university of groningen

faculty of spatial sciences

Geographical inequalities in life expectancy across US States between 1999 and 2020

Assessing contributions of age groups and causes of death

Name	**Bas Koopman**
Student number	**S4454715**
Email address	**B.m.koopman.1@student.rug.nl**
Supervisor	**Dr. A.P.P. Remund**
Course name	**Master Thesis Population Studies**
Master	**Population studies**
Institution	**Population Research Centre, University of Groningen**
Date	**5 July 2022**
Place	**Groningen**

Abstract

Life expectancy in the United States is stalling in the period 1999 to 2020. In this period, US mortality rates were increasing across states, however, not all states have the same developments in life expectancy. In the southeast and up northwards are relatively low values and, in the north-central region and southwest relatively high values of life expectancy. This thesis aims to describe how US life expectancy has developed and which age groups and causes of death contributed the most to developments in life expectancy, which are analyzed using age- and cause-specific decomposition models. The results show that for the states with the lowest life expectancy one pattern in the mortality disadvantage is visible in the age groups. The states with the highest life expectancy show two patterns in the age groups in reaching the mortality advantage. The causes of death which contribute the most to the gap in life expectancy show similarities between the states with the highest and lowest life expectancy, in which especially causes of death I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) are big contributors to the gap, however, also all states have different other causes of death which contribute to the gap. Geographical inequality has increased when looking at the states with the highest and lowest life expectancy, which is related to differences in socioeconomic status, policies, context, composition and lifestyle. Policies to improve health should be focused on state-specific mortality rates to increase national life expectancy again.

Keywords: life expectancy, United States, cardiovascular diseases, deaths of despair, geographical inequality, old age groups, midlife age groups

Preface

Before you, the Thesis "*Geographical inequalities in life expectancy across US States between 1999 and 2020*". I chose this topic for my thesis because the news reported about stalling and increasing life expectancy, especially during the Covid-19 pandemic. However, most of the time no attention is paid to which causes of death and age groups are behind the changes in life expectancy. Therefore, I was interested to look into life expectancy developments to see which causes of death and age groups are the biggest contributors to changes and differences in life expectancy.

Writing the thesis was a journey with ups and downs and many hours invested. In the end, it was all worth it, because I got an interesting view of how life expectancy developed and how national life expectancy is influenced by geographical inequalities and that differences between states, still could lead to an almost similar life expectancy.

I would like to thank dr. Adrien Remund for his support, help with R, quick reactions to questions and feedback during the process of writing my thesis, I would also like to thank my parents and friends for being so supportive in the process of writing the thesis.

I hope you enjoy reading this thesis,

Bas Koopman July 2022

Table of contents

Abstract I

Preface II

List of figures V

List of tables VI

List of abbreviations VII

1. Introduction 1

1.1 Background 1

1.2 Academic relevance 2

1.3 Research objective 2

1.4 Research questions 3

1.4.1 Main research question 3

1.4.2 Subquestions 3

1.5 Structure 3

2. Theoretical framework 4

2.1 Theoretical background 4

2.1.1 Health transition 4

2.1.2 Epidemiological transition theory 4

2.1.3 Divergence and convergence in mortality 5

2.2 Previous studies 6

2.2.1 Previous evolution of US Life expectancy 6

2.3 Conceptual model 7

2.4 Hypotheses 8

3. Data and methodology 9

3.1 Study design 9

3.2 Setting 9

3.3 Databases 10

3.3.1 CDC Wonder database 10

3.3.2 WHO Mortality Database 11

3.4 Data validation 12

3.6 Ethical considerations 14

3.7 Methods 14

3.7.1 Life expectancy 14

3.7.2 Life table analysis 14

3.7.3 Decomposition 15

4. Results and discussion 17

4.1 Geographical inequalities 17

4.1.1 Convergence and divergence 18

4.1.2 Lowest and highest life expectancy 19

4.2 Females laggard *21*
4.3 Females pioneers *23*
4.4 Males laggard *25*
4.5 Males Pioneers *26*
5. Conclusion **34**
5.1 Limitations *34*
5.2 Further research and policy recommendations *35*
References **36**
Appendix 1 Proportion minimum and maximum suppressed cells **45**
Appendix 2 Regression output suppressed cells **49**
Appendix 3 Map of the US states **50**

List of figures

Figure 1: Life expectancy total population US between 1999 and 2020 2
Figure 2: Conceptual model 8
Figure 3: US state life expectancy 2020 17
Figure 4: Gap in life expectancy in years in the period 1999 to 2020 18
Figure 5: States with the lowest and highest life expectancy in the period 1999 to 2020 20
Figure 6: Age decomposition of the difference in life expectancy with national average females 30
Figure 7: Cause specific decomposition of the difference in life expectancy with national average females 31
Figure 8: Age decomposition of the difference in life expectancy with national average males 32
Figure 9: Cause specific decomposition of the difference in life expectancy with national average males 33
Figure 10: Map of the US states 50

List of tables

Table 1: Age groups 9
Table 2: Causes of death according to ICD-10 main chapters 11
Table 3: Percentage of deaths known by causes of death 13
Table 4: Proportion minimum and maximum suppressed cells for every age groups in a year for females 45
Table 5: Proportion minimum and maximum suppressed cells for every age groups in a year for males 47
Table 6: Regression output 49

List of abbreviations

Abbreviation	Meaning
US	United States of America
ICD-10	International Classification of Diseases, tenth revision
OECD	Organisation for Economic Co-operation and Development
CDC	Centers for Disease Control and Prevention
WHO	World Health Organization

1. Introduction

1.1 Background

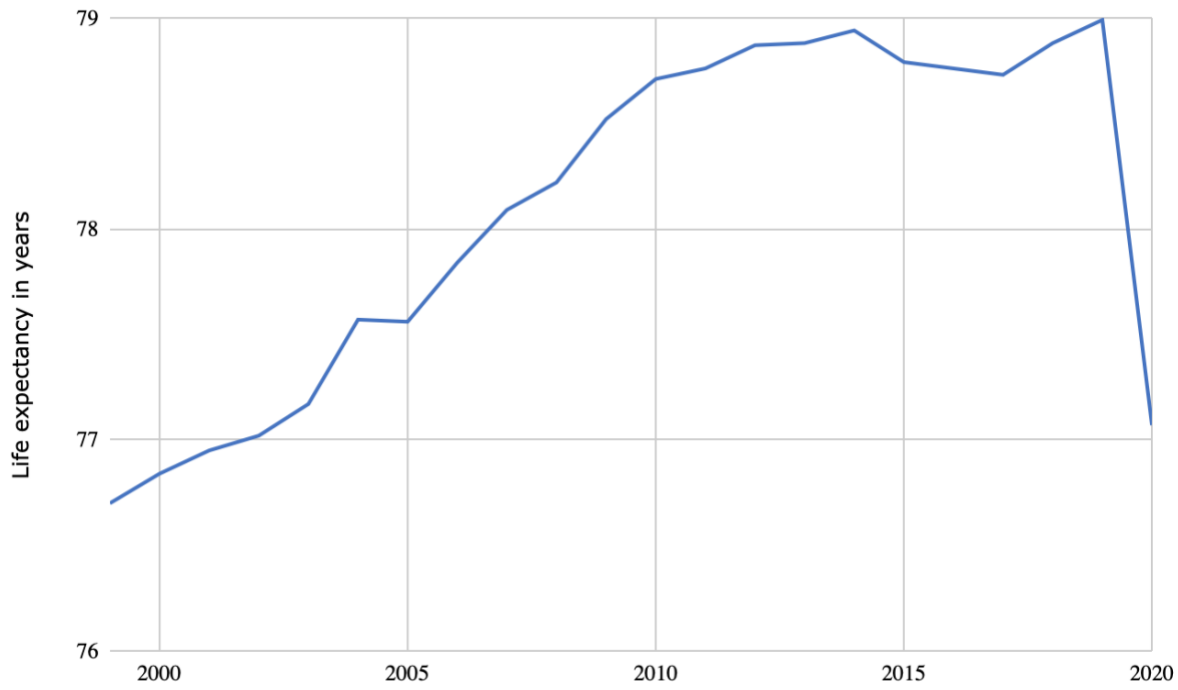
In the United States, life expectancy at birth increased from 76.7 to 77.07 years between 1999 and 2020 (figure 1), meaning that it has almost come to a halt (Murphy et al., 2021; Roser et al., 2013). Even before the drop of 1.92 years due to Covid-19 in 2020, life expectancy already stalled in the period 2014 to 2018 (Human Mortality Database, n.d.; Murphy et al., 2021). Figure 1 shows fluctuations in the development of life expectancy in the US, which displays that, before Covid-19, mortality rates already were increasing. During this period, the main contributors related to this increase were drug overdoses, alcohol abuse, suicide and organ system diseases (Woolf & Schoomaker, 2019).

In the period 1968 to 2019 in every age group, a dispersion of death rates is visible between US states (Couillard et al., 2021). The dispersion of death rates has not had the same pace and influence on life expectancy for males and females. This research focuses mainly on the development of life expectancy in the states with the lowest and highest life expectancy to see how these states contributed to the geographical inequality in life expectancy. In the southeast and up northwards are relatively low values of life expectancy and in the north-central region and southwest relatively high values of life expectancy. Furthermore, not in all states in the US changes in life expectancy are caused by the same cause of death or age group. Differences in the development of life expectancy could be explained by the following factors: migration, socioeconomic factors, lack of access to health care, poor-quality health care, behaviour, metabolic and environmental risks (Crimmins et al., 2011; Dwyer-Lindgren et al., 2017; Wang et al., 2013).

Judging the level of life expectancy requires a reference point, which entails some level of arbitrariness. The reference point in this research is the national life expectancy in the US, to which the state-level life expectancy is compared. The national life expectancy is chosen as a reference point because in every US state the data collection method is similar. Every state is using the ICD-10 chapters to collect information about underlying causes of death. Moreover, the health care system is similar across the US States. Besides that, the US has a large population which means that it reduces the risk of stochasticity, which could be present in small populations. Therefore, the 50 states and the District of Columbia are used in this research. Comparing different countries could influence the results because the data could be collected in different ways, health care systems could be different and different genes and common diseases could be present (Vallin & Meslé, 2008). Heterogeneity also exists among the US states in common diseases and genes.

To sum up, US life expectancy has changed, however, not in every state in the same way. Therefore, it is important to compare the national life expectancy with the state-level life expectancy to see what causes the differences. The outcome of this research could be relevant for policymakers because they could see which specific causes of death need attention and in which age groups the most profits are achievable. Furthermore, the research could help policymakers to give an overview of which states have the highest and lowest life expectancy, what causes this and which approaches could work best to improve life expectancy. Therefore, policies could be made and this research could be used to drive progress in life expectancy again and decrease the geographical inequalities.

Figure 1: Life expectancy total population US between 1999 and 2020



Source: Own figure based on (Human Mortality Database, n.d.; Murphy et al., 2021).

1.2 Academic relevance

From 1800 till the end of the 20th century life expectancy has increased which is related to the health transition. Due to several developments, like economic modernization and rising standard of living mortality rates decreased and life expectancy increased (Riley, 2001). The epidemiological transition theory shows how the pattern of common diseases has changed and has changed the mortality rates (Santosa et al., 2014; Omran, 1971). However, as can be seen in figure 1, recently life expectancy in the US is stalling. Among the different US states, life expectancy developments are different. In some states like Alaska and California, life expectancy is still increasing and in other states like West Virginia life expectancy is declining (Barbieri, 2021; Woolf & Schoomaker, 2019). Looking at geographical inequalities is an important aspect of how the national life expectancy develops because it shows divergence or convergence across US states. Besides that, looking at geographical differences also shows that life expectancy does not have the same developments in every state at the same time (Vallin & Meslé, 2004).

This research aims to be relevant by providing information about how US state life expectancy has developed by focusing on the states with the lowest and highest life expectancy compared to the national life expectancy and how geographical inequality has developed over time by looking at the whole population, in the states with the highest and lowest life expectancy in the period 1999 to 2020. This research is trying to fill this research gap by analyzing life tables and using decomposition models.

1.3 Research objective

The objective of this research is to analyse how life expectancy has developed in the US states with the lowest and highest life expectancy and which age groups and causes of death

contributed the most to these developments. Besides this, the research aims to see which causes of death and age groups drive the geographical inequality in life expectancy in the US.

1.4 Research questions

1.4.1 Main research question

What drives changes and geographical inequalities in life expectancy across US states?

1.4.2 Subquestions

1. How has the life expectancy changed across US states, and especially the gap between states with the lowest and highest life expectancy, between 1999 and 2020?
2. Which age groups contribute the most to the differences in life expectancy in US states with the lowest and highest life expectancy between 1999 and 2020?
3. Which causes of death contribute the most to the differences in life expectancy in US states with the lowest and highest life expectancy between 1999 and 2020?
4. What makes specific US states laggards or pioneers in life expectancy between 1999 and 2020?

1.5 Structure

Chapter 2 discusses the literature and theories to construct a conceptual model. Based on the conceptual model and literature hypotheses are formulated. In chapter 3 the data and the methods used in this research are described. In chapter 4, the results of the research and discussion are described. In the final chapter, the conclusion and recommendations are stated.

2. Theoretical framework

2.1 Theoretical background

2.1.1 Health transition

From 1800 till the end of the 20th century, life expectancy at birth increased from approximately 30 years to approximately 67 years. Already in the 18th century, the change in life expectancy started, first in Northwestern Europe with some pioneering countries. In most countries before the 19th century life expectancy was stable at 30 years. In the 19th century life expectancy began to increase and continued to increase in Northwestern Europe. The increase in life expectancy in the 19th century is related to a decline in death rates, however, death rates did not decline for infant mortality or mortality above 65 years (Riley, 2001). For the 18th century and before the 18th century is not much data available, which makes it difficult to say something about the trends in life expectancy before the 18th century. This increase in life expectancy is called the health transition. The health transition is not a general pattern for all countries. The pattern of the transition is different in every country, however, common for all countries is that mortality rates declined and life expectancy increased. The latter the health transition began in a country, the more compressed the transition is.

Many studies looked into the causes of the health transition. The causes that engender the change were economic modernization, medical knowledge, and a rising living standard which lead to a change in nutrition, although, the respective weight of these explanations is still disputed. The complicated interrelationships between these changes make it difficult to identify the factors behind the changes. Therefore, scholars looked at the changes in causes of death to see how life expectancy has developed (Colgrove, 2002; Riley, 2001). In this research, the change in disease pattern will be the main cause to research the changes in life expectancy. In the end, many countries showed a change in the disease pattern, which lead to a change in the mortality pattern (Riley, 2001). The changes in the disease pattern are also described in the epidemiological transition theory.

2.1.2 Epidemiological transition theory

The epidemiological transition theory was developed by Omran. The theory outlines the changing pattern in leading causes of diseases and the interaction between the changing patterns. Moreover, the epidemiological transition theory focuses on the economic, demographic and sociologic determinants and consequences of the pattern changes. The theory consists of five stages in which there is a shift in the disease and mortality pattern. During this shift, there is a shift from infectious diseases to degenerative and man-made diseases (Santosa et al., 2014; Omran, 1971).

The first stage is the age of pestilence and famine. This stage has high mortality levels which are fluctuating because of living conditions, epidemics and famines. The life expectancy in this stage is on average between 20 and 40 years. In this stage, most causes of death are infectious diseases (Santosa et al., 2014; Omran, 1971).

