniciado em 1920 para projetar a expectativa de vida em São Paulo até 2050 usando os métodos Lee-Carter (LC) e Lee-Miller (LM). Nossa metodologia incluiu o uso de uma combinação de estimativas baseadas em diferentes períodos de ajuste, entre 1920 e 1995 até 2021, e seis cenários para níveis de mortalidade pós-pandemia. Baseado em 73.200 simulações para cada ano entre 2022 e 2050, as projeções de expectativa de vida foram sumarizadas em valores medianos e intervalos de predição (IP) de 95%. Até 2050, estimamos que a expectativa de vida ao nascer na cidade chegará a aproximadamente 81,9 anos para homens e 88,9 para mulheres. Apesar da ocorrência da pandemia de COVID-19, nossos resultados apresentaram estimativas significativamente superiores ao previsto pela projeção mais recente realizada em 2013. Além disso, dentro do IP de 95%, estimamos que até 2045, tanto a expectativa de vida masculina quanto feminina em São Paulo poderia atingir os níveis dos países com melhor desempenho.

Palavras-chave: projeção de mortalidade; covid-19; Lee-Carter; Lee-Miller; São Paulo.

ABSTRACT

The COVID-19 pandemic has caused a significant rise in mortality rates, interrupting historical trends and making it challenging to forecast future life expectancy levels in many places. São Paulo, the first city in Brazil to report a COVID-19 case and death saw a decrease of over four years in life expectancy at birth for males and over three years for females between 2019 and 2021. São Paulo has been at the forefront of the demographic transition in the country and experienced a non-linear mortality decline over the 20th century. The city's historical mortality trajectory and the disruptive effects of COVID-19 have introduced challenges to mortality forecasting. In this study, we used a unique dataset starting from 1920 to forecast life expectancy in São Paulo until 2050 using the Lee-Carter (LC) and Lee-Miller (LM) methods. Our methodology included a combination of forecasts based on different baseline periods, ranging from 1920 to 1995 until 2021, and six scenarios for post-pandemic mortality levels. Based on 73,200 simulations for each year between 2022 and 2050, we synthesized the resulting life expectancy forecasts into median values and 95% prediction intervals (PI). By 2050, we predict that life expectancy at birth in the city will reach approximately 81.9 years for men and 88.9 years for women. Despite the COVID-19 pandemic, our results presented significantly higher forecasts than what was predicted by the most recent forecast in 2013. Additionally, we estimated within the 95% PI that by 2045, both male and female life expectancy in São Paulo could reach the levels of the best-performing countries.

Keywords: mortality forecasting; covid-19; Lee-Carter; Lee-Miller; São Paulo.

LISTA DE ILUSTRAÇÕES

Figure 1 – Life expectancy at birth by sex, São Paulo city 1920-2019 31

Figure 2 – 1-year and smoothed pace of change in life expectancy at birth by sex, São Paulo city 1920-2019 32

Figure 3 – Log of age-specific death rates by sex, São Paulo city 1920-2021 42

Figure 4 – Life expectancy at birth by sex, São Paulo city 1920-2021 and the Best-Practice slope (OEPPEN & VAUPEL, 2002) 43

Figure 5 – Difference in life expectancy at birth to the Oeppen-Vaupel (2002) Best-Practice level by sex, São Paulo city 1920-2019 44

Figure 6 – Age-group contribution to change in life expectancy at birth by period, Males, São Paulo city 46

Figure 7 – Age-group contribution to change in life expectancy at birth by period, Females, São Paulo city 47

Figure 8 – Life expectancy at birth forecasts with 95% prediction intervals until 2050, Males, São Paulo city 48

Figure 9 – Life expectancy at birth forecasts with 95% prediction intervals until 2050, Females, São Paulo city 49

Table 1 – Summary of forecasts for life expectancy at birth in Brazil, São Paulo state, and São Paulo city around 2050 24

Table 2 – Summary of the six scenarios considered for the return of mortality levels after the pandemic 39

LISTA DE ABREVIATURAS E SIGLAS

IBGE – Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística)

HMD – Human Mortality Database

LAC – Latin American countries

LC – Lee-Carter method

LM – Lee-Miller method

SEADE – SEADE foundation (Fundação SEADE)

SUMÁRIO

1 INTRODUCTION 13

2 BACKGROUND 15

2.1. Mortality trends 15

2.1.1. Mortality trends in Brazil and São Paulo 16

2.2. COVID-19 18

2.2.1. COVID-19 in Brazil and São Paulo 20

2.3. Mortality Forecasts 21

2.3.1. Mortality forecasts in Brazil and São Paulo 23

3 DATA AND METHODS 26

3.1. Data 26

3.1.1. Population estimates 26

3.1.2. Mortality estimates 27

3.2. Methods 28

3.2.1. Methods for evaluating trends in mortality 29

3.2.1.1. The Oeppen-Vaupel (2002) Best-Practice countries 29

3.2.1.2. Arriaga's (1984) age decomposition of differences in life expectancy 30

3.2.2. Methods for forecasting 32

3.2.2.1. The Lee-Carter (1992) and the Lee-Miller (2001) methods 33

3.2.2.2. Starting points 36

3.2.2.3. Scenarios for COVID-19 37

3.2.2.4. Final estimate and prediction intervals 39

4 RESULTS 41

4.1. Trends in mortality 1920-2021 41

4.1.1. Death rates and life expectancy at birth 41

4.1.2. Age-decomposition of changes in life expectancy at birth 45

4.2. Forecasts 2022-2050 47

5 DISCUSSION AND CONCLUSION 51

REFERENCES 53

ANNEXES 59

1 INTRODUCTION

São Paulo has been at the forefront of the demographic transition in Brazil. Over the 20th century, the health transition has evolved in the city, improving survival levels and changing mortality patterns (PRATA, 1992). Improvements in public health through the control of diseases, the provision of water, sanitation, and health services, and the combat of violence reduced mortality differently across age groups in the past century, resulting in a non-linear pace of mortality decline (SIMÕES, 2001). More recently, the effects of COVID-19 in 2020-2022 added more complexity to the historical trajectories of the health transition.

The COVID-19 pandemic triggered an unprecedented rise in mortality in the last few years in many different places. Such impact was significant enough to change historical trends, resulting in reduced life expectancy and a surprising number of excess deaths for some populations. The discrepancies across countries were mainly due to differences in the age structure (DOWD et al., 2020) and context-specific factors, including socioeconomic conditions, chronic disease prevalence, and public policies to combat the virus spread (NEPOMUCENO et al., 2020).

In Brazil, the country with the second largest number of excess deaths from COVID-19 (KARLINSKY and KOBAK, 2021), the estimated reduction in life expectancy was 1.3 years between 2019 and 2020 (CASTRO et al., 2021). The impact on mortality varied across Brazilian states, with São Paulo being one that experienced the worst consequences (CASTRO et al., 2021). São Paulo City, the state's capital and the largest and most developed city in Brazil took the lead in receiving and treating the initial cases of COVID-19 while initiating vaccination efforts in January 2021. Estimates from the SEADE foundation (SEADE, 2022) show that until December 2022, more than 1,1 million cases were confirmed in the city, which resulted in more than 44 thousand deaths. The pandemic's long-term effects on life expectancy's trajectory remain unclear.

Estimating and forecasting mortality tendencies is essential for designing public policies, assessing health insurance premiums, and examining the solvency of pension systems. It is also a way to indicate levels of well-being across nations, regions, and other population subgroups. However, to estimate future trends, it is necessary to have robust enough historical data to extrapolate mortality levels for the following decades. Additionally, forecasts must consider the uncertainty caused by variations in the observed trend, such as mortality shocks and non-linear declines. São Paulo may be the only place in Brazil where a comprehensive historical series of age-specific mortality rates by sex has been documented, tracing back to 1920. The data quality of this historical series has been examined before and proved reliable enough to allow demographic exercises, even though it is not error-free.

In the present thesis, we take advantage of this unique dataset to forecast life expectancy in São Paulo city until 2050 by using a combination of estimates based on different baseline periods, starting between 1920 and 1995 until 2021, and considering six scenarios for the recovery of survival levels following the COVID-19 pandemic. To do this, we use Lee-Carter (1992) and Lee-Miller (2001) methods with death rates obtained from the combination of deaths gathered by the SEADE foundation (SEADE) and population collected by the Brazilian Institute of Geography and Statistics (IBGE) data.

We organized this thesis into five chapters, including this introduction. In the second chapter, we review the literature to provide a comprehensive overview of the relevant findings on mortality trends, its possible disruptions during history, such as the COVID-19 pandemic, as well as the existing methods for forecasting them, with a particular focus on the Lee-Carter (1992) and Lee-Miller (2001) methods. The third chapter describes the data sources used in the analysis, including mortality and population data for São Paulo city, Brazil, from SEADE and IBGE. We also present the methods used to analyze the changes in mortality trends in the city and the strategies used to forecast life expectancy until 2050. Chapter four presents the results. Lastly, we provide a discussion in chapter five.

2 BACKGROUND

In this chapter, we will be covering three parts. The first part will delve into the theories and historical estimates for mortality trends across the globe, including Brazil and São Paulo. The second part will focus on the COVID-19 pandemic's main characteristics and possible consequences on mortality levels. Lastly, we will review the methods employed to forecast mortality and the potential implications of any disruptions on their application.

2.1. Mortality trends

In the last two centuries, all countries have experienced a steady decline in mortality rates, one of the primary features of the demographic transition (DYSON, 2010). The mortality transition initially occurred in northwestern countries after the second half of the XVIII century (KIRK, 1996). However, the start and pace of changes differed significantly within and between world regions (DYSON, 2010).

The decline in mortality rates has significantly increased life expectancy (DYSON, 2010; RILEY, 2001). In industrialized societies, there was a slow and irregular increase in life expectancy over the XIX century, followed by a stable and relatively rapid rise over the last century (WILMOTH, 2000). Evidence shows that the female life expectancy of the country holding the highest level in the world in a given calendar year has been rising at a steady pace of almost three months per year over the last century (OEPPEN & VAUPEL, 2002).

Oeppen and Vaupel's (2002) research highlights a consistent pattern in life expectancy across some countries, particularly Japan, Sweden, and Iceland, the best performers in terms of mortality rates in the past century. Among countries with high mortality in the 1960s, life expectancy at birth is rising at a much steeper pace (MCMICHAEL et al., 2004; BLOOM & CANNING, 2007), and there has been a catching up and convergence of the mortality levels with those pioneers (VALLIN & MESLE, 2004). Other suggests that among countries exhibiting poor population health in 1960–1965, some have made modest progress over the last 45 years due to a mortality trap in which high mortality persists (BLOOM & CANNING, 2007).

With the mortality transition, scholars have paid particular attention to describing changes in health and disease patterns and the demographic, economic, and sociological determinants and consequences (OMRAN, 1971). According to Omran (1971), in his proposed framework for the epidemiological transition, most countries experience a typical pattern in which the prevalence of degenerative diseases replaces the deaths from infectious diseases. As a result, mortality levels decrease considerably, and the mortality pattern gradually changes to older ages.

It has been noted that in various European and Northern American countries, the transition toward a new epidemiological pattern took about a century (MACKENBACH, 1994). However, the changes appear to have unfolded differently in other regions. For example, in Latin America, there is evidence of diseases from both pre- and post-transitional stages coexisting at high rates without a clear resolution of the transition process (FRENK et al., 1991). Therefore, countries in this area seem to exhibit a mixed morbidity pattern (FRENK et al., 1991).

These different historical patterns underscore the need to examine the changes in mortality trends in each context. Although the proposed mortality theories significantly improved our understanding of changes in health and mortality patterns over the years, as discussed before, there are still differences among countries that cannot be explained by a standard process or assigned to a theory's proposition.

Therefore, the following section will explore the characteristics of the mortality trend in São Paulo City and Brazil. We discuss how these two contexts can fit the mortality theories discussed before and in which aspects they differ.

2.1.1. Mortality trends in Brazil and São Paulo

There have been notable discrepancies in mortality transition across countries over the last centuries. Typically, low and middle-income nations, including countries in Latin America, have experienced this process later but at a considerably accelerated rate compared to the high-income countries (MCMICHAEL et al., 2004; BLOOM & CANNING, 2007), leading to a gradual process of convergence of mortality regimes (VALLIN & MESLE, 2004).