In the second stage, the age of receding pandemics mortality levels declines and life expectancy increases which is caused by sanitation improvements and medical breakthroughs. The life expectancy is between 30 and 50 years (Omran, 1971). During this stage infectious diseases are still an important cause of death, however, non-communicable diseases are also increasing (Santosa et al., 2014).

The third stage is the age of degenerative and man-made diseases, in this stage, mortality levels continue to decline and become stable at a low level. During this period life expectancy increases to above 50 years. The dominant cause of disease in the third stage are non-communicable diseases (Santosa et al., 2014; Omran, 1971).

The fourth stage of the epidemiological transition theory is called the age of declining cerebrovascular mortality, ageing, lifestyle modifications and resurgent diseases, in this stage life expectancy increases up to 80-85 years. During this period, the levels of mortality caused by cardiovascular diseases declined and stabilized because of improved medical care and lifestyle changes (Santosa et al., 2014).

The fifth stage is called the age of aspired quality of life, with paradoxical longevity and persistent inequities. This stage is about disease control, health promotion and an increase in healthy life expectancy. During this stage, new diseases will emerge. Furthermore, there will be inequalities between people and countries. In this stage, depression is becoming an epidemic because of our stressful lifestyles and competitiveness according to Omran (1998).

Some scholars are mentioning a combination of causes that lead to a change in mortality patterns. Others argue that every country has its causes that lead to changes in death rates and doubt the universality of the epidemiological transition theory because of high variation within stages of the theory, which makes it hard to use the epidemiological transition theory for making predictions. Besides that, the later stages of the model are not universal for all countries because multiple ways to reach low mortality levels are possible (Sudharsanan et al., 2022).

2.1.3 Divergence and convergence in mortality

The epidemiological transition theory by Omran expected a convergence in life expectancy. However, in practice, the convergence did not happen, for example, in the health crisis in Eastern Europe and AIDS crisis in Africa. This demonstrates that not all countries have the same developing pattern and that in some periods, countries experience rises and other countries stagnation or decreases at different paces in life expectancy which could lead to convergence or divergence in life expectancy. Vallin & Meslé (2004) created a model based on the health transition by Frenk et al. (1991). The model consists of three stages, which are based on the epidemiological transition theory, the cardiovascular revolution and the ageing process. In each stage, the transition generates firstly divergence in life expectancy between countries that pick up the innovation earlier and those who lag, and when the innovation is diffused across all countries, the laggards catch up and thus there is convergence in life expectancy (Vallin & Meslé, 2004).

The first stage of the model is similar to the first three stages of the epidemiological transition theory. During this stage, infectious diseases were replaced by chronic diseases. At the beginning of this stage, life expectancy is diverging because some states were pioneers in the epidemiological transition and others followed, which lead to a stage of convergence. At the beginning of the 20th century, the world has a difference of 27 years in life expectancy, which declined to 5 years in 1960 (Vallin & Meslé, 2004).

The second stage consists of a cardiovascular revolution, which started after the 1960s. In this stage, the western world had rapid progress and the eastern countries had a stagnation or a decrease in life expectancy which leads to divergence in life expectancy. This divergence was related to the political and socio-economic systems. After the end of the Cold War, the Eastern countries started to make progress again in their life expectancy which leads to a convergence in life expectancy (Vallin & Meslé, 2004).

In the third stage, Vallin and Meslé (2004) look at ageing in societies which also leads to a new process of divergence and convergence. The new process of divergence and convergence is related to the fact that some countries have already started to battle against an ageing society and other countries are lagging in this process.

However, not all differentials in life expectancy will disappear due to convergence. People's health is also affected by individual characteristics (composition) and by area characteristics (context) (Leyland & Groenewegen, 2020). The composition effect is about how the population is made up, for example, how many people are smoking, and how many people

have a high income. The context is about the setting in which people live, for example, urban or rural or if the area is affected by pollution (Hazen & Anthamatten, 2020). The combination of both factors influences how healthy people are and both factors have an influence on mortality levels. Composition and context could also be an explanation for geographical inequalities in life expectancy. In this research, however, composition and context are not included in the data, thus the reader should be aware of this by drawing conclusions based on the geographical inequalities described in the paper.

2.2 Previous studies

2.2.1 Previous evolution of US Life expectancy

Many researchers have looked into US life expectancy and geographical inequalities in life expectancy, which gives many different views on which age groups and which causes of death contributed to the stalling life expectancy. Focusing on the geographical inequalities helps to shed light on the stalling life expectancy because it could explain which causes of death and age groups contributed the most to the stalling life expectancy in different states.

US life expectancy has not kept the same pace of improvement as other countries in the 1980s. In the US in the 1980s, the death rates were already quite low and since then no big improvements were visible. In the 21st century, the stagnation in the death rates was especially visible for women older than 80 years. However, some trends still show that death rates are declining for the older age groups (80 years and older) (Rau et al., 2008).

Rau et al. (2008) found that the number of octogenarians and nonagenarians increased, because of survival improvements in younger age groups and a decline in mortality levels at age 100 years and older. Furthermore, they found that these improvements did not follow a universal pattern. This could be related to the health transition, and the states that are lagging were not able to go to the next stage (Rau et al., 2008).

The developments in US life expectancy have a different trend than other OECD countries (Harper et al., 2021). In 2010 the life expectancy in the US stopped increasing and since 2014 life expectancy started to decrease. This was caused by an increase in all-cause mortality among young and middle-aged adults. Contrary, a decline in infant and early adolescent mortality levels and for the older age group of 65 years and older were visible. The main causes for the increase in mortality are for example drug overdoses, hypertensive diseases and organ system diseases and injuries (Muennig et al., 2018; Mokdad et al., 2018; Woolf & Schoomaker, 2019; Woolf & Aron, 2018). Deaths caused by drug overdoses, alcohol abuse and suicide are also called deaths of despair (Couillard et al., 2021).

Besides that, compared to the OECD countries, also in the younger age groups, the US is performing worse, which is related to more violent deaths and motor vehicle accidents (Harper et al., 2021). Although, not all researchers agree on what causes the stalling life expectancy. Research by Mehta et al. (2020) shows that by looking at the total population, not deaths of despair but cardiovascular diseases are the main cause of the stalling life expectancy. The reason for this different conclusion is that the research by Mehta et al., (2020) looks at all age groups, whereas Woolf & Schoomaker (2019) only looked at midlife mortality.

Looking at the different US states, differences in life expectancy have increased since the 1980s. The increase in difference is related to different regional mortality trends, which are related to more and less advantaged areas of the country, in which the more advantaged areas had a greater increase in life expectancy than the less advantaged areas (Crimmins et al., 2010). For example, in the south of the US mortality improvements were slower in the 2000s compared to other areas of the country (Harper et al., 2021). The phenomenon of lower life expectancy in the south is particularly present in the states of Alabama, Kentucky, Mississippi and Tennessee (Fenelon, 2013). On the other hand, in the 2000s the life expectancy in Alaska and California had increased (Barbieri, 2021; Woolf & Schoomaker, 2019).

The divergence in life expectancy across US states could be explained by different causes. One of the explanations for the increase in divergence of US life expectancy is the increase in midlife mortality levels, which did not take place in every state at the same pace and in some states, in all age groups, the mortality levels increased (Barbieri, 2021; Woolf & Schoomaker, 2019). The divergence in mortality rates in all age groups could be explained by differences in education levels, labour market prospects and rising income differences. States with an on average higher educated population have lower mortality rates. The research by Couillard et al., (2021), concludes that the divergence in midlife mortality could be explained by decreases in mortality according to educational attainment. Furthermore, differences in how high- and low-income states deal with health policies increased the divergence in mortality levels (Couillard et al., 2021).

Moreover, research by Whitmore Schanzenbach et al., (2016) shows that divergence in life expectancy is caused by that, not all health measures reach everyone in the population at the same time. An example is smoking behaviour, which is higher in older cohorts than in younger cohorts, which also could be the explanation that in older age groups mortality levels for diseases caused by smoking and cancers caused divergence.

Another factor that causes divergence in life expectancy is opioid overdoses, especially in the Northeast and East North Central regions of the US. Besides that, a decline in the decrease of cardiovascular diseases also caused divergence in life expectancy in the US (Harper et al., 2021).

In conclusion, the literature shows that different explanations are possible for the stagnating US life expectancy, notably, the literature shows the pattern that the decrease in cause-specific mortality rates is not found in all age groups. In different age groups, there could be different causes of death generating the stalling in life expectancy. However, some findings show a small improvement in death rates in older age groups, and other research does not show these improvements.

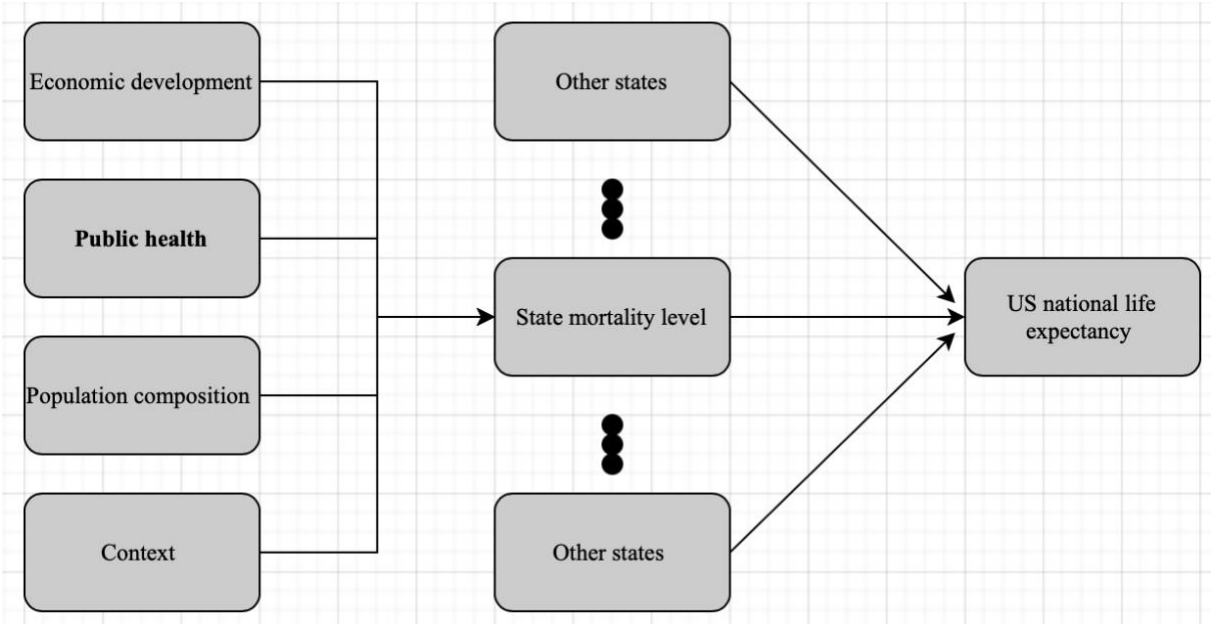
2.3 Conceptual model

From the literature review and the theories, a conceptual model has been constructed, as can be seen in figure 2. The conceptual model shows the main concepts of this thesis and how the concepts are related to each other. State mortality levels are influenced by different factors. One of the factors that influence the state mortality level is economic development, which is related to the wealth situation in a state. In research by Riley (2001), the effect of economic modernization leads to a change in lifestyle, which also influences the most common causes of death, which is also visible in the epidemiological transition theory by Omran (1971). A second factor that influences the state mortality level is population composition, does a state have a young or old population and a third factor is the context, which is related to the living situation in a state. Research by Leyland & Groenewegen (2020) acknowledges these differences and the influence of composition and context on a health level, which influences the mortality levels. The final factor that influences the state mortality level is public health, which is related to the main causes of death and changes in the main causes of death. The changes in the mortality pattern are visible in the epidemiological transition theory by Omran (1971) and the health transition (Riley, 2001). All these factors combined, lead to the state mortality level. 50 states and the district of Columbia are included in this research, which leads to 51 different state mortality levels. All these state mortality levels are different and thus show geographical inequalities. The different state mortality levels combined lead to the national life expectancy in the United States.

This research focuses on public health and how the causes of death in different age groups change over time and how this influences life expectancy. Therefore, the main concept

used in this research is public health. The other factors also influence life expectancy, however, due to data limitations in this thesis, these factors will not be included in the analysis.

Figure 2: Conceptual model



Source: Own figure.

2.4 Hypotheses

Based on the literature the following hypotheses are formulated:

1. Young and middle-aged adult age groups contribute the most to the differences between national life expectancy and state life expectancy, especially due to the causes of death I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality).
2. There is a geographical divergence in US state life expectancy because of differences in midlife mortality levels.
3. The cause of death V01-Y89 (External causes of morbidity and mortality) has had the greatest contribution to the divergence in life expectancy in the US in the period 1999 to 2020.

3. Data and methodology

3.1 Study design

To what extent geographical inequalities in US State life expectancy have developed and which causes of death and age groups contributed most to changes in life expectancy, is investigated by quantitative research. In this research, the numerical data describes and explains how life expectancy and geographical inequality have changed (Babbie, 2021). The research is descriptive, which is related to the purpose. The purpose is to describe what is holding back progress in US life expectancy and which causes of death, age groups and states are contributing to this and which patterns are visible in the states with the lowest and highest life expectancy. According to Babbie (2021), descriptive research is one of the major purposes of social sciences.

3.2 Setting

The study area of this research are the 50 states of the US and the District of Columbia. The research is a diachronic analysis because the data is collected over the period 1999 to 2020 for 50 states and the District of Columbia to see how life expectancy has changed over this period and what caused these changes (Babbie, 2021). The data is collected for the whole period. Due to privacy issues, death counts with 9 or fewer deaths are suppressed (more in section 3.3.1). This research focuses on life expectancy for a longer period because life expectancy always fluctuates because of wars, pandemics or economic crises and because of stochasticity in small populations like in some of the US states (Riley, 2001). By looking at a longer period, a more stable pattern of how life expectancy changed is visible.

Besides the states, the data is collected by sex because the National Center for Health Statistics shows that life expectancy is different for males and females and previous research showed that developments in life expectancy in the US for males and females are also different. The difference between male and female life expectancy depends on biological factors and social factors (Arias et al., 2021; Crimmins et al., 2019; Zarulli et al., 2021).

The data is collected in 10 years age groups (table 1), except for the first two age groups and the last age group. The first two age groups are shorter than 10 years because of the U-shape pattern in age-specific mortality rates (Engelman et al., 2017). The oldest age group is a relatively young open-ended age group, which was imposed by the CDC distribution tool. The open-ended age group is related to the fact that the highest age at which a person could die is unknown.

Table 1: Age groups

Age Groups
<1
1-4
5-14
15-24
25-34
35-44
45-54
55-64
65-74
75-84
85+

Source: Own table.

3.3 Databases

3.3.1 CDC Wonder database

In mortality analysis, high-quality data is important to get reliable results and to be able to get to the right conclusions. Good quality data is important because it shows the real mortality pattern and changes over time. Furthermore, it helps to monitor trends in public health by identifying health risks (Hua et al., 2021: Majewska, 2017).

The Centers for Disease Control and Prevention (CDC) wonder Underlying Cause of Death database is an open-access database on which the data for 20 causes of death is visible. The 20 causes of death are based on the ICD-10 chapters and are visible in table 2. The main advantage of using this dataset is that in the whole study area the data is collected in the same way. In this research, the data is collected via an open-access database, which is reachable via this link (https://wonder.cdc.gov/controller/datarequest/D76). The dataset is produced by the Mortality Statistics Branch, Division of Vital Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention and United States Department of Health and Human Services (CDC, n.d.-b; U.S. Department of Health and Human Services, 2004).