Evidence suggests that the mortality decline began after 1940 in Brazil, with crude death rates dropping from approximately 18 per 1000 in 1940 to roughly 6 per 1000 in 1985 (PRATA, 1992). Similarly, life expectancy at birth has exhibited significant improvements, with variations by gender and geographical regions (SIMÕES, 2001). Results indicate that life expectancy at birth in Brazil rose from 42.8 and 47.2 in 1940 to 64.8 and 72.6 in 2000 for men and women, respectively (SIMÕES, 2001).

Nonetheless, the increase in life expectancy at birth in Brazil was not a linear progression. Evidence suggests that between 1940 and 1960, male life expectancy experienced a relatively stable growth, followed by a rapid increase in the 1970s and, subsequently, a stable improvement until the 2000s (SIMÕES, 2001). Among women, the increase in life expectancy was more pronounced between 1940 and 1970 and began to stabilize after the 1970s (SIMÕES, 2001).

Along with the differences in the level, there were significant changes in the age patterns and the contribution of each cause of death to mortality in Brazil after 1940 (BORGES, 2017). Evidence suggests that infectious and parasitic diseases accounted for 46% of the deaths in 1930, while in 1985, this contribution was reduced to close to 7% (PRATA, 1992). On the other hand, participation of circulatory diseases increased from 12% in 1930 to 33% in 1985, and the external causes' contribution rose from 3% to 12% in the same period (PRATA, 1992). According to SIMÕES (2001), the period between 1940 and 2000 witnessed a change in the contributions from young to old-age groups to the increase in life expectancy at birth. Nevertheless, for men, the age groups of 15-19 and 20-29 made a negative contribution between 1980 and 1991, which can be attributed to external causes.

The mortality transition in Brazil is partly due to the introduction of vaccines, vaccination campaigns, and mass vaccination undertaken by the Brazilian government since the late nineteenth century (HOCHMAN, 2011). The establishment of a culture of immunization that started mainly after the campaign to eradicate smallpox in the country between 1966 and 1973 has remained in recent years, especially considering the COVID-19 pandemic, despite the increasing presence of anti-vax groups.

Although health improvements and vaccinations in Brazil have significantly reduced deaths from infectious diseases, studies indicate that external causes have played a crucial role in the country's mortality rate, particularly among men during the 1980s (BORGES, 2017). This surge is mainly attributed to violence (BORGES, 2017), which caused a reduction in mortality gains for men during that time, a trend that persisted in other Latin American countries (LAC) (CANUDAS-ROMO et al., 2014). This high incidence of violence was pointed out as responsible for the divergence of some LAC towards the mortality profiles of developed countries in recent years (JÉSUS-ADRIÁN et al., 2020).

Regional differences in mortality are another characteristic of the mortality profile in Brazil. Socioeconomic and regional inequalities within the country have resulted in diverging life expectancy trends among the five major geographical regions and between urban centers and rural areas (BORGES, 2017). São Paulo state, recognized as the most developed region in the country, followed the same pattern of mortality decline over the decades, but at higher magnitudes. Evidence shows that life expectancy at birth for São Paulo rose from 44.3 and 46.7 in 1940 to 66.8 and 75.6 in 2000 for men and women, respectively (WALDVOGEL et al., 2003), faster than the life expectancy improvements observed in the country on the same period. São Paulo state's higher income levels has attracted migrants from both within and outside the country (PERILLO & PERDIGÃO, 2005). As a result, it's crucial to take into account the potential impacts of migration on health and mortality rates in the state (PASSARELLI, 2021).

To summarize, the mortality decline in Brazil and São Paulo state has had significant changes over the years, unlike what has been observed among the most developed countries. Life expectancy has shown non-linear increases. These findings emphasize the complex and multifaceted nature of mortality decline in the country during the past century, and underscore the importance of considering these changes while estimating future mortality levels. Recently, in light of the COVID-19 pandemic, it is vital to understand how the new mortality shock affected life expectancy levels and discuss the uncertainty over the long-term effects of the pandemic on life expectancy's trajectory.

2.2. COVID-19

Besides the socioeconomic (PADHAN & PRABHEESH, 2021) and health consequences (DEL RIO, COLLINS & MALANI, 2020), the COVID-19 pandemic triggered an unprecedented rise in mortality levels in the last few years, with discrepancies among countries. Such impact was significant enough to interrupt historical trends, resulting, for example, in a reduction in life expectancy and a surprising number of excess deaths.

Estimates show declines of more than four years in life expectancy at birth between 2019 and 2021 in Bolivia, Botswana, Lebanon, Mexico, Oman, and the Russian Federation (UNITED NATIONS, 2022). Similar analyses show substantial life expectancy reductions in Eastern Europe and the United States between 2019 and 2021, but not among the Western European countries (SCHÖLEY et al., 2022). Regarding excess mortality, estimates show more than 4.0 million excess deaths worldwide until June 2021 (KARLINSKY & KOBAK, 2021). Among the worst-affected countries, such as Peru, Ecuador, Bolivia, and Mexico, the excess mortality exceeded 50% of the expected estimate (KARLINSKY & KOBAK, 2021).

Various factors such as age structure, comorbidity profiles, and contextual factors have led to disparities in the impact of the pandemic across different countries (DOWD et al., 2020; NEPOMUCENO et al., 2020). Additionally, the pandemic has affected individuals differently based on socioeconomic and demographic characteristics. Studies have shown that elderly and middle-aged adults, mainly men, have experienced higher fatality rates (LEVIN et al., 2020; RAMÍREZ-SOTO et al., 2021).

In addition, there have been disparities in mortality rates across different ethnic groups, for example, in the United States during 2020, with Black and Latino populations being the most impacted (ANDRASFAY & GOLDMAN, 2020). Studies have also revealed a connection between obtaining a higher level of education and a reduced risk of severe COVID-19 outcomes among the European populations (YOSHIKAWA & ASABA, 2021).

Nevertheless, estimating the pandemic's impact on mortality remains challenging, even when considering measures such as excess mortality, which mitigates the effect of misidentification of COVID-19 as the underlying cause of death. Age-specific case counts depend highly on the testing capacity, testing strategy, and differences in the definition of cases across sources and over time, especially in low- and middle-income countries (RIFFE et al., 2021). Also, due to the behavioral changes induced by the pandemic and the differential fatality rates associated with the presence of comorbidities, establishing a causal relationship between COVID-19 and other causes of death occurring after 2020 becomes complex (CASTRO et al., 2023). It is possible that the rates of these chronic ailments may have or will decrease if patients succumb to COVID-19, a process known as the "harvesting effect" (SCHWARTZ, 2000; CASTRO et al., 2023). Also, due to the reduced social and work activities, deaths from specific external causes may have declined during the pandemic (CASTRO et al., 2023). Conversely, deaths from other causes may have increased due to several aspects, including inadequate care in clinics and hospitals because of shortages of equipment, staff, and space, worsening of comorbidities owing to the effects of COVID-19, and other factors (DEY & DAVIDSON, 2021).

In summary, it is challenging to assess the impact of COVID-19 on mortality due to its interaction with other causes of death (CASTRO et al., 2023) and data issues (RIFFE et al., 2021). Additionally, it is uncertain when mortality levels will return to pre-pandemic levels.

2.2.1. COVID-19 in Brazil and São Paulo

In Brazil, the estimated reduction in life expectancy was 1.3 years between 2019 and 2020 (CASTRO et al., 2021). Until May 2021, Brazil was considered the second country with the most notable excess deaths after COVID-19 (KARLINSKY & KOBAK, 2021). From March 15 to June 6, 2020, there were over 62,000 excess deaths caused by diseases observed, representing a 22% increase in overall mortality over 12 weeks (MARINHO, 2020). Evidence indicates that the excess deaths followed the spread of COVID-19 throughout the country, presenting regional variations (MARINHO, 2020). In São Paulo state, where the first COVID-19 case and death in Brazil were reported (MELO et al., 2020), there was a 16% increase (10,012 deaths) in mortality from March 15 to June 6, 2020 (MARINHO, 2020). The state also experienced a decline in life expectancy at birth, with reductions of 1.4 years for males and 0.8 years for females between 2019 and 2020 (CASTRO et al., 2021).

Due to the concentration of the population in major cities, the impact of the pandemic on mortality was more pronounced in urban areas, particularly in the capital cities. São Paulo city, the state's capital, recorded a disproportionately higher number of excess deaths, accounting for 36% (6,208) above the expected average from March 15 to June 6, 2020

(MARINHO, 2020). By December 2022, the city had confirmed over 1.1 million cases and more than 44,000 deaths due to COVID-19 (SEADE, 2022).

Despite the uneven distribution of healthcare infrastructure and ongoing cuts in health spending in Brazil, the public health system has remained operational during the COVID-19 pandemic (BARBERIA & GÓMES, 2020). In January 2021, the vaccination campaign started in the country, with São Paulo being the first city to vaccinate. The campaign encountered several challenges, including political instability, lack of national planning, delays in securing supplies such as syringes and needles, and vaccine shortages (CASTRO et al., 2021; BARBERIA & GÓMES, 2020). However, given Brazil's historical culture of immunization, more than 80% of Brazilians aged six months and older have completed the primary vaccination schedule as of 2022 (MINISTRY OF HEALTH, 2022). This high acceptance of immunization among the population led to a large reduction of COVID-19 deaths in 2022 (75,429) compared to 2021 (427,629), despite the similar number of identified cases in the two years: 14,768,820 in 2021 and 14,096,655 in 2022 (CONASS, 2023).

Considering this rapid decline in COVID-19 deaths, it is plausible that mortality levels will return to pre-pandemic levels in the coming years. However, given the challenges in identifying the impact of COVID-19 on mortality discussed in the previous section, there remains uncertainty regarding the mortality level's return.

2.3. Mortality Forecasts

Considering the sustained decrease in mortality rates and recent improvements in life expectancy, future predictions about life expectancy levels are crucial for creating policies, estimating health insurance costs, and designing pension systems. Although longer lifespans can be viewed as a positive development and significant social accomplishment for people, it also raises worries about their impact on public spending for elderly support (TULJAPURKAR & BOE, 1998).

Since the early twentieth century, various reliable techniques have been developed and utilized to estimate mortality rates (POLLARD, 1987). These techniques are usually extrapolative, using age patterns and time trends to forecast the future (BOOTH &

TICKLE, 2008). However, in the last three decades, the introduction and improvement of stochastic methodologies have led to a flourishing of mortality forecasting (BASELLINI, 2020). The Lee-Carter (LC) method (LEE & CARTER, 1992) became popular among demographers due to its innovative and straightforward approach.

The LC method, which we will discuss extensively in Chapter 3, combines demographic mortality and time series models. Its initial application involved forecasting specific death rates in the United States from 1990 to 2065 for comparison with social security forecasts. After its publication, the method has gained popularity, with several applications and critics, leading to the development of new variants. One prominent variant is the Lee-Miller (LM) (LEE & MILLER, 2001). Besides its accuracy (LEE & MILLER, 2001; BOOTH et al., 2006), the LC and its variants present several advantages for mortality forecasting. The method was proposed in a non-complex functional form. It also allows simplified forecasts derived from the projection of the single time index and the derivation of mortality prediction intervals given its probabilistic form (BASELLINI, 2020).

Nonetheless, despite its popularity, the method is not without limitations. The time series model proposed by Lee and Carter (1992) is linear and passes through the first and last point of the baseline period (BASELLINI, CAMARDA & BOOTH, 2022). However, as aforementioned, mortality levels do not always exhibit linear declines, particularly among countries lagging in the demographic transition. Consequently, different choices of the baseline period can lead to various forecasts. In the case of Brazil, life expectancy at birth increased at different paces on the last century (SIMÕES, 2001). Therefore, using a baseline period characterized by a fast rate of mortality decline would lead to a more optimistic estimate for the future. In contrast, a period with a slow decline would lead to a pessimistic outcome.

Moreover, the LC does not incorporate information on future trends (LEE, 2000). This is particularly relevant in the context of mortality shocks, such as the COVID-19 pandemic, where it is crucial to incorporate future information on the expected recovery of historical mortality levels. Additionally, data on the years affected by the mortality shock caused by COVID-19, if considered in modeling applications, can significantly impact point and interval forecasts of death rates (SCHNÜRCH, 2022).

Brazil has been severely impacted by COVID-19, resulting in many excess deaths (as reported by Karlinsky and Kobak in 2021). Despite the progress regarding vaccination rates (as stated by the Ministry of Health in 2022), it remains uncertain when mortality rates will return to pre-pandemic levels. Given this complex and non-linear recent change in mortality, the estimates of future survival levels must also account for the uncertainty stemming from various scenarios for the post-pandemic return of mortality levels.