In the CDC wonder database, the total number of deaths for every age group is known, but, due to privacy, for some categories, when the value is below 9, the real value is suppressed. If these values were considered as zeros without further considerations, this could lead to the problem that the results of which age groups and causes of death contributed the most to the difference in life expectancy are under- or over-estimated. To solve the problem, several options are possible. One of the options is to compare the number of deaths known with the total number of deaths to see if certain age groups or causes of death differ from the national numbers, which could have an influence on the results (more in section 3.4). Another solution could be to spread the deaths which are suppressed in the CDC wonder databases over all 20 causes of death, which could give biased results.

The database presents the underlying cause of death in the United States of America during the period 1999 to 2020. The database consists of mortality and population on different geographical scales. The underlying cause of death is "*the disease or injury which initiated the train of morbid events leading directly to death or circumstances of the accident or violence which produced the fatal injury."* (WHO, 1979, p. 6).

In the absence of a population register in the USA, the data in the database is collected via death certificates that are independent from any previous administrative information about the deceased. The cause on the death certificate is given by a physician, medical examiner or coroner. The death certificates are shared with the states, and the states send the information to the National Center for Health Statistics. If the cause of death is not completely sure, the certifiers may include terms like probably on the certificate. The more complete the data is, the better it is for the quality of the data. Around 20-30% of the death certificates contain missing information, although this does not mean that the certificates are inaccurate. An example of information that is sometimes missing is the age of death, which leads to the death going to the category non-stated. This category, which consists of a small percentage of total deaths, will be excluded from the analysis. Another solution was to spread the non-stated deaths uniformly across all the causes of death, however, causes of death with small numbers could be distorted and therefore influence the final result, because not every cause of death is present in the population in the same way (CDC, n.d.-d).

Another influence on the result could be that in the dataset only one underlying cause of death is noted. In the database, the first cause of death will be used as the underlying cause of death. This could lead to that the influence of some causes of death could be underestimated when those causes of death are often present as comorbidities but not as the causes of death itself (U.S. Department of Health and Human Services, 2004).

Table 2: Causes of death according to ICD-10 main chapters

Cause of death	**Code**
Certain infectious and parasitic diseases	A00-B99
Neoplasms	C00-D48
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanisms	D50-D89
Endocrine, nutritional and metabolic disease	E00-E88
Mental and behavioural disorders	F01-F99
Diseases of the nervous system	G00-G98
Diseases of the eye and adnexa	H00-H57
Diseases of the ear and mastoid process	H60-H93
Diseases of the circulatory system	I00-I99
Disease of the respiratory system	J00-J98
Diseases of the digestive system	K00-K92
Diseases of the skin and subcutaneous tissue	L00-L98
Diseases of the musculoskeletal system and connective tissue	M00-M99
Diseases of the genitourinary system	N00-N98
Pregnancy, childbirth and the puerperium	O00-O99
Certain conditions originating in the perinatal period	P00-P96
Congenital malformations, deformations and chromosomal abnormalities	Q00-Q99
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00-R99
Codes for special purposes	U00-U99
External causes of morbidity and mortality	V01-Y89

Source: Own table based on https://wonder.cdc.gov/controller/datarequest/D76

3.3.2 WHO Mortality Database

To be able to gauge the completeness of the CDC wonder data and see if the suppressed values in the data have a big influence on the results, data on national deaths is collected from the WHO Mortality Database. The data of the WHO Mortality Database is collected by the responsible authorities in the countries. There could be a difference in the data of the countries and the WHO because the WHO use standard classifications to present the data. The data in the WHO mortality database is based on the underlying causes of death which are the same as in the CDC database. The data is only included in the database if the data is coded according to the ICD-10 classification. One caution should be that the data is complete for the population covered but that not all deaths in the country are included in the database (WHO, 2022).

According to the WHO mortality database, the data in the US is 100% complete. 12.1% of the data is ill-defined or non-specific causes of the total deaths which means that the code used does not have a proper medical interpretation. The data has a usability of 87.9% which means that the United States has high usability. Comparing the WHO and CDC data does not lead to a problem because in both datasets the code on the death certificates is used in the same way, which is related to the fact that the data is provided by CDC to the WHO. This number is calculated by using the data from the period 2008 to 2019. According to the WHO the usability is defined as "*the percentage of all deaths which are registered with meaningful cause-of-death*

information" (WHO, 2020). It is calculated by multiplying the completeness with the proportion of deaths that have a meaningful cause of death (WHO, 2020).

3.4 Data validation

Appendix 1 tables 4 and 5 show the suppressed values in the data for females and males. It is visible that the lowest number of suppressed cells for all causes of death is lower in the age group <1 and after that, the number of suppressed cells for all causes of death increases till the age group 5-14. After this age group, the number of suppressed cells is decreasing. This pattern is explained by the age-specific mortality rates, which generally have a U-shape (Engelman et al., 2017). Indeed, age groups with low mortality rates tend to generate low death counts, which are more likely to be suppressed for privacy reasons.

Logistic regression was performed to see which age groups, causes of death, sex and year are more likely to exhibit suppressed death counts. The dependent variable is suppressed cells, with a success if the cell is suppressed. By looking at the regression output (Appendix 2 table 6), the results for every age group are not statistically different from the reference category (5-14). The different causes of death (appendix 2 table 6) have a statistically different effect compared to the reference category A00-B99 (Certain infectious and parasitic diseases). By looking at the log odds, of the causes of death C00-D48 (Neoplasms), E00-E88 (Endocrine, nutritional and metabolic disease), G00-G98 (Diseases of the nervous system), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system), K00-K92 (Diseases of the digestive system), V01-Y89 (External causes of morbidity and mortality) have a negative and significant effect on the 0.001 percent level on the number of suppressed values, compared to the reference category A00-B99 (Certain infectious and parasitic diseases), which means that they decrease the number of suppressed values. In the dataset, those diseases have more often a number instead of a suppressed value, probably because the death rates of these causes make them less likely to result in small death counts.

On the contrary, the log odds of D50-D89 (Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanisms), F01-F99 (Mental and behavioural disorders), L00-L98 (Diseases of the skin and subcutaneous tissue), M00-M99 (Diseases of the musculoskeletal system and connective tissue), N00-N98 (Diseases of the genitourinary system), O00-O99 (Pregnancy, childbirth and the puerperium), P00-P96 (Certain conditions originating in the perinatal period), Q00-Q99 (Congenital malformations, deformations and chromosomal abnormalities), R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified), U00-U99 (Codes for special purposes) have a positive and significant effect on the 0.001 percent level on the number of suppressed values compared to the reference category. This means that those diseases in the dataset more often have a suppressed value instead of a valid death count than A00-B99 (Certain infectious and parasitic diseases). The two insignificant causes of death H00-H57 (Diseases of the eye and adnexa) and H60-H93 (Diseases of the ear and mastoid process) are small groups and therefore they could be too small to say if the effect is caused by these diseases.

The coefficient of the independent variable year also has a negative significant effect on the suppressed values (Appendix 2 table 6). The more recent the year, the less suppressed values are present in the data, which could be related to the slow rise in the population size. By looking at the independent variable sex, the effect for males in comparison to females is negative and significant at the 0.001 percent level. The odds for males are 14% lower than for females. This shows that males have a lower number of suppressed cells than females. This could be related to that for males in the younger age groups the risk of dying is slightly higher, which could decrease the suppressed cells in the younger age groups and causes of death with fewer deaths (Zarulli et al., 2021). Another explanation could be a gender bias by male physicians in the

diagnosis of the cause of death. By adding the new independent variables to the model, the previous independent variables stay constant, which means that the model is robust.

Table 3 shows the percentage of deaths known by cause of death. Two have a percentage of zero, which are H00-H57 (Diseases of the eye and adnexa) and H60-H93 (Diseases of the ear and mastoid process). Additionally, in the cause U00-U99 (Codes for special purposes), only 6,89% of the deaths are known. These categories have a low value because only a few people died from one of these diseases, which does not have a significant influence on calculating the life expectancy and thus does not affect the conclusions. Furthermore, for these causes of death, many age groups and years display no death from this disease. Other causes of death like A00-B99 (Certain infectious and parasitic diseases), E00-E88 (Endocrine, nutritional and metabolic disease), F01-F99 (Mental and behavioural disorders), G00-G98 (Diseases of the nervous system), N00-N98 (Diseases of the genitourinary system), P00-P96 (Certain conditions originating in the perinatal period), R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) and V01-Y89 (External causes of morbidity and mortality) have a percentage which is higher than 97% which means that for those cases, almost all deaths are known and there are only a few suppressed values. However, if the missing 3% all come from a handful of small states, it still could affect the results. It is not possible to solve this. In three categories, there is a percentage which is higher than 100%, which are the categories C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and J00-J98 (Disease of the respiratory system), which means that in the CDC database there are more cases known than in the WHO database. These causes of death are only slightly higher than 100%, which could be caused by that the data transferred to the WHO was later corrected (WHO, 2022).

By comparing the data from the CDC and the WHO, on average the different causes of death have a very high percentage of deaths that are known, with a few exceptions that are not likely to affect the results. The death categories with smaller death counts have lower proportions known, which could be related to the fact that more age groups have 9 or fewer deaths for these causes of death, and thus more suppressed values.

Table 3: Percentage of deaths known by causes of death

Death cause	Percentage of deaths known
A00-B99	98.10%
C00-D48	100.18%
D50-D89	83.06%
E00-E88	99.12%
F01-F99	98.54%
G00-G98	98.83%
H00-H57	0.00%
H60-H93	0.00%
I00-I99	100.47%
J00-J98	105.44%
K00-K92	86.45%
L00-L98	67.99%
M00-M99	90.02%
N00-N98	98.28%
O00-O99	57.52%
P00-P96	98.77%
Q00-Q96	73.14%
R00-R99	97.16%
U00-U99	6.89%
V01-Y89	99.75%

Source: Own table based on CDC wonder database and WHO mortality database.

3.6 Ethical considerations

In the process of quantitative research, it is important to think about ethical considerations. The data in this research is secondary data because the data used in the research is collected by the CDC, which means that the CDC is responsible for the collection and the storage of the data, which the CDC have experience with and has extensive regulations to protect the confidentiality of the data (CDC, n.d.-a). The researcher may assume that the confidentiality and privacy of the respondents have priority in the data collection process. Although, even the researcher should take care of ethical considerations by using the data and preventing harm to the persons while still guaranteeing their anonymity. Furthermore, it is important that the researcher does not misuse the data and only use the data for the purposes for which the data is collected.

Besides the ethical considerations in the data collection and storage process. During the whole research cycle, the research must have social value, which means that the results should provide knowledge in the discipline of demography. Besides that, the research needs to have scientific validity, which means that the research should have followed the scientific method to have reliable and applicable results (Acevedo Pérez et al., 2017). In the analysis and reporting part of the research, it is important that the researcher is more familiar with the limits of the study and that the researcher make them known to the readers (Babbie, 2021).

3.7 Methods

3.7.1 Life expectancy

Life expectancy is one of the measures of the health of a population and its development (Roser et al., 2013). It is defined as "*The average number of years that a newborn is expected to live if current mortality rates continue to apply*" (Hazen & Anthamatten, 2020, p. 95). Worldwide, big differences exist in life expectancy, the more developed a country is, the higher the life expectancy is on average. One factor that has a major influence on life expectancy is infant mortality. The reason that infant mortality has a major influence on life expectancy is that the lifespan for infants is short, which decreases the average lifespan the most years (Hazen & Anthamatten, 2020).

In life expectancy, a difference can be made between cohort and period life expectancy. The cohort life expectancy is how long on average a birth cohort lives. It is only possible to calculate the cohort life expectancy if everyone in the cohort has died. This could lead to outdated information about the health situation in a population because the whole population already passed away (Ortiz-Ospina, 2017).

An alternative way to calculate life expectancy is the period life expectancy. The period life expectancy is the average number of years people in a certain period live. To calculate the period life expectancy a hypothetical cohort is assumed to have the same mortality rates during the whole period. This leads to the disadvantage of this method, which is that it does not take into account changing mortality rates over time (Ortiz-Ospina, 2017).

3.7.2 Life table analysis

The life table is one of the main used tools in demography. A life table is a table that illustrates different pieces of information about mortality in an observed or hypothetical birth cohort. A life table is an easy way to describe the various aspects of the variation of mortality with age and cause of death (Hinde, 1998; Preston et al., 2000). The life table assumes that mortality rates are constant in an age group. The broader chosen the age group, the less likely it is that this assumption holds. This research contains 10 years age groups, except for the age groups <1, 1-4 years and 85+. The youngest age groups are smaller than 10 years because more people die at younger ages than older ages in the youngest age groups. The 85+ is an open-ended age

group because it is not clear to which age someone survives. To solve the problem, with the assumptions in the life table, the nqx can be set equal to 1 (Hinde, 1998).

A life table has different columns with information about mortality in the population. The first column is x, which contains the age groups. The second column is lx, which contains the number of people alive in a certain age group. The third column is ndx, which contains the number of people dying in a certain age group. This column could also be split into multiple causes of death. In the fourth column, nqx, is the probability of dying in a certain age group. The fifth column is the nPx, which contains the probability of surviving in a certain age group. The sixth column is nLx, which is the number of person-years lived. The seventh column is the e0 which is the life expectancy at a certain age. The eighth column is nmx which is the death rate in a cohort in an age group. The last column, nax, contains the average person-years lived in the interval by the people who died in the interval (Preston et al., 2000).

It is possible to have a period or a cohort life table. A cohort life table is a life table based on a cohort. It could be difficult to construct a cohort life table because it is necessary to wait till a complete cohort is deceased, which could lead to the fact that the data is outdated. An alternative is the period life table, a period life table is a life table that shows what would happen to a certain cohort if the cohort follows the mortality pattern for a certain period. A period life table shows what would happen to a hypothetical cohort if mortality conditions stay the same through the whole period (Preston et al., 2000). This thesis used the period life table, which is related to the fact that the data is for a recent period, to get an overview of how mortality developed.

Besides the difference between cohort and period life tables, it is also possible to construct single-decrement life tables and multi-decrement life tables. A single-decrement life table is a life table that has one absorbing state. However, this research contains twenty absorbing states, which means that the life table has various causes on which a person could die (Oeppen, 2008). The multiple decrement life table is often used to see the contribution of a certain cause of death on life expectancy and how life expectancy changed. The main assumptions of the multiple-decrement life table are that multiple causes of death are mutually exclusive of each other and the probability that someone will die is in the existence of the other causes (Carey, 1989; Land et al., 1994).

3.7.3 Decomposition

By generating life tables for different populations, it is useful to know what causes the differences between the populations. However, after creating life tables, it is not possible to compare the life tables directly with each other over time directly from the input rates (Canudas Romo, 2003).