2.3.1. Mortality forecasts in Brazil and São Paulo

The IBGE, the bureau of census in Brazil, is responsible for the official population projections in the country and uses a deterministic approach to forecast mortality. This procedure involves the application of an interpolation technique between the most recent available life table and a limit table (IBGE, 2016). During the first two years of the pandemic, the institution provided mortality estimates without accounting for the impact of COVID-19. Considering that the next census would take place in 2022, mortality estimates for 2020, 2021, and projections until 2060 (IBGE, 2018) were interpolated using data from the previous census conducted in 2010 (IBGE, 2022; IBGE, 2016). Estimates from the IBGE for the country indicate life expectancies at birth of 77.4 and 83.8 years for men and women, respectively, in 2050 (IBGE, 2018). For São Paulo state, the IBGE estimates life expectancies at birth of 79.6 and 85.1 years for men and women in 2050 (IBGE, 2018).

Fundação SEADE, a research institution based in the state of São Paulo, Brazil, produces, analyzes, and disseminates socioeconomic data and indicators related to the state of São Paulo to support public policies and decision-making processes. SEADE regularly projects the population of São Paulo state and its capital, São Paulo. The institution's procedure involves applying the cohort component method (WHELPTON, 1928). The process combines vital registration estimates gathered by the SEADE foundation with assumptions about the future behavior of each demographic component (SEADE, 2017). The last population projections from SEADE were in 2017 and, thus, do not account for the COVID-19 effects. Estimates from the SEADE Foundation for São Paulo state indicate life expectancies at birth of 79.1 and 84.2 years for men and women, respectively, in 2050 (SEADE, 2017). For São Paulo city, SEADE estimates are 79.2 and 84.7 years for men and women, respectively, in 2050 (SEADE, 2013).

In addition to the two primary sources of demographic projections in the country, other sources provide estimates of future life expectancies. Using data between 1974 and 1990, Fígoli (1998) applied the LC method to forecast life expectancy in Brazil until 2040. The estimates suggest life expectancies at birth of 71.9 and 78.4 years for men and women between 2035 and 2040 in the country (FÍGOLI, 1998). For São Paulo city, using eight different baseline periods between 1920 and 2005, Silva (2009) forecasted life expectancy at birth between 2006 and 2100 by applying the LM method. Forecasted life expectancies at birth are 77.8 and 84.4 years in São Paulo city in 2050 for men and women, respectively (SILVA, 2009).

Table 1 compares the forecasted life expectancies at birth for Brazil, São Paulo state, and São Paulo city. The results indicate that by 2050, levels will be around 79 for males and 84 for females in both São Paulo state and city. However, it is crucial to note that the most recent estimates consider as input the last census available, which took place in 2010, potentially leading to underestimations of life expectancy levels.

Table 1 – Summary of forecasts for life expectancy at birth in Brazil, São Paulo state, and São Paulo city around 2050

Place	Male	Female	Year of estimate	Source
Brazil	77.4	83.8	2050	IBGE (2018)
Brazil	71.9	78.4	2035/2040	Fígoli (1998)
São Paulo state	79.6	85.1	2050	IBGE (2018)
São Paulo state	79.1	84.2	2050	SEADE (2017)
São Paulo city	79.2	84.7	2050	SEADE (2013)
São Paulo city	77.8	84.4	2050	Silva (2009)

Source: IBGE (2018), Fígoli (1998), SEADE (2013, 2017), Silva (2009)

These findings show progress in forecasting mortality rates in São Paulo, Brazil. However, they do not fully account for the effects of the COVID-19 pandemic. It is important to acknowledge the uncertainty surrounding mortality rates post-pandemic and incorporate this into future forecasts. Additionally, the non-linear changes in mortality over the past century must be taken into account as separate scenarios to ensure accurate forecasts. To achieve this, it is crucial to use different baseline periods and include more recent data when conducting mortality forecasts.

3 DATA AND METHODS

In this study, we estimate future life expectancies in São Paulo city until 2050, incorporating the uncertainty caused by COVID-19 and non-linear changes in mortality trends over the last century. We use the Lee-Carter (1992) and Lee-Miller (2001) methods to do this. These two approaches (discussed in the methods section) extrapolate historical trends and age patterns in mortality using observed death rates as input (LEE, 2000).

To estimate future mortality in São Paulo, we need population and death data by age and sex for multiple years. São Paulo is known for its reliable demographic data, including precise death registration and population enumeration. We gathered death counts and population estimations by sex and age groups from 1920 to 2021 from SEADE and the IBGE. We used a modified version of the data that Souza (2023) applied in her dissertation. Similar datasets have been used before by Silva (2009) and Siviero (2009).

Using data since 1920 improves our understanding of the historical mortality trends and enhances the models' accuracy, enabling us to make more informed predictions. Therefore, using death rates for São Paulo is the ideal way to apply forecasting methods to analyze future survival levels in Brazil.

3.1.Data

3.1.1. Population estimates

We use data from the Brazilian census for 1920, 1940, 1950, 1960, 1970, 1980, 1991, 2000, and 2010. For 2020, we use the population projections from the IBGE. The population data are organized by sex and 5-year age groups for almost all years, except 1920, 1940, 1950, and 1960, in which some age groups are in 10-year intervals (between ages 30 and 80 in 1920, 1950 and 1960, and for all age groups in 1940).

To obtain 5-year age groups up to 80+ for all census years, Souza (2023) applied the Karup-King interpolation method (SIEGEL & SWANSON, 2004) to break the 10-year into 5-year age groups. The Karup-King was chosen for its simplicity and accuracy (SIEGEL & SWANSON, 2004). Also, we decided not to model ages older than 80 because of the lack of information and the risk of age misreporting at advanced ages in Brazil (GOMES & TURRA, 2009; NEPOMUCENO & TURRA, 2020; PAES & ALBUQUERQUE, 1999).

After obtaining population numbers by 5-year age groups for both men and women in the censuses years (1920, 1940, 1950, 1960, 1970, 1980, 1991, 2000, and 2010) and 2020, we used a linear interpolation method to estimate population by age and sex for the other calendar years up to the year of 2021. The approach uses the inter-censitary growth rates of each age group between two censuses to obtain estimates for the population in the years between them. We used the same method to estimate the population in 1930 - a year the census was not collected - and to extrapolate the population from 2020 to 2021.

3.1.2. Mortality estimates

The SEADE provides death records for São Paulo city between 1920 and 2021. The death counts are distributed by sex and by 5-year age groups in almost all years (except between 1920 and 1923). The records with ignored information on age were proportionally distributed among the age groups. Souza (2023) also used the Karup-King method (SIEGEL & SWANSON, 2004) to disaggregate deaths from the 10-year into 5-year age groups from 1920 to 1923.

Until 1969, death registration in Brazil was done by place of occurrence instead of the place of residence. Therefore, there was no distinction in the death records between residents and non-residents of São Paulo city. Given the magnitude of people attracted to the city given its development (PERILLO & PERDIGÃO, 2005), it is essential to remove the effect of the so-called "invasion of deaths" (GONÇALVES, 1974; LAPREGA & MANÇO, 1999) on the death records. Evidence shows that in 1968 they accounted for 11% of the deaths in the city (GONÇALVES, 1974).

In order to correct for the invasion of deaths in São Paulo city, Berquó (1974) proposed to remove from the death records the proportion due to non-resident deaths, estimated by a longitudinal study that she had conducted. Berquó (1974) estimated proportions were used by Silva (2009) and Siviero (2009) to obtain death estimates free from the effects of non-city residents between 1920 and 1969 in São Paulo city.

Besides the invasion of deaths, another possible disruption in the data quality could be due to a lack of coverage of the death records. Although there were improvements in data quality in Brazil in the late 1900s, evidence shows that the under-registration of deaths still occurred in the country in the twentieth century, even in São Paulo state, known for its higher quality (VASCONCELOS, 1998).

In order to correct possible issues with lack of coverage, the General Growth Balance method (HILL, 1987) was applied to the death records between 1920 and 2000. After 1990 evidence shows that the coverage in São Paulo state was close to 100% (PAES & ALBUQUERQUE, 1999). With this, it was possible to obtain estimates for the death counts between 1920 and 2021 by 5-year age groups and sex corrected by the possible invasion of deaths and lack of coverage. Again, it is essential to emphasize that data until 2000 with the corrections described above was shared by Souza (2023) in July 2022.

The next step was to divide the death count estimates by the population estimates to obtain the age-specific death rates for this period. Hence, it was also possible to construct life tables for all the years using standard demographic methods (PRESTON et al., 2001) and obtain life expectancy estimates.

3.2.Methods

The main goal of this thesis is to forecast mortality rates for São Paulo city, considering various uncertainties caused by changes in observed trends, including the effects of COVID-19 and advancements in the city's healthcare system. Before forecasting, it is crucial to examine the mortality changes that have happened over the years and discuss their potential impacts on future life expectancies. Therefore, the methodology section will be split into two parts. The initial sub-section will analyze the techniques used to assess mortality trends within the city, while the second sub-section will detail the methodology applied to forecast life expectancy up to 2050.

3.2.1. Methods for evaluating trends in mortality

3.2.1.1. The Oeppen-Vaupel (2002) Best-Practice countries

In addition to analyzing age-specific death rates, it's important to compare the mortality rates of São Paulo with other countries, such as those considered Best-Practice by experts in the field like Oeppen and Vaupel (2002). They found a consistent increase in life expectancies across the globe, with the country holding the highest level experiencing an average increase of three months per year over the last century (RIFFE, NEPOMUCENO & BASELLINI, 2020).

To obtain such a result, Oeppen and Vaupel (2002) used period life expectancies for all populations of the Human Mortality Database (HMD) between 1840 and 2000 to identify the highest estimate in each year. After identifying the Best-Practice countries, the authors applied a linear model to compute the slope of the increase in life expectancy at birth over the years.

Other scholars have studied the trends in mortality rates over time. For instance, Vallin and Meslé (2009) proposed a segmented trend line to showcase the increase in life expectancy in the last two centuries. Their recommendation was to use four segments to represent the maximum life expectancies trajectory worldwide between 1840 and 2000, considering countries in and outside the HMD. The results found by Vallin and Meslé (2009) are more comprehensive in the study of life expectancy trajectories, as they include other mortality contexts besides HMD. However, for this work, we will only compare the life expectancy levels in São Paulo with the Oeppen and Vaupel (2002) line to keep the analysis simple.

Using a similar dataset, this study computed the slope of Best-Practice countries and compared it with trends observed in São Paulo city over the years. The analysis considered data from HMD until 2019 for both sexes, instead of limiting it to 2000. Since the COVID-19 pandemic had a significant impact on life expectancy after 2019, including such data could distort the analysis.

The estimated slopes for women and men were 0.250 ($R^2 = 0.977$) and 0.230 ($R^2 = 0.971$), respectively. These results are slightly higher than those obtained by Oeppen and Vaupel (2002), where the slope was 0.243 ($R^2 = 0.992$) for women and 0.222 ($R^2 =$

0.980) for men. The difference between the results could be due to data adjustments and the inclusion of other countries and more recent time periods after Oeppen and Vaupel's publication.

3.2.1.2. Arriaga's (1984) age decomposition of differences in life expectancy

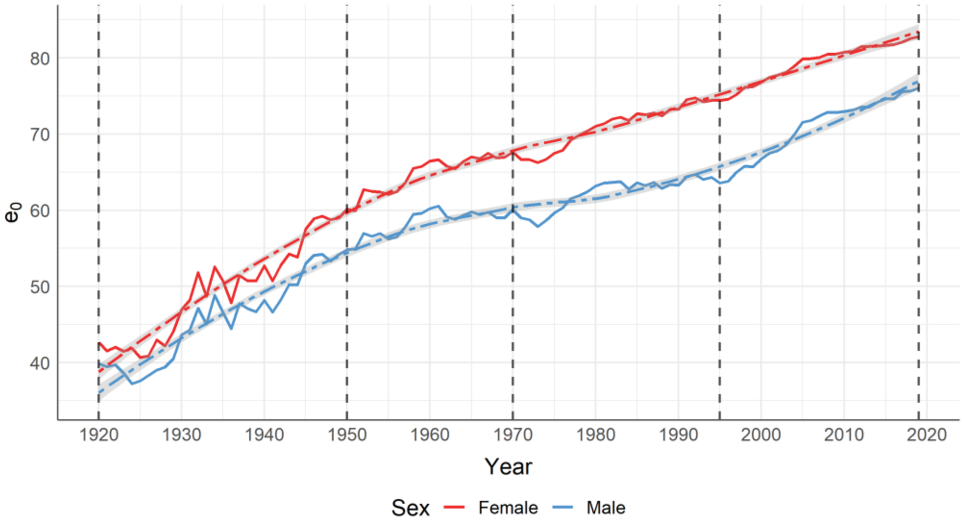
When analyzing mortality rates, it's important to examine how the age pattern of mortality changes and which age groups contribute to the differences in life expectancy. Our study utilized Arriaga's (1984) methodology to break down changes in life expectancy at birth and understand how each age group contributes to the overall mortality experience in São Paulo city. The author suggests two types of effects, direct and indirect (POLLARD, 1988). The direct effect refers to the change in years of life within a specific age group due to mortality changes in that age group. The indirect effect is the number of additional life years added to a given life expectancy because mortality changes within a particular age group will impact the number of survivors at the end of that age interval (ARRIAGA, 1984).

Arriaga's (1984) study noted that mortality changes occur simultaneously across all age groups, which led to the proposal of an interaction effect that explains overall mortality change on life expectancy. However, this work focuses on estimating the effect of each age group, so the interaction effect will not be estimated. Instead, the estimated direct and indirect effects will be used to decompose the total effect of each age group on changes in life expectancy during defined intervals. The baseline period will be divided into four intervals: 1920-1950, 1950-1970, 1970-1995, and 1995-2019. Arriaga's (1984) decomposition method will then be applied to determine each age group's contribution to life expectancy changes between the start and end points of each interval.

As shown in Figure 1, there are differences in the pace of the increase in life expectancy in São Paulo city for both men and women with cut points in 1950, 1970, and 1995. As discussed in subsection 2.1 and later in the results, there are some possible explanations for such changes, and it is crucial to understand the contribution of each age group to the observed differences in life expectancy along with the variation over time. Figure 1 shows that the period between 1920 and 1950 was marked by a rapid increase in life expectancy at birth, followed by a deceleration between 1950 and 1970, especially among males.

After 1970, life expectancy increased steadily, followed by a more stable increase after 1995.

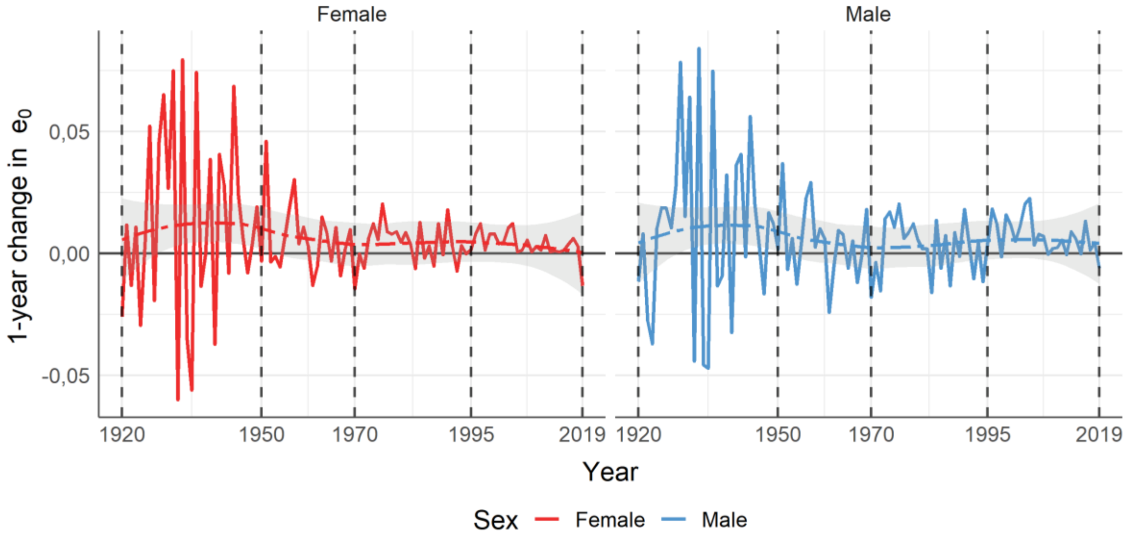
Figure 1 – Life expectancy at birth by sex, São Paulo city 1920-2019



Source: SEADE, IBGE

The findings in Figure 2 support the previous observations. By examining the one-year and smoothed rate of change in life expectancy, we can see that between 1920 and 1950, there was a significant increase in life expectancy. However, after 1950, there was a noticeable decrease in the rate of change, followed by a slight increase from 1970 to 1995. For the last period (1995-2019), we can observe different dynamics by sex, with a steady increase among males, and a deceleration in the pace of change among females, with changes approaching zero.

Figure 2 – 1-year and smoothed pace of change in life expectancy at birth by sex, São Paulo city 1920-2019



Source: SEADE, IBGE

To identify potential changes in the distribution of life expectancy at birth, we used a method for detecting multiple change points, as outlined by MATTESON & JAMES (2013). This involved applying the method separately for men and women, which revealed two change points for both genders in 1952 and 1992. These findings align with our proposed timeframe, although we included the 1970 change point in our analysis for more comprehensive results.

3.2.2. Methods for forecasting

To forecast life expectancy at birth up to the year 2050, we used the seminal method of Lee-Carter (1992) and the extension proposed by Lee-Miller (2001). We considered different baseline periods and various scenarios for the return of mortality levels after COVID-19, which we will discuss in the next sub-sections.

3.2.2.1. The Lee-Carter (1992) and the Lee-Miller (2001) methods

The Lee-Carter (LC) is a stochastic method developed by Ronald D. Lee and Lawrence R. Carter in 1992 to model and forecast age-specific death rates, combining demographic mortality and time series models. The model was proposed as follows:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

where $\ln(m_{x,t})$ denotes the logarithm of the death rates by age group x at year t; a_x describes the average shape of the mortality by age; b_x describes the extent to which mortality at age x changes given the overall temporal change in the general level of mortality; k_t is the mortality level index at time t; and $\varepsilon_{x,t}$ is an error term reflecting residual age-specific influences not captured by the model.

In demography, the LC can be seen as a relational method (LEE & MILLER, 2001) since it involves the transformation of a standard mortality profile (a_x) fixed over time. Mortality dynamics over age and time are described by the interaction of the b_x and k_t parameters that modify the standard time-fixed profile a_x (BASELLINI, CAMARDA & BOOTH, 2022). If k_t declines linearly, each age-specific death rate declines at its exponential rate, proportional to b_x and depending on how fast k_t declines.

The LC model does not have a unique solution. To address this, Lee and Carter introduced two constraints ($\sum_x b_x = 1$ and $\sum_t k_t = 0$) to ensure model identification. The model's parameters can then be estimated using singular value decomposition (SVD), as done by Lee and Carter (1992), to derive an ordinary least squares solution. The estimation presented on the original LC involves computing a matrix of deviations of $\ln(m_{x,t})$ from the a_x (considered in this case as the mean of all the $\ln(m_{x,t})$ over time) and then applying the SVD to this residual matrix.

In this case, b_x is considered the proportional pace of change and is obtained from the first left singular vector (adjusting for its constraint). Similarly, k_t can be obtained by the product of the leading singular value, the first right singular vector, and the sum of the first left singular vector (BASELLINI, CAMARDA & BOOTH, 2022).

In the original version, Lee and Carter (1992) proposed an adjustment on the k_t parameter, so the number of deaths estimated by the method using fixed a_x and b_x would match the ones observed each year t. After this adjustment, they proposed forecasting the k_ts using a random walk process with drift. This time series model can be expressed as follows:

$$k_t = c + k_{t-1} + u_t$$

Where c is the drift and u_t is a white noise process with mean zero and variance σ^2.

Lee and Carter (1992) introduced an innovative and straightforward approach to forecasting mortality rates, yet the method has limitations and disadvantages. The method assumes that the changes in the mortality pattern are constant over time (fixed b_x). This premise fails to account for the possibility that the decline may vary in pace for specific age groups over time. Especially for long-term projections, it is unlikely that such a pattern would hold proportionally to all age groups for more than some decades into the future (LI, LEE & GERLAND, 2013).

The method also relies on $\hat{k}_t$s linearity and the fitting period's choice. The proposed random walk with drift time series model, as suggested by Lee and Carter (1992), is linear and passes through the first and last point of the series (BASELLINI, CAMARDA & BOOTH, 2022). However, mortality levels do not always decrease linearly, particularly among countries considered behind in the demographic transition over the last century, as shown before in the case of São Paulo. Consequently, different choices of the baseline period can lead to various forecasts, and the starting year's influence lies in its position in the historical series (BASELLINI, CAMARDA & BOOTH, 2022).

Other limitations are the adjustment of the $\hat{k}_t$ to match the observed deaths, the method gives greater weight to ages at which numbers of deaths are large (LEE & MILLER, 2001), and the jump-off error, in which the forecasted death rates would not fit the observed death rates in the jump-off year. Given that a_x is considered the mean of all the $\ln(m_{x,t})$ over time, previous years with higher mortality levels would also influence the predicted levels, leading to higher forecasted mortality levels that don't necessarily reflect the recent trends. Finally, the LC method does not consider information on future trends

(LEE, 2000). This is particularly relevant in the context of mortality shocks, such as the COVID-19 pandemic, where it is crucial to incorporate future information on the expected return of mortality levels.

The LC method was developed to forecast age-specific death rates in the U. S. from 1990 to 2065 and compare them with the social security forecasts. After its publication, the method became very popular, with several applications and critics, leading to the development of new variants. Lee and Miller proposed one of them in 2001 (LM) to address some of the main limitations of the LC, specifically: (i) the jump-off error, (ii) the assumption of fixed changes in the mortality pattern over time, and (iii) the adjustment of the time index $\hat{k}_t$. After identifying that the LC tended to underestimate gains in life expectancy, Lee and Miller (2001) proposed using the last available year for the average shape of mortality by age (a_x) instead of considering the mean of all the $\ln(m_{x,t})$ over time. According to the authors, the underestimation of gains in life expectancy induced by the influence of higher mortality levels at the beginning of the baseline period would be reduced with this change. This approach also eliminates completely the jump-off error (i.e. the error in the last year of the fitting period), but likely introduces idiosyncrasies and additional variability in the forecast rates, since the specific mortality pattern observed in the last year is carried forward (and potentially amplified) in the forecast.

Additionally, Lee and Miller (2001) discussed the possibility that changes in the mortality pattern are not constant over time (varying b_x). The authors found that for the United States, Sweden, France, Canada, and Japan, the age pattern of mortality decline has shifted systematically, comparing periods before and after 1950. To address this issue, the authors proposed the use of data only after 1950, since according to them, and following Tuljapurkar et al. (2000), this would be a period in which the change in the age pattern of decline (non-fixed) has relatively weak effects on the forecast of e0. Note that this approach still relies on a time-constant set of b_x.

The authors also proposed changing the k_t parameter adjustment from matching the number of deaths estimated by the method with the ones observed in each year t to matching the estimated life expectancy levels to those observed at time t. This adjustment of k_t by fitting to e_0 was adopted to avoid the use of population data as required for fitting to D_t (LEE & MILLER, 2001).

Other variants also tried to account for different limitations of the model. For instance, Booth et al. (2002) proposed changes to the LC method to adjust the time component to reproduce the age distribution of deaths rather than total deaths and to determine the optimal fitting period to address non-linearity in the time component. Renshaw and Haberman (2006) extended the model to account for cohort effects, while Hyndman and Ullah (2007) proposed using the functional data paradigm (RAMSAY & SILVERMAN, 2005) to capture additional dimensions of change in mortality rates from the log death rates. Li, Lee, and Gerland (2013) proposed using a rotation in the projected age pattern of mortality decline to account for the changes observed in the b_x over time.

In this thesis, we apply the LC and LM methods to forecast life expectancy at birth in São Paulo city until 2050. These models were chosen due to their simplicity and widespread acceptance in mortality forecasting. To determine a_x for the LC method, we calculated the mean of all the $\ln(m_{x,t})$ up to 2019, while for the LM, the a_x was defined as equal $\ln(m_{x,t})$ of 2019. Essentially, we did not take into account the impact of COVID-19 on the average shape of mortality by age. All other estimated parameters and adjustments were made in accordance with the proposals of the authors for both methods.