To work around this disadvantage, decomposition methods are a useful way to analyze mortality patterns. Decomposition is the principle of separating demographic measures into different components to see how each of these components contributes to the understanding of the phenomenon under study. Different ways of decomposition exist. One of them is age decomposition, which helps to estimate the contribution of each age group to the total change over time. Another decomposition technique that could be used is the cause of death decomposition. Using this technique makes it possible to see the contribution of every cause of death to the total gap in life expectancy. A third way to decompose is to use different categories, for example, males and females (Canudas Romo, 2003). So, decomposition models help to estimate what mortality differences a specific age group or cause of death contribute to the gap in life expectancy (Preston et al., 2000). Decomposition models are a way to see how each age group and cause of death contribute to differences. Another advantage of decomposition models is that decomposition models show which cause of death and age group contributes the most to

the change and what kind of action policymakers could implement to make further progress in life expectancy (Auger et al., 2014).

Different decomposition models exist. The two broad methods to decompose life expectancy are the analytical approach, i.a. by Pollard and Arriaga, and the other method is the numerical approach by Andreev and Horiuchi (Andreev et al., 2002; Preston et al., 2000).

The Pollard decomposition method is based on the force of mortality concept. In the method, the interaction effect and the main effect are combined. The interaction terms are the effects that are rising because of changes in mortality in two or more age groups. However, a disadvantage of this method, is that it needs detailed information in life tables about the age groups 85+ (Murthy, 2005).

Arriaga's decomposition model is based on temporary life expectancy. A change in mortality in a specific age group affects the life expectancy of other age groups. This could have a direct or an indirect effect on life expectancy. A direct effect on life expectancy is that there are changes in life years within an age group because of changes in mortality in that age group. The indirect effect is that the number of years lived increases due to mortality changes within a specific age group. Therefore, this change will lead to a change in the number of survivors till the end of the age group. It is assumed that mortality does not change in other age groups. The interaction effect is that there will be changes in other age groups, because of the changes in mortality in a certain age group. Furthermore, an interaction effect is always present because of mortality changes in an age group, which influences the number of survivors in a certain age group. The direct and indirect effects came because of the change only within a certain age group (Arriaga, 1984; Murthy, 2005).

In this research the package DemoDecomp is used in R. The package DemoDecomp offers two different decomposition methods (Riffe, 2018). The first method uses stepwise replacement (Andreev et al., 2002). The stepwise replacement method estimates the effect of replacing a cell in one matrix with a cell in the other matrix. This method could be used to decompose the difference in life expectancy. The package performs all the possible replacements to calculate the results. This method is quite similar to the Arriaga decomposition method (Andreev et al., 2002; Andreev & Shkolnikov, 2012). The second method is Pseudo-continuous decomposition (Horiuchi et al., 2008). In this thesis, the stepwise replacement method is used for the analyses because the results were more stable than with an analytical approach.

4. Results and discussion

4.1 Geographical inequalities

Geographical inequalities in life expectancy are analysed by looking at the life expectancy developments in the period 1999 to 2020. To analyse the developments in life expectancy different maps and graphs are created based on life tables. The data to make the maps and graphs comes from the CDC Wonder database and WHO Mortality database. Firstly, maps of the US states are created, in which the states are sized according to the life expectancy to investigate how geographical inequalities in life expectancy are visible in the US. Thus, the bigger the size of the state, the higher the life expectancy will be. Alaska which has on average a higher life expectancy, look small, however, compared to the other states it still contains a big surface, which is related to the map projection. In appendix 3 figure 10 a map of the US is visible which includes all the state names.

Figures 3.1 and 3.2 show that the states in the north, central and southwest have a high life expectancy for females and males. Moreover, Hawaii also has a high life expectancy compared to the other states in the US. Additionally figures 3.1 and 3.2 show that the southeast of the US, the states of Louisiana, Mississippi and Alabama have low values of life expectancy compared to the other states in the US.

Figure 3: US state life expectancy 2020

Figure 3.1: US state life expectancy for females 2020

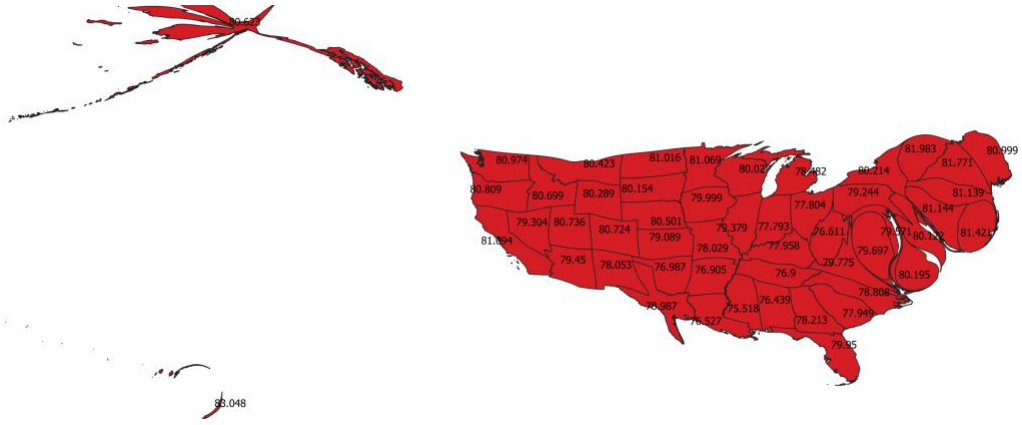
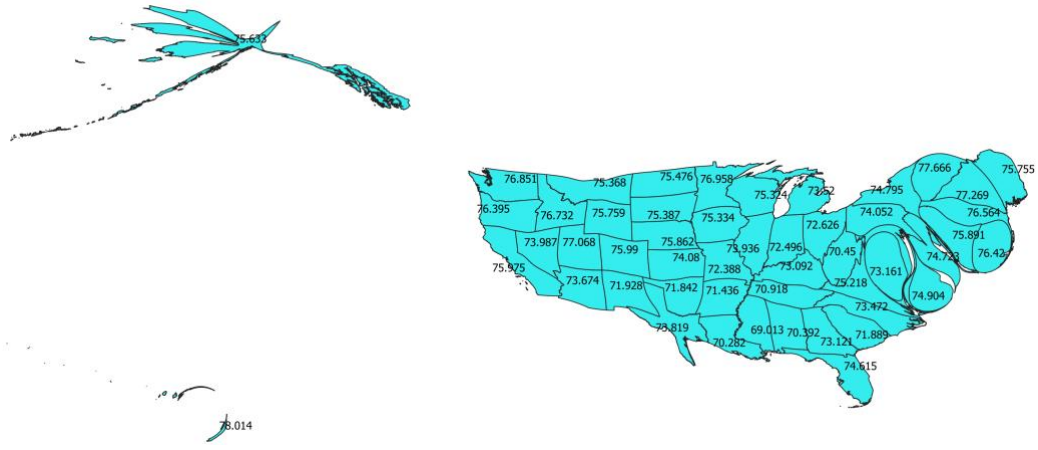


Figure 3.2: US state life expectancy for males 2020



Source: Own figure based on CDC wonder database.

4.1.1 Convergence and divergence

In figure 4.1 the difference between the lowest and highest life expectancy for females in the period 1999 to 2020 is visible. Figure 4.1 shows that in 1999 the gap in life expectancy was 5.06 years, till 2019 the gap increases to 5.88 years. The pattern in the period 1999 to 2014 is quite constant. After 2014 the gap increases from 5.36 years to 5.88 years, which shows a divergence in life expectancy. In 2020 the gap in life expectancy is 7.53 years, which is most likely related to the Covid-19 pandemic and could therefore be an outlier. More data for years after 2020 is needed to see how this gap develops. During the period 1999 to 2020 for females, the states with the lowest life expectancy are Mississippi, Louisiana, Alabama and West Virginia. All those states are located in the southeast of the US. The states with the highest life expectancy are Hawaii, Vermont and North Dakota.

Figure 4.2 shows the gap in life expectancy between the states with the lowest and highest life expectancy for males in the period 1999 to 2020. The pattern shows on average a convergence in life expectancy in the period 1999 to 2013. In this period, the gap in life expectancy decreased from 8.3 years to 6.28 years, which is a decrease of 2.02 years. In the period 2014 to 2019, the gap in life expectancy stabilized. The gap in life expectancy in 2014 was 6.72 years and in 2019 6.77 years. In 2020 the gap in life expectancy was 9 years, which most likely is caused by the Covid-19 pandemic, therefore data from more recent years is needed to see if 2020 is an outlier or if the gap in life expectancy is increasing. The states with the lowest life expectancy are the District of Columbia and Mississippi and the states with the highest life expectancy are Hawaii, Vermont, North Dakota and Utah.

Figure 4: Gap in life expectancy in years in the period 1999 to 2020

Figure 4.1: Gap in life expectancy in years for females

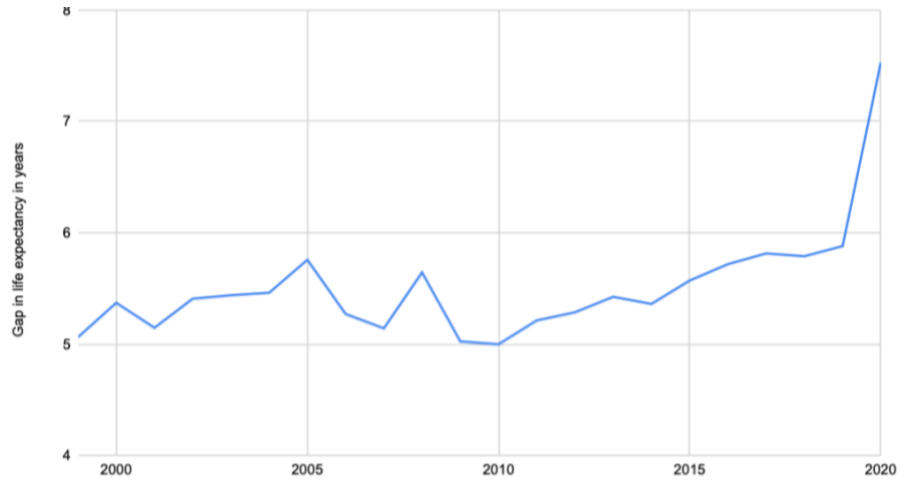
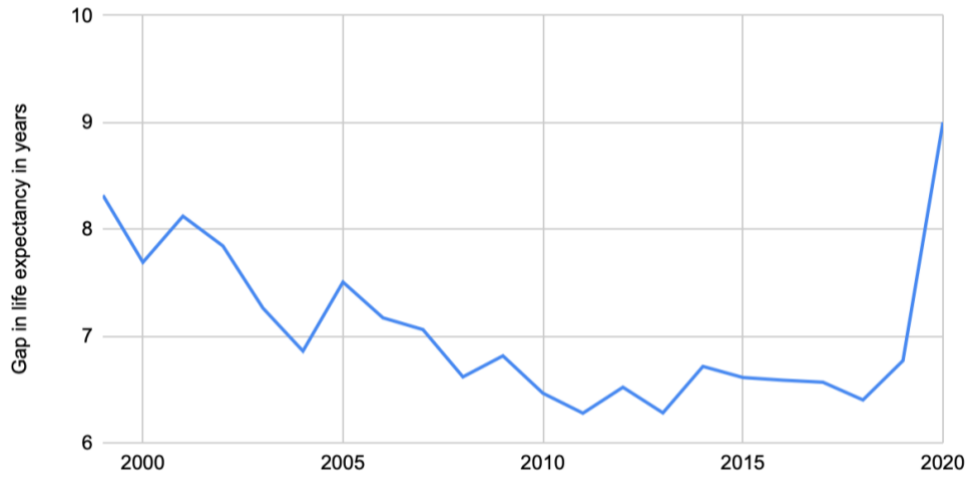


Figure 4.2: Gap in life expectancy in years for males



Source: Own figure based on CDC wonder database.

4.1.2 Lowest and highest life expectancy

Figure 5.1 shows the states with the lowest and highest life expectancy for females in the period 1999 to 2020. In the period 1999 to 2020, the highest life expectancy has increased from 81.95 years to 83.08 years. In the period 1999 to 2020 the states with the highest life expectancy have an increase in additional years of life of 0.05 years per year. The increase in life expectancy for the states with the highest life expectancy seems linear with a small drop around 2011 and 2020. After 2011 the gap in life expectancy between Hawaii and North Dakota increased as can be seen in figure 5.1. In the same period, the lowest life expectancy for females has decreased from 76.89 years to 75.52 years, however, 2020 is most likely influenced by the Covid-19 pandemic, until 2019 the lowest life expectancy increased to 77.37 years. In the period 1999 to 2019, the increase in life expectancy for the states with the lowest life expectancy seems also to show a linear pattern with 0.02 years of additional years of life per year. Thus, in the period 1999 to 2020, the states with the highest life expectancy made a progress and have an increased life expectancy. The states with the lowest life expectancy have an increase in life expectancy in the period 1999 to 2019, however, in the period 1999 to 2020 a decrease, which is caused by the year 2020. During the whole period, the same states with the highest life expectancy are keeping the lead, which is also the case for the states with the lowest life expectancy. The general pattern shows that the states with the highest life expectancy have a higher increase in life expectancy than the states with the lowest life expectancy, although the difference in increase is really small.

Figure 5.2 shows the states with the lowest and highest life expectancy for males in the period 1999 to 2020. The highest life expectancy in 1999 was 77.05 years and increased to 78.01 years in 2020. In this period, the states with the highest life expectancy have an increase in additional years of life of 0.04 years per year. The increase in life expectancy for the states with the highest life expectancy seems linear, with a small drop in 2020. In the same period for the states with the lowest life expectancy, the life expectancy increased from 68.73 years in 1999 to 69.01 years in 2020. In this period the states with the lowest life expectancy have an increase in additional years of life of 0.01 years per year. The increase in life expectancy for the states with the lowest life expectancy seems linear, with a drop in 2020. Thus, in the period 1999 to 2020, the states with the highest life expectancy made a progress and have an increase in life expectancy. The states with the lowest life expectancy have an increase in life expectancy, however, the increase in life expectancy is lower than for the states with the highest life expectancy. Both, the states with the highest and lowest life expectancy only made a small

increase in life expectancy, which also influence the stalling national life expectancy, both the highest and lowest states contributed to the stagnation.

In the period 1985 to 2010, life expectancy developments in the US show that not every state has the same developments or the same pace of development in life expectancy for males and females, which could be related to five different factors: migration, socioeconomic factors, lack of access to healthcare, poor-quality health care and behaviour, metabolic and environmental risks (Wang et al., 2013). These differences in life expectancy developments could also explain why the states with the highest and lowest life expectancy are not the same but differ across gender.