3.2.2.2. Starting points

As we discussed earlier, there has been a significant change in living conditions in São Paulo over the last century. This has resulted in non-linear changes in the mortality pattern of the city (SIMÕES, 2001). When it comes to forecasting, historical trends are usually extrapolated into the future, including the Lee-Carter (1992) and Lee-Miller (2001) methods. However, this can lead to considerably different forecasted results depending on the chosen baseline period, especially if there are non-linear changes in the mortality level over time. To address this issue, Lee and Miller (2001) suggested using data after 1950 since this period has relatively weak effects on the forecast of e0 due to the non-fixed age pattern of decline. Similarly, Hyndman and Booth (2008) also adopted the starting year of 1950 to avoid external influences in the mortality trend, such as the two world wars and the Spanish Influenza in 1918, as well as structural changes over the last century.

The choice of using 1950 as a starting point is subjective and may not be applicable to all countries. While some studies such as those by Lee and Miller (2001), Tuljapurkar et al. (2000), and Hyndman and Booth (2008) suggest that 1950 is a suitable starting point, it is uncertain whether this year accurately represents mortality patterns in less developed countries. For instance, in Brazil, mortality rates began to decrease after 1940 (PRATA, 1992), and there were non-linear declines throughout the last century, resulting in varying rates of increase in life expectancy at birth (SIMÕES, 2001).

This study used different starting points to forecast life expectancy until 2050, aiming to prevent the subjectivity of choosing periods with fast or slow mortality declines. The approach was to use baseline periods starting at each year between 1920 and 1995 to forecast life expectancy in São Paulo city from 2020 to 2050. For the LM method, baseline periods started between 1950 and 1995. As such, for each COVID-19 scenario and choice of the final year of the fitting period (which will be presented in the next subsection), we run the LC method for 76 different baseline periods and the LM for 46, separating men and women.

3.2.2.3. Scenarios for COVID-19

In our study, we faced a challenge in measuring mortality trends due to the COVID-19 pandemic. This led to a significant increase in deaths, particularly in Brazil, which had the second-highest number of excess deaths after COVID-19 (KARLINSKY and KOBAK, 2021). It's uncertain whether mortality levels will return to pre-pandemic levels despite the decline in COVID-19 deaths and the widespread acceptance of vaccinations among the population (MINISTRY OF HEALTH, 2022). Although it is plausible that mortality will return to pre-pandemic levels, there is still uncertainty due to the challenges of identifying the impact of COVID-19 on mortality (see section 2.2). To address this issue, we propose using different scenarios for the recovery of mortality levels in the years following COVID-19. These scenarios were computed by including or not the observed death rates of the two years of COVID-19 (2020 and 2021) and the creation of assumptions for the years after 2021.

We used six different scenarios to forecast life expectancies at birth until 2050. The first scenario utilized data until 2019 and assumed that the mortality level observed before the pandemic would continue in the following years. The second scenario used data until 2021, factoring in the pandemic's impact on mortality rates over the past two years. The third scenario considered the two years of COVID-19 and a hypothetical year of 2022, assuming that mortality levels would return to pre-pandemic rates (i.e., 2019). Moreover, the fourth scenario also included the two years of COVID-19 and a hypothetical year of 2022, basing mortality levels on the forecasted levels from the first scenario.

Considering the challenges faced during the COVID-19 vaccination campaign in Brazil, which began in 2021 (CASTRO, 2021; BARBERIA & GÓMES, 2020), it is likely that the effects of the pandemic may only be fully realized in 2023. To explore this further, we have applied the methodology used by the United Nations in their World Population Prospects (UNITED NATIONS, 2022) to two alternative scenarios, where mortality levels return to pre-pandemic levels in 2023 instead of 2022. Scenario five assumes two years of COVID-19 (2020 and 2021), followed by a hypothetical year in 2023 with mortality levels equal to those observed in 2019, and a hypothetical year in 2022 with mortality levels between those of 2021 and 2023. In this scenario, it is assumed that mortality levels will return to pre-pandemic levels in 2023, with 2022 being an intermediary year. Scenario six also assumes two COVID-19 years (2020 and 2021), but with a hypothetical year in 2023 based on the forecasted year from scenario one (following the same starting year iteration), and a hypothetical year in 2022 with mortality levels between those of 2021 and 2023. In this scenario, it is assumed that mortality levels will return to the expected levels for 2023 amidst the pandemic, with 2022 being an intermediary year. Finally, we recall here that for each of these six scenarios, we employ a variety of starting years for the fitting period (see previous subsection), resulting in 456 (76*6) different fitting periods for the LC method and 276 (46*6) fitting periods for the LM.

Table 2 summarizes all six scenarios. It is important to note that other mortality outcomes may occur in São Paulo city in the next years. However, by including the six scenarios into the forecasts, it was possible to account for some of the uncertainty regarding the mortality level's return post-pandemic.

Table 2 – Summary of the six scenarios considered for the return of mortality levels after the pandemic

Scenario	Data used	Assumption
I	Until 2019	We consider that the pattern observed before COVID-19 will continue to be seen until 2050.
II	Until 2021	We consider that there will be a shift in mortality level due to COVID-19 that won't be absorbed.
III	Until 2021 2022 = 2019	With a hypothetical year 2022 equal to 2019 we consider that mortality will return in 2022 to the same level observed in 2019.
IV	Until 2021 2022 = 2022 forecasted in I	With a hypothetical year 2022 equal to the forecast obtained in scenario I, we consider that mortality will return in 2022 to the level expected before COVID-19.
V	Until 2021 2022 = mean of 2021, 2023 2023 = 2019	With a hypothetical year 2023 equal to 2019 we consider that mortality will return in 2023 to the same level observed in 2019.
VI	Until 2021 2022 = mean of 2021, 2023 2023 = 2023 forecasted in I	With a hypothetical year 2023 equal to the forecast obtained in scenario I, we consider that mortality will return in 2023 to the level expected before COVID-19.

3.2.2.4. Final estimate and prediction intervals

To compute prediction intervals for life expectancy, we use a simulation approach for the forecasted k_ts. Specifically, we simulated the forecasted k_ts 100 times and computed the life expectancy at birth for each simulation over the forecasted horizon (2022-2050). This resulted in a total of 45,600 forecasts of life expectancy at birth between 2022 and 2050 for the LC method and 27,600 forecasts for the LM method for men and women separately. From these estimates, we obtained the median and the 2.5 and 97.5 quantiles. The quantiles allowed us to compute the 95% prediction interval for each forecasted year, with the median used as the final estimate.

Our final approach considered different baseline periods, six scenarios of survival recovery after the pandemic, and 100 simulations of the time-index parameter, allowing us to better account for the forecast uncertainty in the context of rapid changes in mortality levels. However, it's important to note that our analysis does not account for other potential mortality outcomes that may occur in São Paulo city in the coming years. The effect of the pandemic on mortality and other changes in mortality patterns induced by future shocks are yet to be predicted.

4 RESULTS

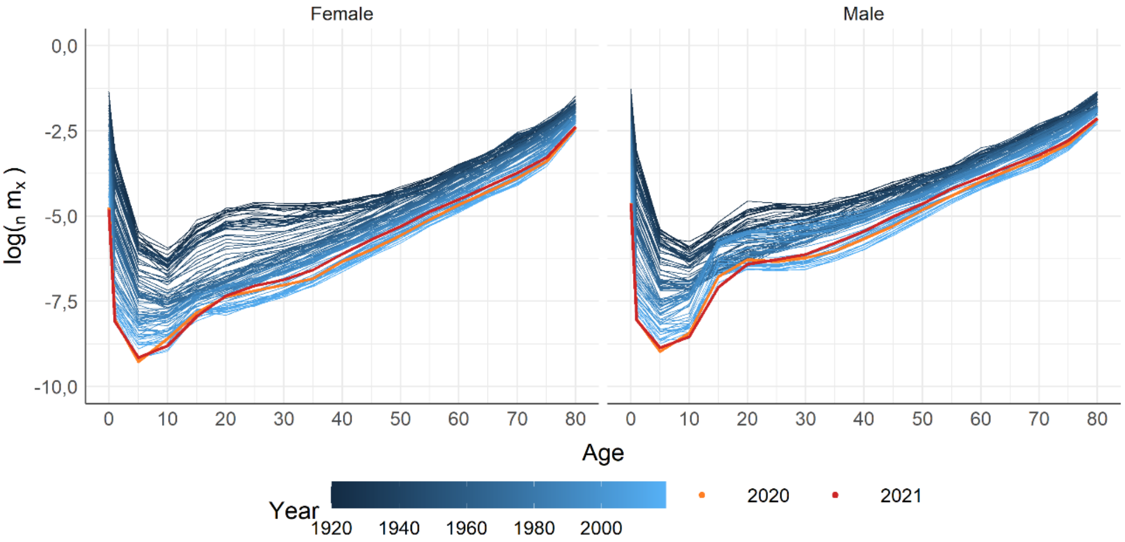
We present the results of this thesis in two parts, following the structure outlined in Chapter 3. The first sub-section contains a descriptive analysis of the changes in death rates and life expectancy since 1920. The results include comparisons with mortality trends for other populations, including the Best-Practice approach proposed by Oeppen and Vaupel (2002). We also decompose the contribution of each age group to the observed changes in life expectancy in São Paulo. The second sub-section shows forecasts for the life expectancy at birth until 2050.

4.1. Trends in mortality 1920-2021

4.1.1. Death rates and life expectancy at birth

Figure 3 shows the logarithm of age-specific death rates by sex for São Paulo from 1920 to 2021. The figure reveals a consistent mortality decline over the years, with the highest level observed in 1920. Comparing the age groups, the most considerable improvement in mortality, especially among females, occurred at younger ages (up to age 50). Our results align with previous research, which has consistently shown higher mortality rates among men than women. Also, the age patterns follow previous studies (EBELING, 2018; WILMOTH, 1997), with decreases over time that led to a structure of higher mortality rates at birth that decreases until around age ten and increases again with a hump between ages 15-30. This peak, more pronounced among men, is attributed to external causes of death, such as accidents and violence, which have become more prevalent in Brazil over the last decades, affecting especially men at young ages (BORGES, 2017).

Figure 3 – Log of age-specific death rates by sex, São Paulo city 1920-2021

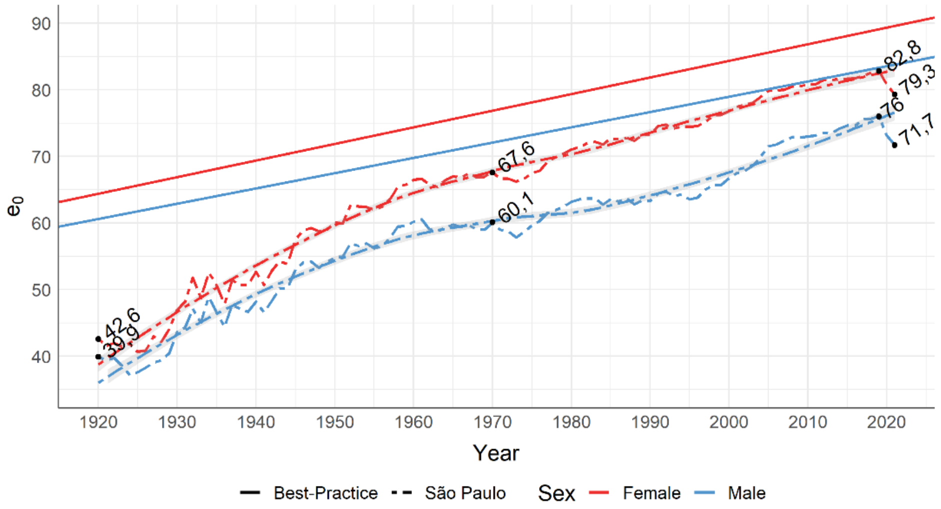


Source: SEADE, IBGE

Figure 3 highlights the years 2020 and 2021 in orange and red, respectively, to indicate the period affected by COVID-19. The pandemic disrupted the pattern of decreasing death rates, causing a significant increase in mortality, especially for those above 20 years old. It is evident that the year 2021 had a more substantial impact on mortality than 2020, with a higher number of reported deaths. However, with the start of vaccination in 2021, followed by a subsequent decline in fatalities in 2022, it is highly likely that mortality will revert to pre-pandemic levels in 2022 or the following years. This finding aligns with the report from CONASS in 2023. Regarding the age pattern of mortality, the pandemic changed the hump: the peak shifted to older age groups, reflecting variations in fatality rates across different ages, particularly the concentration of fatalities among older individuals (LEVIN et al., 2020).

Life expectancy at birth in São Paulo city has increased over time due to the decrease in age-specific death rates. This is shown in Figure 4, which reveals that life expectancy for both men and women has improved significantly since 1920. In that year, life expectancy was around 40 and 43 years for men and women, respectively, whereas in 2019, it was close to 76 and 83 years.

Figure 4 – Life expectancy at birth by sex, São Paulo city 1920-2021 and the Best-Practice slope (OEPPEN & VAUPEL, 2002)



Source: SEADE, IBGE, HMD

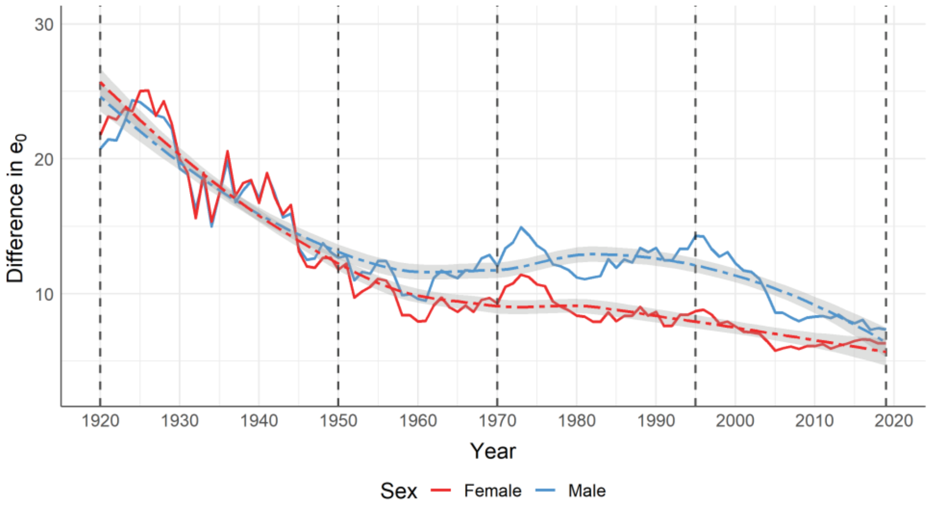
Nevertheless, over this period, life expectancy at birth did not follow a linear trend. As mentioned in the Methods section, there were at least four distinct periods of mortality decline. The first period, from 1920 to 1950, saw a rapid increase in life expectancy for both males and females, with convergence to Best-Practice levels (OEPPEN & VAUPEL, 2002) as depicted in Figure 5.

The second period, between 1950 and 1970, witnessed a deceleration in the increase of life expectancy, especially among men, resulting in a constant difference between São Paulo city and Best-Practice males' levels (Figure 5). Additionally, the divergence between men's and women's levels increased during this time (Figures 4 and 5).

The third period, from 1970 to 1995, saw a slight increase in life expectancy at birth, particularly among females who experienced steadier progress. This period also had the largest difference in life expectancy between the sexes. However, male life expectancy started increasing at a stable and faster pace after 1995, reducing the gap to Best-Practice (Figure 5) and narrowing the difference between male and female levels. Meanwhile, female life expectancy's pace of change at birth decelerated after 1995 (Figure 2).

After 2019, is possible to note decreases in life expectancy at birth due to the COVID-19 pandemic, as shown in Figure 4. Between 2019 and 2021, life expectancy at birth has dropped by more than four years for males and more than three years for females.

Figure 5 – Difference in life expectancy at birth to the Oeppen-Vaupel (2002) Best-Practice level by sex, São Paulo city 1920-2019



Source: SEADE, IBGE, HMD

Life expectancy at birth for women in São Paulo was already close to the estimated levels for men by Oeppen and Vaupel (2002) after 2000 (see Figure 4). This result expresses the relatively fast convergence of mortality levels to the ones in the wealthiest countries, and it is in line with results by Vallin and Meslé (2004). By 2019, the difference between São Paulo and Best-Practice life expectancy at birth was around six years for females and seven years for males.

São Paulo presented a complex and multifaceted mortality decline in the past century, with differences by sex and age groups. Recently, due to the COVID-19 pandemic, the mortality trend presented substantial variations, with decreases in life expectancy at birth. It remains unclear the lasting effects of the pandemic on mortality. Considering the vaccination rates and the reductions in deaths in 2022, it is plausible that mortality levels in the country will return to pre-pandemic levels in the coming years.

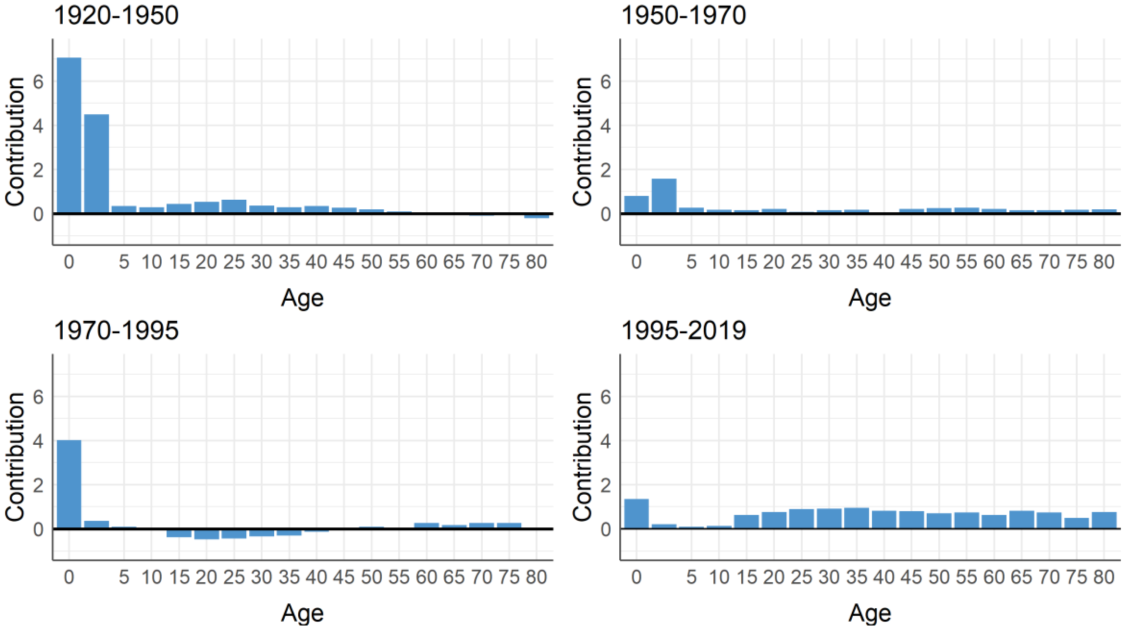
4.1.2. Age-decomposition of changes in life expectancy at birth

To examine the extent to which different age groups contributed to the changes in life expectancy at birth in each of the four discussed periods in São Paulo city, we applied Arriaga's (1984) decomposition for men and women separately. The results can be seen in Figures 6 and 7.

From 1920 to 1950, the mortality rate decreased at its fastest pace, and the age groups of 0-1 and 1-4 contributed the most to the increase in male life expectancy at birth. However, positive contributions can also be observed in the other age groups, except for the last group, where the contribution was negative. This could be due to the poor quality of declarations on old age in Brazil, as noted by Gomes and Turra in 2009 and Nepomuceno and Turra in 2020.

From 1950 to 1970, younger age groups made significant contributions, though to a lesser extent. Meanwhile, older age groups saw an increase in contributions. Between 1970 and 1995, a period of relatively stable increase in life expectancy at birth, we observed a negative contribution from young adults aged 15 to 45. This may be due to the increase in external causes of death among men in the 1980s (BORGES, 2017). However, during this same period, the first age group made a significant positive contribution to the increase in life expectancy at birth. From 1995 to 2019, all age groups between 15 and 80 years old made similar positive contributions to the increase in life expectancy at birth.

Figure 6 – Age-group contribution to change in life expectancy at birth by period, Males, São Paulo city



Source: SEADE, IBGE

The contribution of various age groups to the differences in life expectancy at birth for males in the four periods shifted from young to old ages. It is worth noting the distinct trend observed among young adults between 1970 and 1995, where mortality appeared to have increased. Figure 7 shows a similar shift from young to old-age patterns of contribution to the difference in life expectancy at birth for females across the four periods, with a more explicit shifting pattern. Old-age groups contributed significantly more than others between 1995 and 2019, and there was no negative contribution among the young adult age groups in any period. It is important to remember that the negative contribution to the 80+ age group between 1920 and 1950 was likely due to the poor quality of declarations regarding old age in Brazil (GOMES & TURRA, 2009; NEPOMUCENO & TURRA, 2020).

Figure 7 – Age-group contribution to change in life expectancy at birth by period, Females, São Paulo city

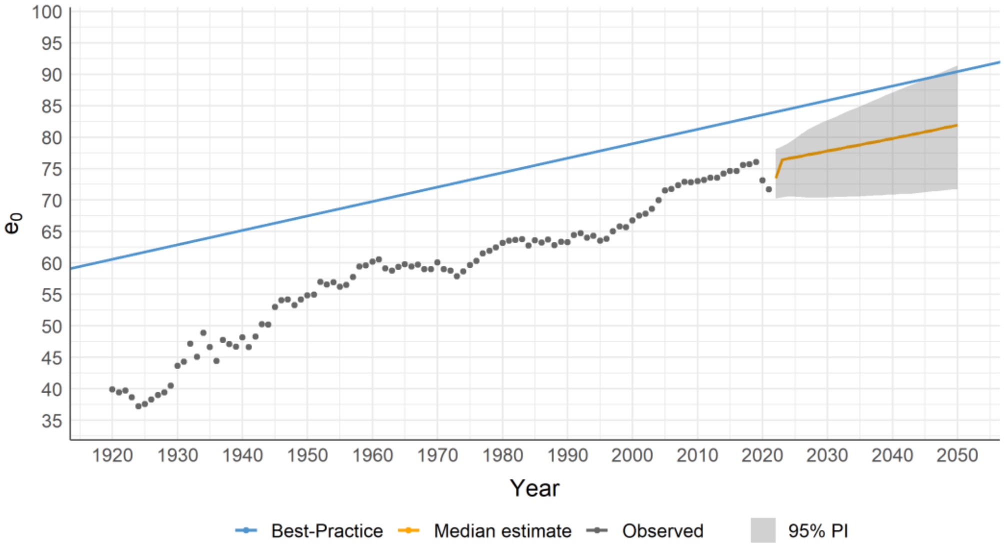
Source: SEADE, IBGE

The change in the age contribution to the differences in life expectancy over the period 1920-2019 highlights one of the limitations of the LC and LM in contexts of rapid changes in mortality trends, such as the one that occurred in São Paulo city in the last century. This result can indicate a varying b_x, or non-constancy of the changes in the mortality pattern over time and, therefore, must be accounted for in the forecasts of life expectancy at birth. In this analysis, we address this issue by applying the two forecasting methods to different baseline periods, varying from 1920 to 1995 until 2021, and combining the forecasted life expectancy obtained from different estimated b_xs.

4.2. Forecasts 2022-2050

We analyzed mortality trends up to 2021 and used the Lee-Carter (1992) and Lee-Miller (2001) methods to forecast life expectancy in São Paulo city until 2050. Our methodology which included accounting for different baseline periods and scenarios for the return of mortality levels after COVID-19 (see Methods), led to 73,200 estimates of life expectancy at birth for each year between 2022 and 2050. We combined the results into median and 95% prediction intervals for both males and females.

Figure 8 – Life expectancy at birth forecasts with 95% prediction intervals until 2050, Males, São Paulo city

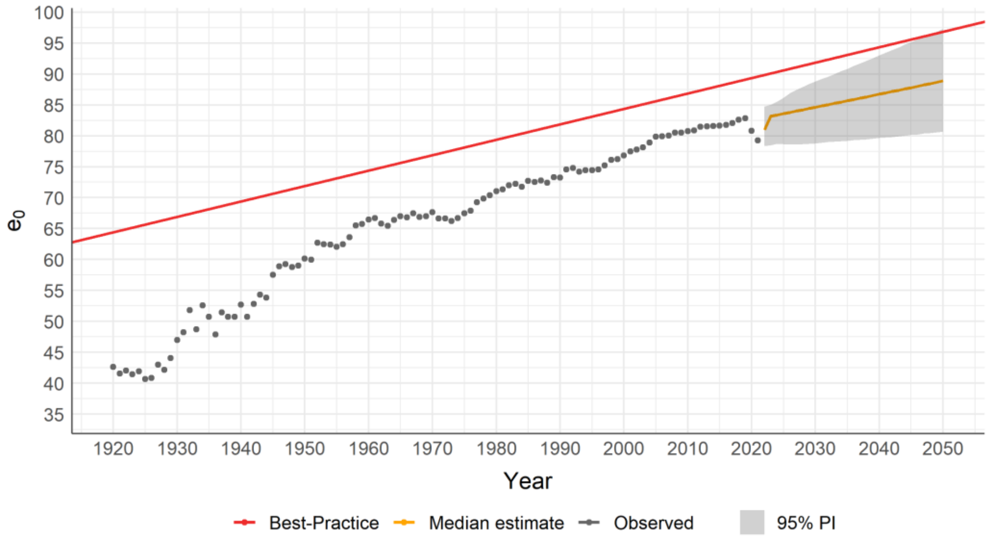


Source: SEADE, IBGE, HMD

Figure 8 shows forecasts for male life expectancies. The median forecast for 2022 is 73.6, with a 95% PI between 70.3 and 78.2. Although this is higher than the life expectancy observed in 2021, it still falls short of the pre-pandemic level. However, for 2023, the median estimate of 76,4 is close to the level seen in 2019, with a 95% PI ranging from 70.4 to 78.5. From 2023 onwards, the estimates show a relatively stable increase until 2050, with estimated life expectancy at birth of 81.9 and a 95% PI ranging from 71.7 to 91.3.