Figure 5: States with the lowest and highest life expectancy in the period 1999 to 2020

Figure 5.1: States with the lowest and highest life expectancy females

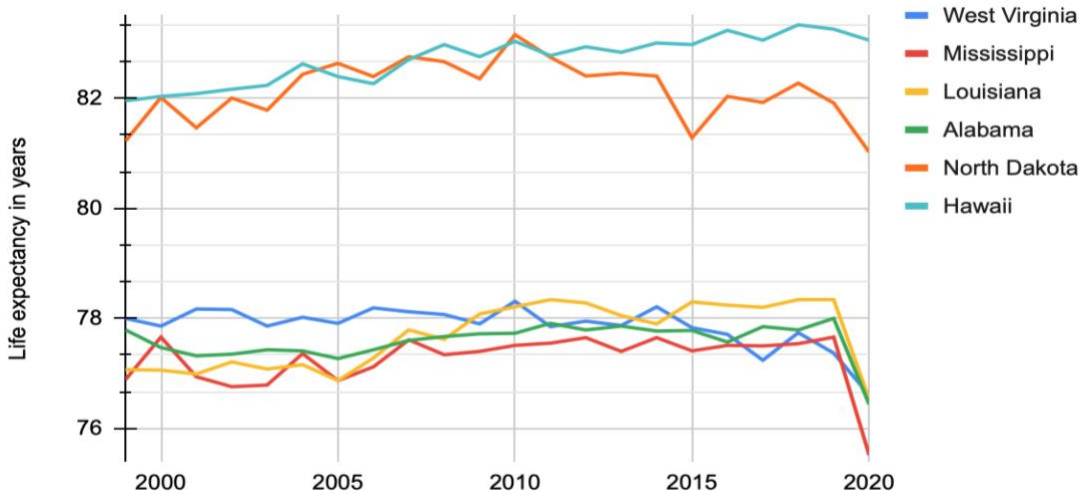
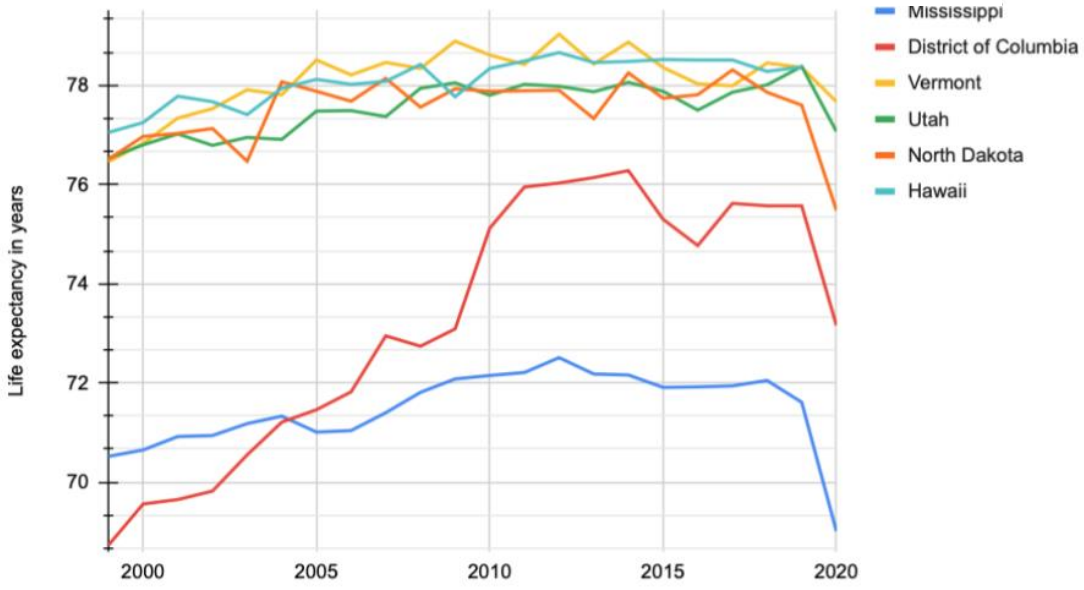


Figure 5.2: States with the lowest and highest life expectancy males



Source: Own figure based on CDC wonder database.

4.2 Females laggard

Figure 6.1 shows that in Alabama in the period 2003 to 2019 the age groups 1-4, 5-14 and 15-24 have on average a decreasing effect on the gap in life expectancy in comparison with the national life expectancy. The age groups 45-54, 55-64 and 75-84 have the biggest contribution to the gap in life expectancy.

Besides that, figure 6.2 shows that in Louisiana in the period 2003 to 2019 the age groups 1-4, 5-14 and 15-24 on average are decreasing the difference in life expectancy or have a small contribution to the gap in life expectancy in comparison with the national life expectancy. The age groups 35-44, 45-54, 55-64, 65-74 and 75-84 have the biggest contribution to the gap in life expectancy.

Additionally, figure 6.3 shows that in Mississippi in the period 2003 to 2019 the age groups 1-4, 5-14 and 15-24 have in some years a small contribution to the difference in life expectancy, however, in most years these age groups have a decreasing effect on the difference in life expectancy in comparison with the national life expectancy. The age groups 45-54, 55-64, 65-74 and 75-84 are the biggest contributors to the difference in life expectancy.

In West Virginia, in the period 2003 to 2019 as can be seen in figure 6.4 is that the age groups <1, 1-4, 5-14 and 15-24 have on average over time a decreasing effect on the gap in life expectancy. The age groups 55-64, 65-74 and 75-84 have the biggest and relatively stable contribution to the difference in life expectancy. The age groups 35-44 and 45-54 have an increasing contribution to the gap in life expectancy.

Figure 7.1 shows that in Alabama in the period 2003 to 2019 the causes of death I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) are the biggest contributors to the difference in life expectancy. The cause of death C00-D48 (Neoplasms) has also a constant contribution to the gap in life expectancy, although, a lower contribution than the ones mentioned previously. The other causes have on average a small or no contribution to the gap in life expectancy.

In Louisiana, as figure 7.2 shows, in the period 2003 to 2019, the causes of death with the biggest contribution to the gap in life expectancy are C00-D48 (Neoplasms), E00-E88 (Endocrine, nutritional and metabolic disease), I00-I99 (Diseases of the circulatory system) and V01-V89 (External causes of morbidity and mortality). The other causes have on average a small or no contribution to the gap in life expectancy.

Figure 7.3 shows that in Mississippi in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) are the biggest contributors to the gap in life expectancy. The causes of death I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) have a quite constant contribution to the gap in life expectancy. The causes of death C00-D48 (Neoplasms) and J00-J98 (Disease of the respiratory system) have a slightly increasing contribution to the gap in life expectancy. The cause of death C00-D48 (Neoplasms) is especially after 2013 a big contributor to the gap in life expectancy. The other causes have on average a small or no contribution to the gap in life expectancy.

In West Virginia, in figure 7.4, in the period 2003 to 2019, the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) are during the whole period big contributors to the difference in life expectancy. The cause of death R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) is in the years 2004 to 2011 and 2016 to 2017 a big contributor, and in the other years a small contributor. The cause of death E00-E88 (Endocrine, nutritional and metabolic disease) shows an increase in the contribution to the gap in life expectancy. The other causes have on average a small or no contribution to the gap in life expectancy.

The states with the lowest life expectancy for females all show the same pattern in the mortality disadvantage. In all states, the younger age groups have a small but non-negligible contribution to the difference in life expectancy, and the contribution peaks around the older midlife age groups (50s and 60s), with a clear increase in some states, and the old age groups (60+), with a more stable pattern. Besides that, the four states with the lowest life expectancy show similarities in the pattern of the contribution of the different causes of death to the gap in life expectancy. In Alabama, Louisiana, Mississippi and West Virginia the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) are big contributors to the gap in life expectancy.

The mortality disadvantage in those four states could be related to socioeconomic status. Research by Fenelon (2013) shows that Mississippi, Louisiana and Alabama have high levels of poverty and rural isolation, which affects social and economic opportunities. The lower socioeconomic status also influences the obesity rate which especially in older age groups has an influence on the mortality rates (Fenelon, 2013). The high obesity rates are related to unhealthy lifestyles, like high rates of tobacco use, poor diet, physical inequality and low fruit and vegetable consumption, this could be an explanation for the high contribution of the cause of death I00-I99 (Diseases of the circulatory system). Research by Akil & Ahmad (2011) shows that the low prices of unhealthy food make it more affordable for people with a low socioeconomic status, which could decrease health. Other reasons for the high contribution could be low access to food stores and low access to physical care facilities. The high contribution of the cause of death C00-D48 (Neoplasms) could be caused by pollution and smoking, which increases the chances of developing neoplasms. The states Mississippi, Louisiana and Alabama show higher smoking rates, compared to other states in the US, which leads to a higher prevalence of lung cancer in older age groups. The effect of neoplasms is strengthened by poor screening behaviour and low socioeconomic status (Aly et al., 2021; Mississippi State Department of Health, n.d.). The negative health effects caused by the previous factors are especially visible in the age groups 50 years and older. In West Virginia, Mississippi and Alabama the cause of death J00-J98 (Disease of the respiratory system) is also a big contributor. In West Virginia as research by Bennet et al., (2019) show the big contribution of the cause of death J00-J98 (Disease of the respiratory system) could be explained by the high smoking rates and air pollution. The high contribution of causes of death C00-D48 (Neoplasms) and/or J00-J98 (Disease of the respiratory system) could also be related to local habits by physicians in how they note down the causes of death on the death certificate because in both causes of death, diseases related to smoking and air pollution are included. The big contribution of the cause of death V01-Y89 (External causes of morbidity and mortality) to the gap in life expectancy could be explained that in rural areas health care facilities are far away. Another explanation could be policies, which are less focused on safety or economic stability. Moreover, alcohol-impaired drivers could also be an explanation for the high number of motor vehicle accidents (Dollar et al., 2020; Mokdad et al., 2018). Besides that, in rural regions, people drive more, which increases the risk of accidents. Research by Woolf et al., (2019) shows that the deaths of despair in midlife age groups are a big contributor to the stalling life expectancy. In West Virginia especially the midlife age groups are contributing to the gap, which could be related to deaths of despair V01-Y89 (External causes of morbidity and mortality). Another factor that could explain the contribution of the older midlife age groups and the old age groups to the gap in life expectancy is the knowledge about the health risks and the belief in science as mentioned by Fenelon (2013). Additionally, inequalities in educational attainment could explain the big contribution of the cause of death E00-E88 (Endocrine, nutritional and metabolic disease) in Louisiana and West Virginia (Amarasinghe et al., 2009; Srikanthan et al., 2016; Tomblin & Lewis, 2011). Montez et al., (2020) analyzed the influence of policies on public health, which could also be an explanation for the states with the lowest life expectancy.

In these states, social security, like labour protection and income security is less profound, which could have negative health effects.

These findings support partially hypotheses 1 and 2 for the states with the lowest life expectancy for females. The middle-age adult age groups have a big contribution to the gap in life expectancy, however, the young adult age groups have a decreasing effect on the gap in life expectancy, which is related to the risks of the factors that cause the mortality disadvantage is only visible later in life. Besides that, the old age groups also are a part of the explanation of the gap in life expectancy. Additionally, in middle and young adult age groups, I00-I99 (Diseases of the circulatory system) is a big contributor to the gap in life expectancy and also deaths of despair (V01-Y89) is a big contributor in these age groups, however, in this cause of death are all external causes of death included, which makes it also possible that motor vehicle accidents, for example, could be the biggest contributors to the gap in life expectancy instead of deaths of despair. The results show that the cause of death V01-Y89 (External causes of morbidity and mortality) is a big contributor to the gap in life expectancy, but the cause of death I00-I99 (Diseases of the circulatory system) has on average a bigger contribution to the gap in life expectancy, which shows that the results are not in line with hypothesis 3. This is caused by lifestyle factors, which could cause cardiovascular diseases later in life.

4.3 Females pioneers

Figure 6.5 shows that in North Dakota in the period 2003 to 2019 almost all age groups in the whole period are contributing to the gap in life expectancy. Especially the age groups 35-44, 45-54 and 55-64 are contributing to the gap in life expectancy. However, the contribution of these age groups to the gap in life expectancy is decreasing in more recent years.

Meanwhile, figure 6.6 shows that in Hawaii in the period 2003 to 2019 all age groups are contributing to the gap in life expectancy. During the whole period, all age groups have lower mortality rates than the national mortality rates. Especially the age groups 65-74 and 75-84 are contributing on average around 50% to the difference in life expectancy.

Figure 7.5 shows that in North Dakota in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) have a big and constant contribution to the gap in life expectancy. The other causes have on average a small or no contribution to the gap in life expectancy.

Figure 7.6 shows that in Hawaii in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) are the biggest contributors the gap in life expectancy. The causes of death I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) show a quite constant contribution to the gap in life expectancy during the whole period. The cause of death C00-D48 (Neoplasms) shows a decreasing effect in the contribution to the gap in life expectancy towards the end of the period. The causes of death A00-B99 (Certain infectious and parasitic diseases), E00-E88 (Endocrine, nutritional and metabolic disease), G00-G98 (Diseases of the nervous system), K00-K92 (Diseases of the digestive system) and R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) also have a notable, although a bit lower, contribution to the gap in life expectancy. During the whole period, the contribution of the cause of death Q00-Q99 (Congenital malformations, deformations and chromosomal abnormalities) is fluctuating, in some years with a low contribution, close to zero and in other years a contribution higher than 0.1 years.

Comparing the life expectancy of the states with the highest life expectancy for females, two different patterns are visible in Hawaii and North Dakota. In Hawaii, the age groups 65-74

and 75-84 are the biggest contributors to the gap in life expectancy. In North Dakota, the age groups 35-44, 45-54 and 55-64 are the biggest contributors to the mortality advantage. Besides that, in the pattern of the contribution of the causes of death in the gap in life expectancy Hawaii and North Dakota show similarities and differences. In both states, the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) have a big contribution to the gap in life expectancy. The differences are visible in that the causes of death A00-B99 (Certain infectious and parasitic diseases), E00-E88 (Endocrine, nutritional and metabolic disease) and G00-G98 (Diseases of the nervous system) contribute to the gap in life expectancy in Hawaii, however, in North Dakota, these causes of death have a much smaller contribution to the gap in life expectancy.

The two clusters of pioneer states could be explained by different factors. The reason that in Hawaii the older age groups are contributing the most to the gap in life expectancy could be related to migration, wealthier and healthier retirees migrate to spend their old days in Hawaii. The contribution of many causes of death such as A00-B99 (Certain infectious and parasitic diseases), C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), E00-E88 (Endocrine, nutritional and metabolic disease), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) to the gap in life expectancy also show the migration pattern of healthy and wealthy migrants. Besides that, research by Wu et al., (2017) also shows the 'salmon effect', which shows outmigration of old sick immigrant groups, which go back to their home country, to be close to family and have access to cheaper health care. The pattern of outmigration could, therefore, have an effect on the mortality rates in the older age groups. North Dakota shows a different pattern compared to Hawaii, which could be explained by a more active lifestyle due to the rural nature of the state or a positive selection of migrants. In more recent years a decreasing trend in the midlife age groups is visible, which could be related to the cause of death V01-Y89 (External causes of morbidity and mortality), which is also mentioned in research by Harris et al., (2021). Another explanation for the big contribution of the causes of death I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) to the gap in life expectancy could be related to the Healthy Hawaii Initiative. This programme strives for making the healthiest choice the easiest one, to promote a healthy lifestyle, it includes taxes on smoking, promoting physical activity and eating healthy, which decreases the risks of cardiovascular diseases and cancers, which are mainly present in the ages 50 years and older. Besides that, the programme includes policies about avoiding car traffic around schools, which decreases car use and increases physical activity. This could also be an explanation for the big contribution of the cause of death V01-Y89 (External causes of morbidity and mortality) (Agner et al., 2020). In North Dakota, the big contribution of causes of death C00-D48 (Neoplasms) and I00-I99 (Diseases of the circulatory system) could be related to composition and context. For North Dakota it is not sure why North Dakota has low cancer rates, it could be related to the environment or related to less income inequality and social norms (Williamson & Ahmed, 2021; Xu et al., 2021).