It is essential to note that the prediction intervals are quite broad. This expresses the uncertainty surrounding the life expectancy levels when the different baseline periods and levels of mortality post-pandemic are accounted for in the forecasts. Depending on the chosen baseline period and COVID-19 scenario, results with differences of up to 20 years in 2050 can be found within the 95% range.

Figure 9 – Life expectancy at birth forecasts with 95% prediction intervals until 2050, Females, São Paulo city



Source: SEADE, IBGE, HMD

The results for females are displayed in Figure 9. The median forecast for 2022 is 80.9, with a 95% PI ranging from 78.3 to 84.7. Like males, there is an increase in female life expectancy from the previous year, but it has not yet reached pre-pandemic levels. For 2023, however, the median estimate is 83.2, which is close to the 2019 level, and the 95% PI ranges from 78.5 to 85.1. From 2023 onwards, there is a steady increase in life expectancy until 2050, with estimated life expectancy at birth equal to 88.9 and 95% PI between 80.6 and 97.4.

Compared to males, the prediction intervals for females are narrower, indicating less uncertainty in the forecasts. This is likely due to the more consistent decrease in mortality over the last century and the shorter impact of COVID-19 on female mortality. The widest 95% prediction interval is estimated in 2050, with a range of approximately 17 years.

It should be noted here that our forecasts – despite the occurrence of COVID-19 – are significantly higher than those that have been made in the past (presented in Table 1). Specifically, we forecast that life expectancy at birth in 2050 for males and females will be 2.7 and 4.2 years higher, respectively, than what was predicted by the most recent forecast in 2013 (SEADE, 2013).

Also, within the 95% PI of our forecasts, both male and female São Paulo life expectancy could be reaching the Oeppen-Vaupel (2002) Best-Practice levels by 2045. For this analysis, we assumed that the Best-Practice increase will follow the same pace of increase until 2050. Annex 1 summarizes all median estimates and 95% prediction intervals for males and females between 2022 and 2050.

5 DISCUSSION AND CONCLUSION

The aim of this thesis was to forecast life expectancy at birth in São Paulo city until 2050 while taking into account the impact of the COVID-19 pandemic and the non-linear changes in mortality trends that have occurred over the past century. São Paulo has played a pioneering role in managing the pandemic in Brazil, and it also has a comprehensive historical record of age-specific mortality rates dating back to 1920. Therefore, using death rates for São Paulo was ideal for applying forecasting methods to analyze future survival levels in Brazil.

We used the Lee-Carter (1992) and Lee-Miller (2001) methods with death rates obtained from the combination of deaths gathered by the SEADE foundation (SEADE) and population collected by the Brazilian Institute of Geography and Statistics (IBGE) data. These methodologies have gained popularity among demographers due to their innovative and straightforward approach to mortality forecasting, although they are not without limitations. Most forecasting methods, including the LC and the LM, are extrapolative. Therefore, these methods are sensitive to the choice of the baseline period and fluctuations in the rate of mortality decline in the past.

As emphasized in our study, from 1920 to 2019, life expectancy at birth in São Paulo increased non-linearly. Moreover, there was a substantial reduction in life expectancy during the first two years of the COVID-19 pandemic. To mitigate the dependency on the fitting period's choice and better incorporate the effects of the recent mortality shock caused by COVID-19, we used different baseline periods, from 1920 and 1995 until 2021, and six scenarios for post-pandemic mortality levels. Also, we simulated the forecasted time-index parameters to compute prediction intervals.

Based on 73,200 simulations for each year between 2022 and 2050, we summarized the resulting life expectancy forecasts into median values and 95% prediction intervals. These outcomes improved the accuracy of life expectancy forecasts for São Paulo. We are unaware of any estimates that have accounted for pandemic effects and explored the non-linear changes in mortality patterns over the past century in São Paulo. Also, adding recent data enhanced the forecasts by considering the city's most recent mortality experiences. Despite the negative effects of the COVID-19 pandemic, our results presented significantly higher estimates than predicted by the most recent forecast in 2013 (SEADE, 2013). By 2050, we estimate that life expectancy at birth in São Paulo city will reach approximately 81.9 (95% PI ranging from 71.7 to 91.3) for men and 88.9 (95% PI ranging from 80.6 to 97.4) for women. Also, within the 95% PI, we estimated that by 2045, male and female life expectancy could reach the Best-Practice levels established by Oeppen and Vaupel (2002), extrapolated for the future.

Our forecasting approach – based on a combination of different fitting periods and future scenarios – is among the first attempts to forecast mortality in the presence of shocks. However, the long-term effects of the pandemic on health and mortality remain unclear. Although vaccination rates are increasing in São Paulo city, the worsening of the interactions of COVID-19 with other illnesses, such as influenza, or the occurrence of new variants is yet to be predicted. Therefore, it is essential to note that different mortality outcomes - not accounted for in our six scenarios - may occur in the city in the coming years.

Additionally, as seen with the decomposition of differences in life expectancy at birth over this period, there was a shift from young to old age contributions to survival levels. This result can indicate a non-constancy of the changes in mortality pattern over time or varying b_x. By including different baseline periods, starting from 1920 to 1995 until 2021, we were able to combine forecasts obtained over different estimated b_xs, one for each baseline period, and consequently, account for some possible changes in the age pattern over time. Future research could, however, focus on applying methods that account for variations in the b_x, such as Li, Lee, and Gerland (2013). These methods could be more appropriate for non-linear changes in mortality over time, such as in São Paulo city.

Finally, the analysis and forecasts presented here provide a valuable resource for policymakers and researchers working to address public health challenges and plan for the future well-being of the city's population. However, our forecasts share the same limitations as other extrapolation methods. Historical patterns may not hold in the future, and unpredicted structural changes, such as new mortality shocks, may occur.

REFERENCES

Andrasfay, T. & Goldman, N. (2021). Reductions in 2020 U.S. life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. Proc. Natl Acad. Sci. USA 118, e2014746118.

Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. Demography February 1; 21 (1): 83–96. doi: https://doi.org/10.2307/2061029.

Barberia, L. G. & Gómez, E. J. (2020). Political and institutional perils of Brazil's COVID-19 crisis. Lancet 396, 367–368.

Basellini, U. (2020). New Approaches in Mortality Modelling and Forecasting. Syddansk Universitet. Det Samfundsvidenskabelige Fakultet.

Basellini, U., Camarda, C. G., & Booth, H. (2022). Thirty years on: A review of the Lee–Carter method for forecasting mortality. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2022.11.002.

Berquó, E. S. (1974). A invasão de óbitos no município de São Paulo. Cadernos de Pesquisa - CEBRAP, São Paulo, v. 19, p. 1-157. Available in: <http://www.cebrap.org.br/index.asp?Fuseaction=Conteudo&ParentID=389&Menu=377,389,0,0&Materia=1614&str_busca>.

Bloom, D. E. and D. Canning. (2007). Mortality traps and the dynamics of health transitions, Proceedings of the National Academy of Sciences 104(41): 16044–16049.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. Population Studies, 56(3):325–336.

Booth, H., Hyndman, R. J., Tickle, L., and De Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. Demographic Research, 15:289–310.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. Annals of Actuarial Science, 3(1-2):3–43.

Borges, G. M. (2017). Health transition in Brazil: Regional variations and divergence/convergence in mortality, Cadernos de Saude Publica 33(8): 1–15.

Canudas-Romo V, García-Guerrero VM, Echarri-Cánovas CJ. (2014). The stagnation of the Mexican male life expectancy in the first decade of the 21st century: the impact of homicides and diabetes mellitus. J Epidemiol Community Health. 2015 Jan;69(1):28-34. doi: 10.1136/jech-2014-204237. Epub 2014 September 24. PMID: 25252678.

Castro, R. (2021). Vacinas contra a Covid-19: o fim da pandemia?. Revista de Saúde Coletiva, Rio de Janeiro, v. 31(1), e310100.

Castro, M.C., Gurzenda, S., Turra, C.M. et al. (2021). Reduction in life expectancy in Brazil after COVID-19. Nat Med 27, 1629–1635. https://doi.org/10.1038/s41591-021-01437-z.

Castro, M.C., Gurzenda, S., Turra, C.M. et al. (2023). Research Note: COVID-19 Is Not an Independent Cause of Death. Demography April 1 2023; 60 (2): 343–349. doi: https://doi.org/10.1215/00703370-10575276.

Conselho Nacional de Secretários de Saúde (CONASS). (2023). Painel Conass Covid-19. Available in: <https://www.conass.org.br/painelconasscovid19/>. Accessed 26 May 2023.

Dey, S., & Davidson, J. (2021). The determinants of non-COVID-19 excess deaths during the COVID-19 pandemic: A cross-country panel study. Studies in Microeconomics, 9, 196–226.

Del Rio, C., Collins, L.F., Malani, P. (2020). Long-term Health Consequences of COVID-19. JAMA. 2020;324(17):1723–1724. doi:10.1001/jama.19719.

Dowd, J.B., Andriano L, Brazel D.M., Rotondi V., Block P., Ding X., et al. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. PNAS; 117:9696-8.

Dyson, T. (2010). Population and Development. The Demographic Transition. Zed Books.

Ebeling, M. (2018). How Has the Lower Boundary of Human Mortality Evolved, and Has It Already Stopped Decreasing?. Demography 55, 1887–1903. https://doi.org/10.1007/s13524-018-0698-z.

Frenk J., Frejka T., Bobadilla J.L., et al. (1991). La transición epidemiológica en América Latina. Boletin de la Oficina Sanitaria Panamericana. Pan American Sanitary Bureau; 111(6):485-496. PMID: 1838685.

Fígoli, M. G. B. (1998). Modelando e projetando a mortalidade no Brasil. Revista Brasileira De Estudos De População, 15(1).

Fundação SEADE (SEADE). (2013). 459 anos de São Paulo Retrato da população hoje e as necessidades rumo aos 500 anos. Resenha de Estatísticas Vitais do Estado de São Paulo Ano 13 – nº 1. Available in: <https://economia.seade.gov.br/wp-content/uploads/sites/18/2021/12/SPDemografico-459-anos-sao-paulo-retrato-populacao-hoje-necessidades-rumo-500-anos.pdf>. Accessed 30 May 2023.

Fundação SEADE (SEADE). (2017). A População Regional Paulista em Perspectiva Histórica: projeções demográficas até 2050. Available in: <https://produtos2.seade.gov.br/produtos/midia/2017/04/SPDemografico_Num-01_2017-2.pdf>. Accessed 29 May 2023.

Fundação SEADE (SEADE). (2022). SP contra o novo coronavírus: Boletim Completo. Available in: <https://www.seade.gov.br/coronavirus/>. Accessed 17 December 2022.

Gomes, M. M. F., and Turra, C. M. (2009). The Number of Centenarians in Brazil: Indirect Estimates Based on Death Certificates." Demographic Research 20(20): 495–502.

Gonçalves, M. A. (1974). A invasão de óbitos no município de São Paulo. São Paulo: CEBRAP, 1974. (Cadernos CEBRAP, 19). Available in: <https://bibliotecavirtual.cebrap.org.br/arquivos/a_invasao_de_obitos_no_municipio.g.pdf >.

Hill, K. (1987). Estimating Census and Death registration Completeness. Asian and Pacific Population Forum, Honolulu, v. 1, n. 3, p.8, 23-13.

Hochman, G. (2011). Vacinação, varíola e uma cultura da imunização no Brasil. Ciência & Saúde Coletiva, 16(2):375-386.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis, 51(10):4942–4956.

Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. International Journal of Forecasting, 24(3):323–342.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2016). Procedimentos para obtenção de uma Tábua Completa de Mortalidade a partir de uma Tábua Abreviada – Brasil 2014. Available in: <https://ftp.ibge.gov.br/Tabuas_Completas_de_Mortalidade/Textos_metodologico_e_de_analise/Metodologia_para_transformar_uma_tabua_abreviada_em_completa_de_mortalidade.pdf>. Accessed 29 May 2023.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2018). Tábuas de Mortalidade 2010/2060. Available in: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html>. Accessed 30 May 2023.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2022). Tábuas Completas de Mortalidade para o Brasil 2021: Nota técnica n. 01/2022. Tábuas completas de mortalidade em ano de pandemia de COVID-19. Available in: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101981.pdf>. Accessed 29 May 2023.

Jesús-Adrián Alvarez, José Manuel Aburto & Vladimir Canudas-Romo (2020) Latin American convergence and divergence towards the mortality profiles of developed countries, Population Studies, 74:1, 75-92, DOI: 10.1080/00324728.2019.1614651.

Karlinsky, A. & Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. eLife 10, e69336.

Kirk, D. (1996). Demographic transition theory. Population studies, 50(3), 361-387.

Laprega, M. R. and Manço, A. R. X. (1999). Tendência histórica da invasão de óbitos em município da Região Sul do Brasil: período de 1936 a 1982”. Cad. Saúde Pública, Rio de Janeiro, v. 15, n. 3, p. 513-519, Jul./Set.

Lee, R. D. and Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality [with Rejoinder]. Journal of the American Statistical Association, 87(419):659–671.

Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. North American actuarial journal 4.1: 80-91.

Lee, R. D. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. Demography, 38:537–549.

Levin, A.T., Hanage, W.P., Owusu-Boaitey, N. et al. (2020). Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol 35, 1123–1138. https://doi.org/10.1007/s10654-020-00698-1.

Li N., Lee R., Gerland P. (2013). Extending the Lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. Demography. Dec;50(6):2037-51. doi: 10.1007/s13524-013-0232-2. PMID: 23904392; PMCID: PMC4550589.

Mackenbach, J. P. (1994). The epidemiologic transition theory. Journal of Epidemiology and Community Health; 48:329-332.

Marinho, F. et al. (2020). Excess mortality in Brazil: a detailed description of trends in mortality during the COVID-19 pandemic. Vital Strategies. https://www.vitalstrategies.org/resources/excess-mortality-in-brazil-a-detailed-descriptionof-trends-in-mortality-during-the-covid-19-pandemic/.

Matteson, D. and James, A. N. (2014) A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data, Journal of the American Statistical Association, 109:505, 334-345, DOI: 10.1080/01621459.2013.849605.

McMichael A.J., McKee M., Shkolnikov V., Valkonen T. (2004). Mortality trends and setbacks: global convergence or divergence? Lancet. 363(9415):1155-9. doi: 10.1016/S0140-6736(04)15902-3. PMID: 15064037.

Melo, C. M. L., Silva, G. A. S., Melo, A. R. S., and Freitas, A. C. (2020). COVID-19 pandemic outbreak: the Brazilian reality from the first case to the collapse of health services. Health Sciences. An. Acad. Bras. Ciênc. 92 (4). https://doi.org/10.1590/0001-3765202020200709.

Ministério da Saúde (Ministry of Health). (2022). Programa Nacional de Imunizações. Available in: <https://www.gov.br/saude/pt-br/coronavirus/vacinas/pni>. Accessed May 26 2023.

Nepomuceno, M. R. and Turra, C. M. (2020), The Population of Centenarians in Brazil: Historical Estimates from 1900 to 2000. Population and Development Review, 46: 813-833. https://doi.org/10.1111/padr.12355.

Nepomuceno, M. R., Acosta, E., Alburez-Gutierrez, D., Aburto, J. M., Gagnon, A., & Turra, C. M. (2020). Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. Proceedings of the National Academy of Sciences of the United States of America, 117(25), 13881–13883. https://doi.org/10.1073/pnas.2008760117.

Oeppen J., Vaupel J. (2002). Broken limits to life expectancy. Science (New York, NY) 296:1029–1031. https://doi.org/10.1126/science.1069675.

Omran, A. R. (1971). The epidemiologic transition: a theory of the epidemiology of population change. Milbank Mem. Fund Q. 49, 509–538.

Padhan, R., Prabheesh, K.P. (2021). The economics of COVID-19 pandemic: A survey. Economic Analysis and Policy 70: 220-237. ISSN 0313-5926. https://doi.org/10.1016/j.eap.2021.02.012.

Paes, N. A. and Albuquerque, E. E. (1999). Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras". Revista de Saúde Pública, São Paulo, v. 33, n. 1, p.33-43, fev.

Passarelli, H. A. (2022).Sociodemographic profiles and mortality patterns of foreign-born adults living in São Paulo, Brazil. [Unpublished master's thesis]. Federal University of Minas Gerais.

Perillo, S. R. & Perdigão, M. L. (2005). Percursos migratórios no estado de São Paulo: uma análise do período 1995-2000. São Paulo em Perspectiva, v. 19, n. 3, p. 97-109.

Pollard, J. H. (1987). Projection of age-specific mortality rates. Population Bulletin of the United Nations, 21-22:55–69.

Pollard, J. H. (1988). On the Decomposition of Changes in Expectation of Life and Differentials in Life Expectancy. Demography, 25(2), 265–276. https://doi.org/10.2307/2061293.

Prata, P. R. (1992). A Transição Epidemiológica no Brasil. Cad. Saúde Públ., Rio de Janeiro, 8 (2): 168-175, abr/jun.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). Demography: Measuring and modeling population processes. Oxford: Blackwell Publishers Ltd.

Ramírez-Soto MC, Ortega-Cáceres G, Arroyo-Hernández H. (2021). Sex differences in COVID-19 fatality rate and risk of death: An analysis in 73 countries, 2020-2021. Infez Med 29(3):402-407. doi: 10.53854/liim-2903-11. PMID: 35146345; PMCID: PMC8805484.

Ramsay J. O. and Silverman B. W. (2005), “Functional data analysis.” New York: Springer-Verlag, 2nd edition.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: Mathematics and Economics, 38(3):556–570.

Riffe T., Acosta E., the COVerAGE-DB team. (2021). Data Resource Profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths, International Journal of Epidemiology, Volume 50, Issue 2, Pages 390–390f, https://doi.org/10.1093/ije/dyab027.

Riffe, T., Nepomuceno, M.R., Basellini, U. (2020). Mortality Modeling. In: Gu, D., Dupre, M. (eds) Encyclopedia of Gerontology and Population Aging. Springer, Cham. https://doi.org/10.1007/978-3-319-69892-2_556-1.

Riley, J. C. (2001). Rising Life Expectancy: A Global History. Cambridge University Press. https://doi.org/10.1017/CBO9781316036495.

Schnürch, S., Kleinow, T., Korn, R., & Wagner, A. (2022). The impact of mortality shocks on modelling and insurance valuation as exemplified by COVID-19. Annals of Actuarial Science, 16(3), 498-526. doi:10.1017/S1748499522000045.

Schöley, J., Aburto, J.M., Kashnitsky, I. et al. (2022). Life expectancy changes since COVID-19. Nat Hum Behav. https://doi.org/10.1038/s41562-022-01450-3.

Schwartz, J. (2000). Harvesting and long-term exposure effects in the relation between air pollution and mortality. American Journal of Epidemiology, 151, 440–448.

Siegel, J. S.; Swanson, D. A. (2004). The methods and materials of demography. 2. ed. San Diego, Califórnia: Elsevier Academic. 732p.

Silva, F. S. (2009). Tendências e projeção da mortalidade do município de São Paulo – 1920 a 2100. [Unpublished master’s thesis]. Federal University of Minas Gerais.

Simões, C. C. S. (2001). Perfis de saúde e de mortalidade no Brasil: uma análise de seus condicionantes em grupos populacionais específicos. Brasília, 1ª edição.

Siviero, P. C. L. (2009). Níveis e padrões do diferencial de mortalidade por sexo no município de São Paulo, 1920 – 2005. [Unpublished master’s thesis]. Federal University of Minas Gerais.

Souza, L. G. (2023). Diferencial de gênero na mortalidade no município de São Paulo, 1920 a 2020: padrões por idade e causas de morte. [Unpublished dissertation]. Federal University of Minas Gerais.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. Nature, 405(6788):789–792.

Tuljapurkar, S. & Boe, C. (1998). Mortality change and forecasting: how much and how little do we know? North American Actuarial Journal, Schaumburg, v. 2, n. 4, p. 13-47.

United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022. Retrieved from https://population.un.org/wpp/.

Vallin, J. & F. Mesle. (2004). Convergences and divergences in mortality: A new approach to health transition, Demographic Research 10(2): 11–44.

Vallin, J., & Meslé, F. (2009). The Segmented Trend Line of Highest Life Expectancies. Population and Development Review, 35(1), 159–187. http://www.jstor.org/stable/25487645.

Vasconcelos, A M. (1998). Qualidade de estatísticas de óbito no Brasil: uma classificação das unidades da federação. Revista Brasileira de Estudos de População, Campinas, v. 15, n. 1, p. 115-124, jan./jul.

Waldvogel, B. C. et al. (2003). Projeção da população paulista como instrumento de planejamento. São Paulo em Perspectiva. São Paulo, v. 17, n.3/4, p. 67-79.

Whelpton, P. K. (1928) Population of the United States, 1925 to 1975. Am. J. Sociol., 34, 253-270.

Wilmoth, J. R. (1997). In Search of Limits: What Do Demographic Trends Suggest about the Future of Human Longevity, in Kenneth W. Wachter and Caleb E. Finch, eds., Be- tween Zeus and the salmon: The biodemog- raphy of longevity. Washington, DC: National Academy Press, pp. 38-64.

Wilmoth, J. R. (2000). Demography of longevity: past, present and future trends. Experimental Gerontology, Berkeley, v. 35, n. 9/10, p. 1111-1129.

Yoshikawa M., & Asaba K. (2021). Educational Attainment Decreases the Risk of COVID-19 Severity in the European Population: A Two-Sample Mendelian Randomization Study. Front Public Health 9:673451. doi: 10.3389/fpubh.2021.673451. PMID: 34150709; PMCID: PMC8212884.

ANNEXES

Annex 1 – Summary of life expectancy estimates and 95% prediction intervals by sex, São Paulo city 2022-2050

Year	Males			Females		
	Median	**95% Lower bond**	**95% Upper bond**	**Median**	**95% Lower bond**	**95% Upper bond**
2022	73,58	70,29	78,16	80,88	78,31	84,74
2023	76,41	70,40	78,54	83,17	78,48	85,13
2024	76,62	70,66	79,07	83,40	78,70	85,51
2025	76,81	70,56	79,84	83,60	78,61	86,17
2026	77,01	70,49	80,66	83,81	78,59	86,85
2027	77,21	70,41	81,29	84,03	78,59	87,43
2028	77,40	70,40	81,82	84,23	78,66	87,90
2029	77,60	70,36	82,25	84,44	78,71	88,34
2030	77,80	70,39	82,70	84,64	78,77	88,75
2031	77,99	70,42	83,15	84,85	78,84	89,17
2032	78,19	70,46	83,61	85,04	78,89	89,61
2033	78,40	70,57	84,01	85,26	78,95	90,04
2034	78,60	70,57	84,42	85,46	79,05	90,44
2035	78,80	70,57	84,88	85,68	79,13	90,84
2036	79,01	70,62	85,31	85,88	79,25	91,28
2037	79,22	70,71	85,76	86,10	79,32	91,71
2038	79,42	70,79	86,20	86,31	79,47	92,12
2039	79,62	70,84	86,66	86,53	79,53	92,55
2040	79,83	70,88	87,06	86,74	79,62	92,98
2041	80,03	70,96	87,48	86,96	79,72	93,46
2042	80,24	70,99	87,92	87,18	79,82	93,88
2043	80,43	71,11	88,30	87,39	79,90	94,31
2044	80,66	71,19	88,78	87,60	80,03	94,78
2045	80,85	71,29	89,15	87,81	80,10	95,17
2046	81,05	71,34	89,58	88,03	80,21	95,59
2047	81,26	71,47	90,03	88,25	80,32	96,02
2048	81,46	71,52	90,47	88,47	80,43	96,48
2049	81,67	71,62	90,88	88,69	80,54	96,91
2050	81,89	71,7	91,34	88,91	80,63	97,35