The results for the pioneer states for females are partially in line with hypothesis 1, in some clusters of pioneer states especially the midlife age groups are contributing to the gap in life expectancy. However, the young adult age groups have a small contribution to the gap in life expectancy. Additionally, the results show that causes of death V01-Y89 (External causes of morbidity and mortality) and I00-I99 (Diseases of the circulatory system) have a big contribution to the gap in life expectancy. The results are also partially in line with hypothesis 2, in one cluster the midlife age groups are the biggest contributors to the gap in life expectancy, however, in the other cluster especially the oldest age groups are the biggest contributors to the gap in life expectancy. The results for the pioneer states for females are not in line with

hypothesis 3 because the cause of death V01-Y89 (External causes of morbidity and mortality) is a big contributor to the gap in life expectancy but the cause of death I00-I99 (Diseases of the circulatory system) has on average a bigger contribution to the gap in life expectancy. The big contribution of I00-I99 (Diseases of the circulatory system) is caused by lifestyle factors.

4.4 Males laggard

Figure 8.1 shows that in the District of Columbia in the period 2003 to 2019 the age groups 1-4 and 5-14 have a decreasing effect on the difference in life expectancy compared to the national life expectancy. The age groups 55-64 and 65-74 have the biggest contribution to the difference in life expectancy. The age group 35-44 has in the period 2003 to 2009 also a big contribution to the difference in life expectancy, however, after 2009, this age group has a decreasing effect on the gap in life expectancy.

Besides that, figure 8.2 shows that in Mississippi in the period 2003 to 2019 the age groups 1-4 and 5-14 have a small effect on the difference in life expectancy in comparison with the national life expectancy. The age groups 45-54, 55-64 and 65-74 are the biggest contributors to the difference in life expectancy. In Mississippi, the contribution of all age groups stays relatively constant over time.

Figure 9.1 shows that in the District of Columbia in the period 2003 to 2019 the cause of death I00-I99 (Diseases of the circulatory system) has a high contribution to the gap in life expectancy. The cause of death V01-Y89 (External causes of morbidity and mortality) has a U-shape contribution to the gap in life expectancy, the contribution is high in the period 2003 to 2010, in the period 2011 to 2014 the cause of death has a decreasing contribution to the gap in life expectancy and in the period 2015 to 2019 an increasing contribution to the gap in the life expectancy. The causes of death A00-B99 (Certain infectious and parasitic diseases) and P00-P96 (Certain conditions originating in the perinatal period) have a strongly decreasing contribution to the gap in life expectancy. The cause of death R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) has on average in the years 2003 to 2009 a relatively high contribution to the difference in life expectancy, however, after 2009 the cause of death R00-R99 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) has a decreasing contribution to the gap in life expectancy.

Figure 9.2 shows that in Mississippi in the period 2003 to 2019 the cause of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) have a big and constant contribution to the gap in life expectancy. The cause of death V01-Y89 (External causes of morbidity and mortality) also has a big and constant contribution, although much smaller than the previous causes of death.

The states with the lowest life expectancy show one pattern in the mortality disadvantage in the age groups. Both states show the pattern that the age groups 55-64 and 65-74 are the biggest contributors to the mortality disadvantage. After 2009 the District of Columbia starts to deviate from this pattern and shows an increase in life expectancy, which could be related to the one city action plan, in which different health improvement actions are stated. Examples of these plans are reducing infant mortality, lowering the obesity rate and expanding access to quality health care. Another explanation could be that the District of Columbia is becoming more and more a transient city, thus an increase in immigration of white affluent migrants, which could have a positive effect on the mortality levels. Gentrification could also be an explanation, with more affluent white people entering the city and a decrease in the lower socioeconomic population, which affects mortality levels (Government of the District of Columbia, 2013; Jackson, 2014).

The mortality disadvantage in Mississippi could be explained by socioeconomic status. Mississippi has high poverty rates and rural isolation, which according to research by Fenelon (2013) affects the social and economic opportunities which could affect the health situation and

the mortality levels. The cause of death I00-I99 (Diseases of the circulatory system) has a big contribution to the gap in life expectancy which could be explained by socioeconomic status, low access to food stores, access to physical care facilities, and low prices of unhealthy food, which makes it more affordable, especially for people with low socioeconomic status (Akil & Ahmad, 2011). An unhealthy lifestyle could lead to higher obesity rates, which especially leads to an increase in old-age mortality levels, which is also shown in the research by Fenelon (2013). In the District of Columbia, it could be explained by physical inactivity, unhealthy diet and tobacco use as research by Kret et al., (2014) shows, which especially lead to an increase in old-age mortality levels. To decrease the effect of the influence of cardiovascular diseases on the contribution to the gap in life expectancy, a policy programme called Cardiovascular Disease and Diabetes Prevention Program is implemented (Kret et al., 2014).

The cause of death V01-Y89 shows a different pattern in the contribution to the gap in life expectancy. In the District of Columbia, the pattern is fluctuating and in Mississippi the pattern is constant. The contribution to the gap in life expectancy in Mississippi as research by Dollar et al., (2020) and Mokdad et al., (2018) show could be related to the accessibility of health care facilities in rural areas and alcohol-impaired drivers. Besides that, Mississippi is a rural state, which leads to higher car use, which could lead to a higher risk for traffic accidents. In the District of Columbia, the contribution to the gap in life expectancy is fluctuating, which could be related to the small population size in the District of Columbia. Another difference is that in the District of Columbia the causes of death A00-B99 (Certain infectious and parasitic diseases) and P00-P96 (Certain conditions originating in the perinatal period) have a decreasing contribution to the gap in life expectancy, and in Mississippi the contribution of these causes of death is constant. The decrease in the cause of death A00-B99 (Certain infectious and parasitic diseases) could be explained by more attention to the prevention, and treatment of HIV, and due to a needle exchange programme (Government of the District of Columbia department of health, n.d.). Besides that, an explanation for the differences in the contribution of causes of death between states could be the composition of the population and the implementation of health policies (Riddell et al., 2018).

Analyzing the results shows that the states with the lowest life expectancy for males are only partly in line with hypotheses 1 and 2. The results show that especially the older midlife age groups and old age groups are contributing the most to the gap in life expectancy and the young age groups are decreasing or have a small effect on the gap in life expectancy. This could be related to implemented policies in the laggard states and in lifestyle, which leads to diseases which are only visible later in life. In the midlife age groups, the causes of death I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) are big contributors to the gap in life expectancy and have a big influence on the gap in life expectancy.

Additionally, the results are not in line with hypothesis 3 which states that the cause of death V01-Y89 (External causes of morbidity and mortality) is a big contributor to the gap in life expectancy, however, the cause of death I00-I99 (Diseases of the circulatory system) has on average a bigger contribution to the gap in life expectancy. This is caused by lifestyle factors, which cause cardiovascular diseases later in life.

4.5 Males Pioneers

Figure 8.3 shows that in Vermont in the period 2003 to 2019 all age groups have on average lower mortality rates compared to the national mortality rates, which could be related to the small size of the state's population. The age groups 45-54 and 55-64 have the biggest contribution to the difference in life expectancy. On average the contribution of the higher middle age groups declined during the whole period.

In Utah, as can be seen in figure 8.4 in the period 2003 to 2019 all age groups have lower mortality rates compared to the national mortality rates. The age groups 55-64, 65-74 and 75-84 are the biggest contributors to the gap in life expectancy. On average all age groups have a quite constant contribution to the gap in life expectancy.

Besides that, figure 8.5 shows that in North Dakota in the period 2003 to 2019 all age groups have on average lower mortality rates compared to the national mortality rates. The age groups 45-54 and 55-64 have the biggest contribution to the gap in life expectancy. In 2018 and 2019 the effect of the contribution of these age groups decreased, and the contribution of the age groups 25-34 and 35-44 increased.

Finally, figure 8.6 shows that in Hawaii in the period 2003 to 2019 all age groups have lower mortality rates compared with the national mortality rates. The age groups 65-74 and 75-84 are the biggest contributors to the gap in life expectancy. In the period 2003 to 2008 the age groups 55-64 is also a big contributor to the difference in life expectancy, however, after 2008 the contribution of this age group decreases. All the other age groups show during the period a quite constant contribution to the difference in life expectancy.

Figure 9.3 shows that in Vermont in the period 2003 to 2019 the cause of death I00-I99 (Diseases of the circulatory system) has the biggest contribution to the gap in life expectancy. The contribution of I00-I99 (Diseases of the circulatory system) decreased during the period. The causes of death K00-K92 (Diseases of the digestive system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) have a fluctuating contribution to the gap in life expectancy, in some years the contribution to the gap is high and in other years the contribution is low.

Figure 9.4 shows that in Utah in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms) and I00-I99 (Diseases of the circulatory system) have the biggest contribution to the gap in life expectancy, both causes of death show during the period a decrease in the contribution to the gap in life expectancy. The cause of death V01-Y89 (External causes of morbidity and mortality) has in the period 2003 to 2015 a negative or a small contribution to the gap in life expectancy. After 2015 the contribution to the gap in life expectancy is increasing. A stunning fact is that figures 9.2 and 9.4 show that Utah and Mississippi show the same pattern, in which both states show a strong contribution of the causes of death C00-D48 (Neoplasms) and I00-I99 (Diseases of the circulatory system), however, in opposite directions.

Moreover, figure 9.5 shows that in North Dakota in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) have the biggest contribution to the gap in life expectancy. However, over time the contribution of these causes of death to the gap in life expectancy is fluctuating. The cause of death A00-B99 (Certain infectious and parasitic diseases) has also a big influence on the gap in life expectancy, this cause of death shows on average a decrease in the contribution to the gap in life expectancy over time.

Figure 9.6 shows in Hawaii in the period 2003 to 2019 the causes of death C00-D48 (Neoplasms) and V01-Y89 (External causes of morbidity and mortality) have the biggest contribution to the gap in life expectancy. The cause of death I00-I99 (Diseases of the circulatory system) has at the beginning of the period a high contribution to the gap in life expectancy, which is decreasing over time. Besides that, it is striking that in Hawaii the mortality advantage is spread over many causes of death.

The four states with the highest life expectancy in the period 2003 to 2019 show two different patterns in the contribution to the gap in life expectancy in the age groups. In North Dakota and Vermont the age groups 45-54 and 55-64 are the main contributors to the mortality advantage. In Hawaii and Utah, the age groups 65-74 and 75-84 are the main contributors to the mortality advantage. In the contribution of the causes of death, similarities and differences are visible in every state. The different patterns could be explained by different factors. The

mortality advantage in the midlife age groups in North Dakota and Vermont could be related to a more active lifestyle, which could be related to the rural nature of the state or a positive selection of migrants. The decreasing effect of the midlife age groups could be related to the increases in midlife mortality levels, related to deaths of despair (Harris et al., 2021). In Vermont, the low midlife mortality rates could also be related to the small population or the composition of the population (Woolf & Schoomaker, 2019). In Hawaii, the low mortality levels in the old age groups could be related to migration because wealthier and healthier retirees are migrating to Hawaii and old sick immigrants causing outmigration according to Wu et al., (2017). In figure 9.6 is also visible that in Hawaii the mortality advantage is spread over many causes of death like C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality), which is in line with a general selected population of wealthy retirees, no specific cause of death is responsible for the big gap in life expectancy. A similarity among Vermont, Utah, North Dakota and Hawaii is that the cause of death I00-I99 (Diseases of the circulatory system) is a big contributor to the gap in life expectancy. In Hawaii the big contribution of the cause of death I00-I99 (Diseases of the circulatory system) to the gap in life expectancy could be explained by the Healthy Hawaii Initiative, which is a programme in which they making the healthiest choice the easiest one, to promote a healthy lifestyle. This includes taxes on smoking, promoting physical activity and eating healthy (Agner et al., 2020). In North Dakota, this could be explained by the environment, less income inequality and social norms (Xu et al., 2021). In Utah, the big contribution could be explained by low smoking rates and physically active people, which reduces the negative health effect of cardiovascular diseases (Yorgason et al., 2018). By reducing the mortality levels of the cause of death I00-I99 (Diseases of the circulatory system) the effects of the mortality advantage are especially visible in the age groups 50 years and older.

A difference is that in Hawaii and North Dakota the cause of death V01-Y89 (External causes of morbidity and mortality) is a big contributor to the gap in life expectancy. In Hawaii, this could be explained by the Healthy Hawaii Initiative, which includes policies about avoiding car traffic around schools, which decreases car use and increase activity. This could also be an explanation of the big contribution of the cause of death V01-Y89 (External causes of morbidity and mortality). However, in Vermont and Utah, this cause of death shows a more fluctuating pattern in the contribution to the gap in life expectancy, which in Vermont could be related to the small population size. Besides that, the rural environment could also have an influence on the fluctuating pattern because in a rural environment people drive more and could have more accidents. Another difference is that in North Dakota, Utah and Hawaii the cause of death C00-D48 (Neoplasms) has a big contribution to the gap in life expectancy. In Utah, this could be related to low smoking prevalence and physically active people, which lead to lower mortality levels in old age groups (Yorgason et al., 2018). In Vermont, the cause of death C00-D48 (Neoplasms) does not have a big contribution to the gap in life expectancy, which could be related that no policies are implemented in the previous period. In the next period, the policy the 2025 Vermont Cancer plan is implemented to reduce the burden of neoplasms (Vermont Department of Health, 2021). For North Dakota it is not sure why low cancer rates are present, it could be related to the environment or related to less income inequality and social norms (Williamson & Ahmed, 2021; Xu et al., 2021).

The results show that hypothesis 1 is partly accepted, which states that young and middle-aged adult age groups cause the gap in life expectancy, especially due to the causes of death V01-Y89 (External causes of morbidity and mortality) and I00-I99 (Diseases of the circulatory system). In one cluster of states, the middle-aged adult age groups are the biggest contributor however in the other cluster the older age groups are the biggest contributor to the gap. In the younger and adult age groups, indeed the causes of death I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality) are the biggest

contributors to the difference in life expectancy, which could be related to lifestyle and the socioeconomic status of the states. Hypothesis 2 is also partly accepted because in one cluster of pioneer states that midlife mortality is the biggest contributor to the gap in life expectancy, however, the other cluster shows that the oldest age groups are the biggest contributor. hypothesis 3 is rejected which is related to the fact that the contribution to the gap in life expectancy is higher for I00-I99 (Diseases of the circulatory system) than for V01-Y89 (External causes of morbidity and mortality).

Figure 6: Age decomposition of the difference in life expectancy with national average females

States with the lowest life expectancy for females in the period 2003 to 2019

Figure 6.1: Alabama

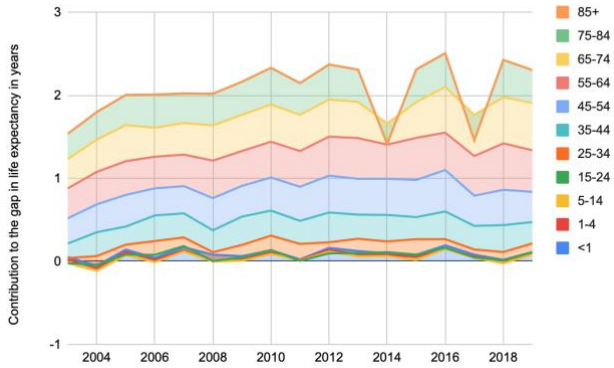


Figure 6.2: Louisiana

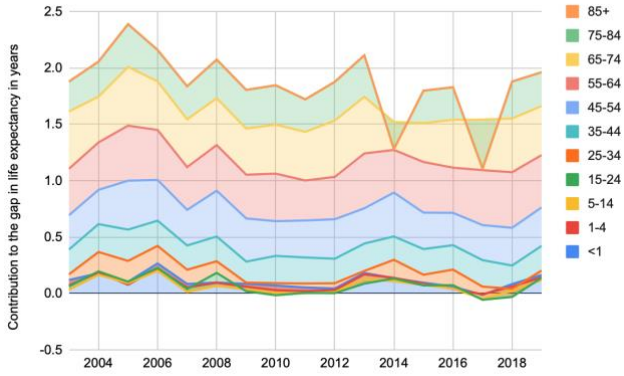


Figure 6.3: Mississippi

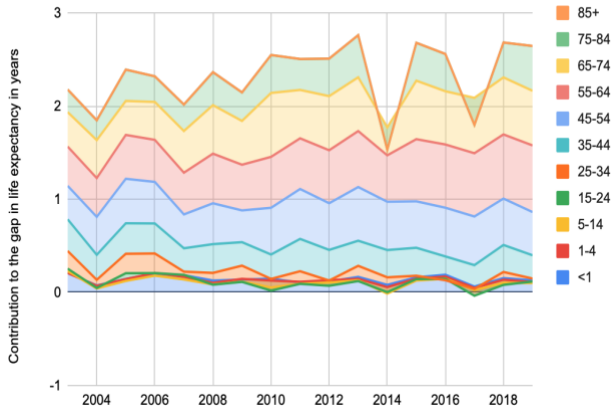
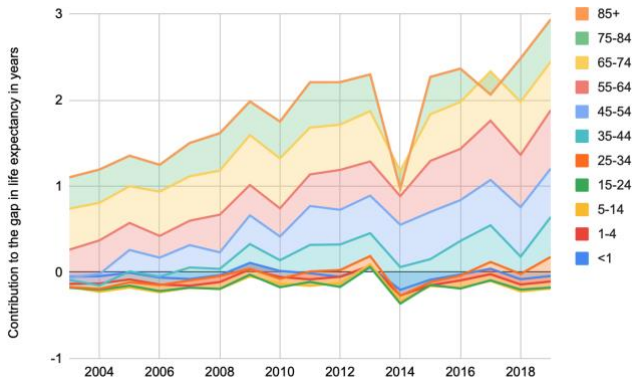


Figure 6.4: West Virginia



States with the highest life expectancy for females in the period 2003 to 2019

Figure 6.5: North Dakota

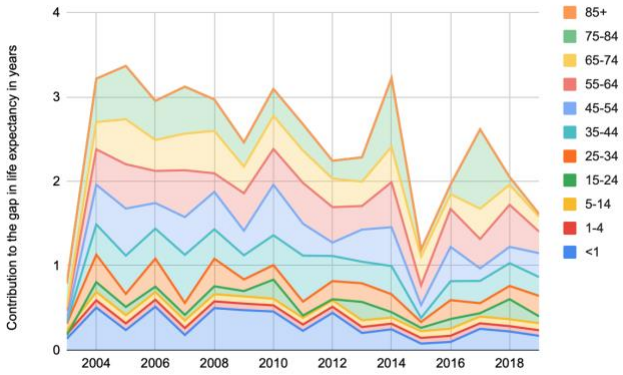
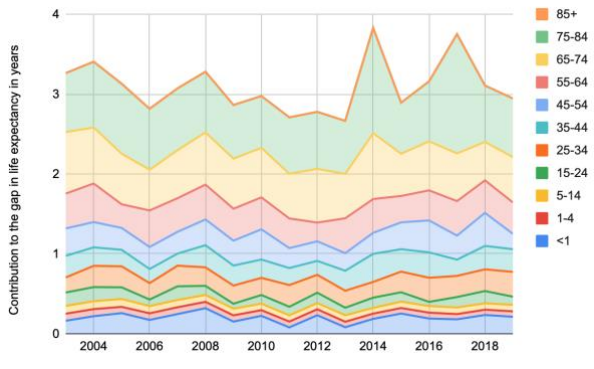


Figure 6.6: Hawaii



Source: Own figure based on CDC wonder and WHO Mortality database.

Figure 7: Cause specific decomposition of the difference in life expectancy with national average females

States with the lowest life expectancy for females in the period 2003 to 2019

Figure 7.1: Alabama

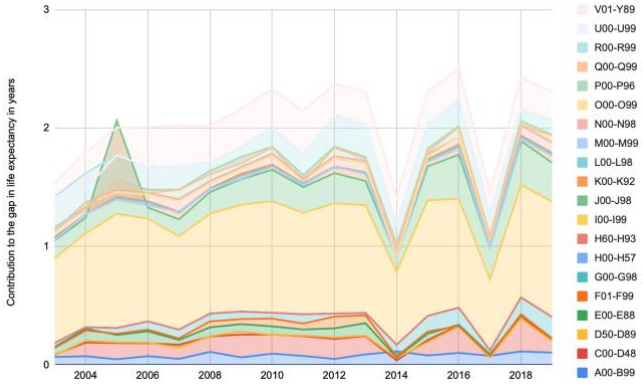


Figure 7.2: Louisiana

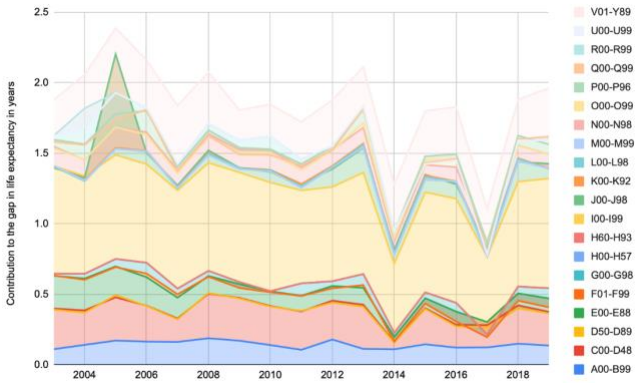


Figure 7.3: Mississippi

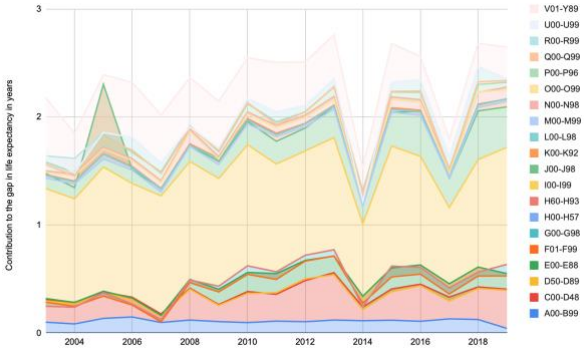
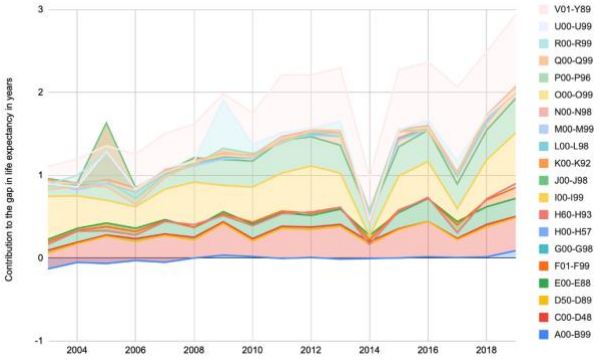


Figure 7.4: West Virginia



States with the highest life expectancy for females in the period 2003 to 2019

Figure 7.5: North Dakota

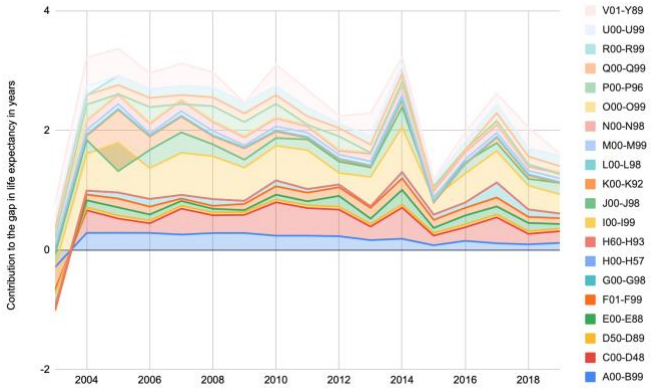
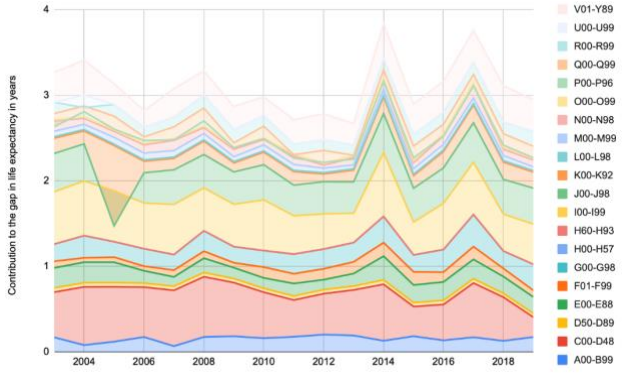


Figure 7.6: Hawaii



Source: Own figure based on CDC wonder and WHO Mortality database.

Figure 8: Age decomposition of the difference in life expectancy with national average males

States with the lowest life expectancy for males in the period 2003 to 2019

Figure 8.1: District of Columbia

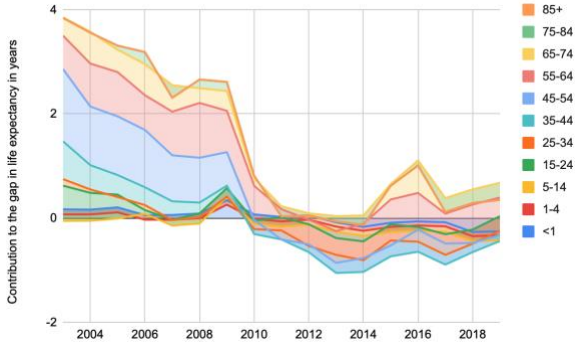
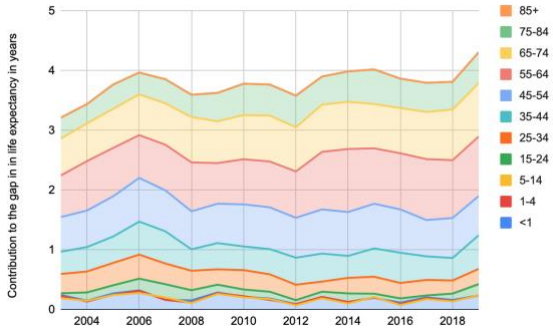


Figure 8.2: Mississippi



States with the highest life expectancy for males in the period 2003 to 2019

Figure 8.3: Vermont

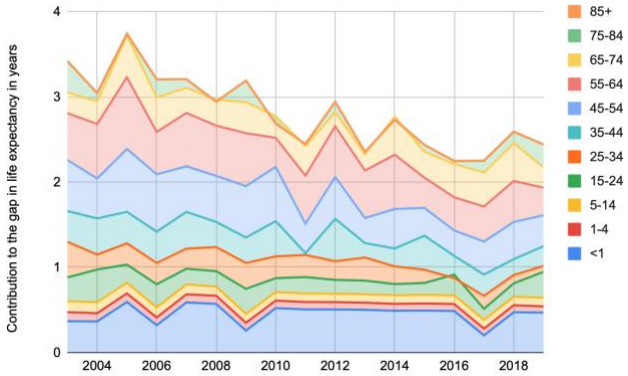


Figure 8.4: Utah

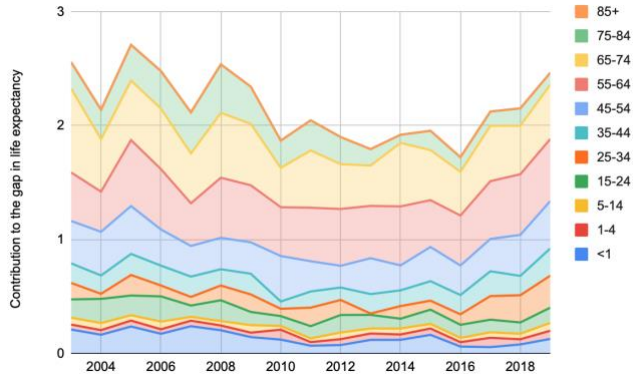


Figure 8.5: North Dakota

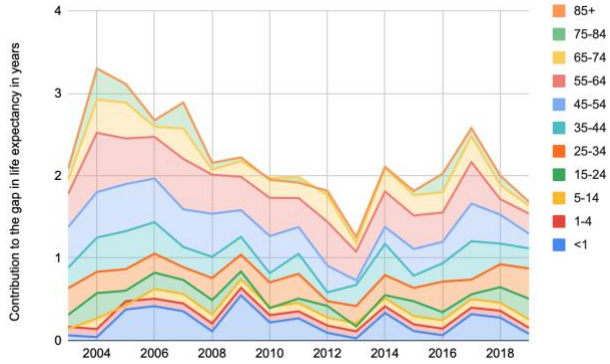
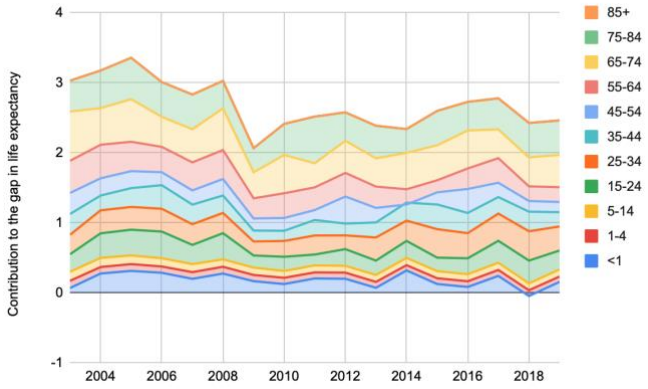


Figure 8.6: Hawaii



Source: Own figure based on CDC wonder and WHO Mortality database.

Figure 9: Cause specific decomposition of the difference in life expectancy with national average males

States with the lowest life expectancy for males in the period 2003 to 2019

Figure 9.1: District of Columbia

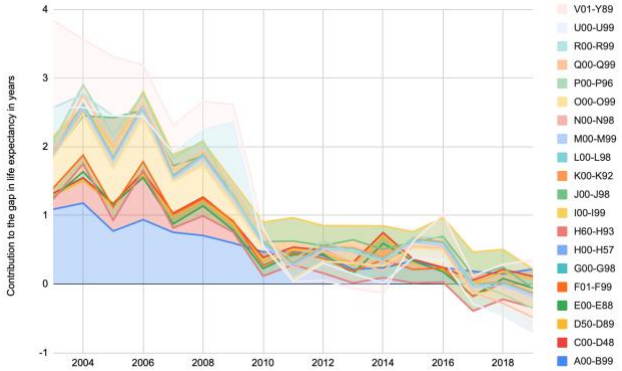
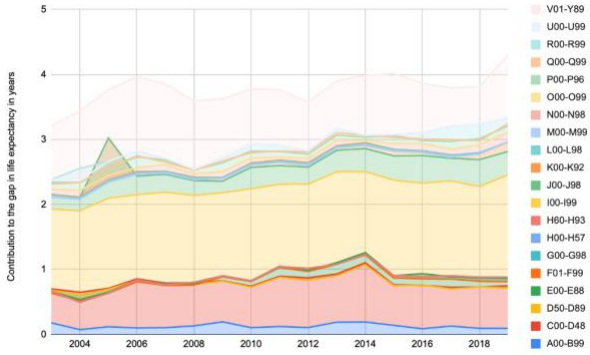


Figure 9.2: Mississippi



States with the lowest life expectancy for males in the period 2003 to 2019

Figure 9.3: Vermont

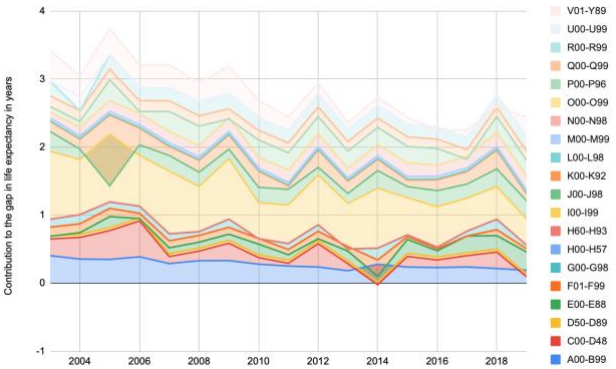


Figure 9.4: Utah

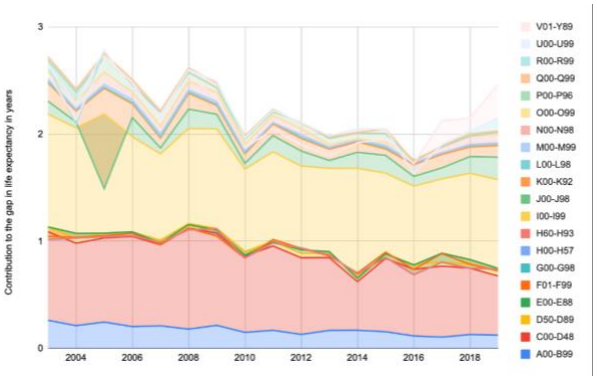


Figure 9.5: North Dakota

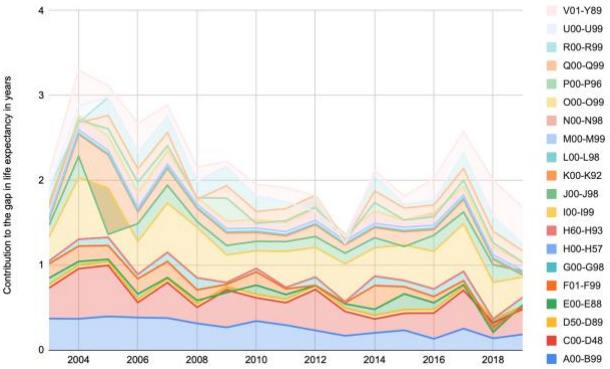
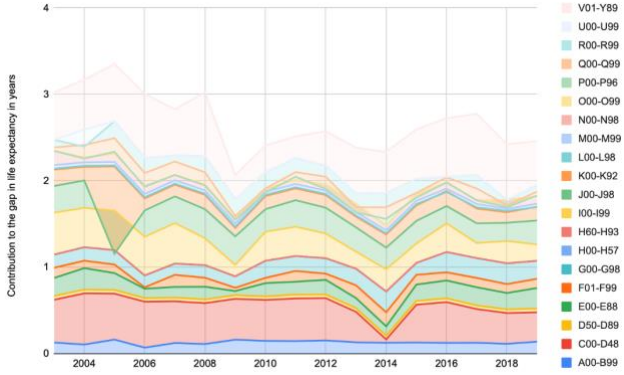


Figure 9.6: Hawaii



Source: Own figure based on CDC wonder and WHO Mortality database.

5. Conclusion

The research focused on developments in life expectancy in the US and how geographical inequalities have developed in the period 1999 to 2020 by looking into the laggard and pioneer states. In this chapter, the research questions will be answered, and the limitations and recommendations of this research are mentioned. The main research question is "*What drives changes and geographical inequalities in life expectancy across US states?*" This thesis focused mainly on what makes specific US states laggards or pioneers in life expectancy between 1999 and 2020.

In conclusion, this research has found that geographical inequality for females has increased for males the geographical inequality has decreased in the period 1999 to 2020. For females, the increase in the gap in life expectancy could be explained by that the states with the highest life expectancy made more increase in life expectancy than the states with the lowest life expectancy, although it is a small difference. The decrease in the gap in life expectancy for males could be explained by policies, which increased the life expectancy in the laggard states.

The differences in life expectancy in the laggard states for females are caused by one pattern in the age groups. The gap is mainly caused by older midlife and old age groups and the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality). The pattern of being a laggard is caused by the socioeconomic situation, which affects lifestyle factors. Besides that, also the living context and the socioeconomic composition of the population have an effect on why a state is a laggard.

The pioneer states for females show two different patterns in which age groups are contributing to the gap in life expectancy. One cluster of the pioneer states is caused by midlife age groups and the other cluster is caused by old age groups. In the pioneer states the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system), J00-J98 (Disease of the respiratory system) and V01-Y89 (External causes of morbidity and mortality) are big contributors to the gap in life expectancy. The two clusters of being a pioneer are related to migration and lifestyle factors, which influence the health situation.

The laggard states for males, the mortality pattern in the age groups is similar in the period 2003 to 2009 and after that, it starts to deviate. The gap is mainly caused by older midlife and old age groups and the causes of death C00-D48 (Neoplasms), I00-I99 (Diseases of the circulatory system) and V01-Y89 (External causes of morbidity and mortality). The difference in the pattern is caused by differences in policies, one state is making progress in life expectancy and the other state is not making progress. The laggard position is also related to the socioeconomic situation and lifestyle factors, like food consumption and physical activity.

The pioneer states for males show two patterns in the age groups in the contribution to the gap in life expectancy for males. One cluster is caused by midlife age groups and the other cluster is caused by old age groups. In the pioneer states for males, the causes of death C00-D48 (Neoplasms) and I00-I99 (Diseases of the circulatory system) are the biggest contributors. The different patterns are related to lifestyle factors, like living context and migration.

In all the states with the lowest and highest life expectancy differences are visible in the causes of death, which contributed the most to the difference in life expectancy besides the previously mentioned causes of death.

5.1 Limitations

Some points should be taken into account by interpreting the final results and conclusion of this research. An important point to consider is the geographical scale of the study. This study looked at differences in life expectancy at the state level. A disadvantage of looking at the state level is that the inequalities in life expectancy within states are not visible in this research anymore, which is related to the Modifiable Areal Unit Problem. This could affect that the

counties with the lowest or highest life expectancy are not in the states with the highest or lowest life expectancy, which influence the pattern of geographical inequality.

Another limitation of the research is that this research only focused on the states with the highest and lowest values of life expectancy, however, geographical inequality could also be increasing if the variability among the other states is increasing. When the variability is increasing, the states that are not laggards or pioneers have bigger gaps in life expectancy between them, which has an effect on the geographical inequality of life expectancy.

Besides that, a limitation of this research is that the data contains suppressed values, which did not have a big influence on the results by comparing the number of suppressed values with the national death numbers. The suppressed values are related to limited access to the data. However, by treating the data as mentioned in chapter 3, it was possible to do the analysis, although, it could lead to minor differences in the contribution of every cause of death and age group to the gap in life expectancy.

Another limitation is that the package DemoDecomp contains an issue which gives no contribution to the gap in life expectancy in the 85+ age group. The total gap in life expectancy which is calculated by the package will be the same when the 85+ age group is included, only the values of the other age groups will slightly change. Unfortunately, this could not be fixed for the final version of the thesis.

In this research, the geographical inequalities in life expectancy are only researched by looking into public health as an indicator for life expectancy, which is related to data limitations. However, more factors, like population composition, context and economic development have an influence on how life expectancy develops. By looking into the literature, it shows also that the composition, context and socioeconomic situation have an influence on why some states are laggards or pioneers in life expectancy. By including more factors in the analyses, it will be better possible to see which other factors also influence life expectancy and in which causes of death and age groups these factors have a big contribution to the gap in life expectancy.

5.2 Further research and policy recommendations

Further research should be on which factors explain the gap in life expectancy, especially for the states which are pioneers in life expectancy. This research shows that for the laggard states, much research is done into why these states are lagging. On the other hand, for some of the pioneering states, few explanations are available to explain why these states are pioneers in life expectancy. Focusing more on why states are pioneer states, it could help policymakers to make policies to decrease the gap in life expectancy or it could show policymakers if the policies are causing the mortality advantage or not.

For policymakers, it is recommended to look into which age groups and which causes of death contribute to the decreasing or stalling or increasing life expectancy, to make policies to increase life expectancy or to see how policies are over time influencing life expectancy developments. Especially in the states who are lagging in life expectancy policymakers should focus on decreasing the effect of the cause of death I00-I99 (Diseases of the circulatory system) on the gap in life expectancy. Policymakers could use this knowledge to implement policies which are making healthier lifestyles possible.

For the pioneer states, policymakers could use this research to see which age groups and causes of death already have a mortality advantage and in which age groups, and compare themselves with other pioneer states which show a different pattern in being a pioneer state and learn from each other. Policymakers could use this to implement specific plans to increase the healthy lifestyle for the old age groups or for the midlife age groups or to improve traffic safety. By implementing the policies at the state level, policymakers could focus first on what is the biggest contributor to the gap in life expectancy in a specific state and try to solve this.

References

Acevedo Pérez, I., Eugenia Rapiman, M., Cáneo Orellana, M., & Rueda Castro, L. (2017). Seven Ethical Requirements for Quantitative and Qualitative Research in Nursing: Experiences of three Research Ethics Committees from Santiago, Chile. *International Journal of Humanities and Social Science*, 19–24. https://www.ijhssnet.com/journals/Vol_7_No_7_July_2017/3.pdf

Agner, J., Pirkle, C. M., Irvin, L., Maddock, J. E., Buchthal, O. V., Yamauchi, J., Starr, R., & Sentell, T. (2020). The Healthy Hawai'i Initiative: insights from two decades of building a culture of health in a multicultural state. *BMC Public Health*, 20(1). https://doi.org/10.1186/s12889-019-8078-1

Akil, L., & Ahmad, H. A. (2011). Relationships between Obesity and Cardiovascular Diseases in Four Southern States and Colorado. *Journal of Health Care for the Poor and Underserved*, 22(4A), 61–72. https://doi.org/10.1353/hpu.2011.0166

Aly, S., Daniel, C. L., Bae, S., Scarinci, I. C., Hardy, C. M., Fouad, M. N., Baskin, M. L., Hoenemeyer, T., Acemgil, A., & Demark-Wahnefried, W. (2021). Cancer-related Beliefs and Preventive Health Practices among Residents of Rural versus Urban Counties in Alabama. *Cancer Prevention Research*, 14(5), 593–602. https://doi.org/10.1158/1940-6207.capr-20-0458

Amarasinghe, A., D'Souza, G., Brown, C., Oh, H., & Borisova, T. (2009). The Influence of Socioeconomic and Environmental Determinants on Health and Obesity: A West Virginia Case Study. *International Journal of Environmental Research and Public Health*, 6(8), 2271–2287. https://doi.org/10.3390/ijerph6082271

Andreev, E., Shkolnikov, V., & Begun, A. Z. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, 7, 499–522. https://doi.org/10.4054/demres.2002.7.14

Andreev, E. M., & Shkolnikov, V. M. (2012). *An Excel spreadsheet for the decomposition of a difference between two values of an aggregate demographic measure by stepwise*

replacement running from young to old ages. Max Planck Institute for Demographic Research. https://www.readcube.com/articles/10.4054%2Fmpidr-tr-2012-002

Arias, E., Tejada-Vera, B., & Ahmad, F. (2021). *Provisional Life Expectancy Estimates for January through June, 2020*. https://www.cdc.gov/nchs/data/vsrr/VSRR10-508.pdf

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

Auger, N., Feuillet, P., Martel, S., Lo, E., Barry, A. D., & Harper, S. (2014). Mortality inequality in populations with equal life expectancy: Arriaga's decomposition method in SAS, Stata, and Excel. *Annals of Epidemiology*, 24(8), 575–580. https://doi.org/10.1016/j.annepidem.2014.05.006

Babbie, R. (2020). *The Practice of Social Research*. Cengage Learning.

Barbieri, M. (2021). *United States Mortality DataBase*. Retrieved on February 21, 2022 from https://usa.mortality.org

Bennett, J. E., Tamura-Wicks, H., Parks, R. M., Burnett, R. T., Pope, C. A., Bechle, M. J., Marshall, J. D., Danaei, G., & Ezzati, M. (2019). Particulate matter air pollution and national and county life expectancy loss in the USA: A spatiotemporal analysis. *PLOS Medicine*, 16(7), e1002856. https://doi.org/10.1371/journal.pmed.1002856

Canudas Romo, V. (2003). *Decomposition Methods in Demography*. s.n.

Carey, J. R. (1989). The multiple decrement life table: a unifying framework for cause-of-death analysis in ecology. *Ecologia*, 78(1), 131–137. https://doi.org/10.1007/bf00377208

CDC. (n.d.-a). *Standards to Facilitate Data Sharing and Use of Surveillance Data for Public Health Action*. Retrieved on February 19, 2022 from https://www.cdc.gov/nchhstp/programintegration/sc-standards.htm

CDC. (n.d.-b). *Underlying Cause of Death 1999–2020*. Retrieved on February 23, 2022 from https://wonder.cdc.gov/wonder/help/ucd.html

CDC. (n.d.-c). *Underlying Cause of Death, 1999–2020 Request*. Retrieved on February 18, 2022 from https://wonder.cdc.gov/ucd-icd10.html

CDC. (n.d.-d). *Understanding Death Data Quality: Cause of Death from Death Certificates*. https://www.cdc.gov/nchs/data/nvss/coronavirus/cause-of-death-data-quality.pdf

Colgrove, J. (2002). The McKeown Thesis: A Historical Controversy and Its Enduring Influence. *American Journal of Public Health*, 92(5), 725–729. https://doi.org/10.2105/ajph.92.5.725

Couillard, B. K., Foote, C. L., Gandhi, K., Meara, E., & Skinner, J. (2021). Rising Geographic Disparities in US Mortality. *Journal of Economic Perspectives*, 35(4), 123–146. https://doi.org/10.1257/jep.35.4.123

Crimmins, E. M., Cohen, B., & Preston, S. H. (2011). *International Differences in Mortality at Older Ages: Dimensions and Sources*. National Academies Press.

Crimmins, E. M., Preston, S. H., & B, C. (2010). *International differences in Mortality at Older ages: Dimensions and sources.* National academies Press. https://www.ncbi.nlm.nih.gov/books/NBK62591/

Crimmins, E. M., Shim, H., Zhang, Y. S., & Kim, J. K. (2019). Differences between Men and Women in Mortality and the Health Dimensions of the Morbidity Process. *Clinical Chemistry*, 65(1), 135–145. https://doi.org/10.1373/clinchem.2018.288332

Dollar, N., Gutin, I., Lawrence, E. M., Braudt, D. B., Fishman, S. H., Rogers, R. G., & Hummer, R. A. (2020). The persistent southern disadvantage in US early life mortality, 1965–2014. *Demographic Research*, 42, 343–382. https://doi.org/10.4054/demres.2020.42.11

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., Van Lenthe, F. J., Mokdad, A. H., & Murray, C. J. L. (2017). Inequalities in Life Expectancy Among US Counties, 1980 to 2014. *JAMA Internal Medicine*, 177(7), 1003. https://doi.org/10.1001/jamainternmed.2017.0918

Engelman, M., Seplaki, C. L., & Varadhan, R. (2017). A Quiescent Phase in Human Mortality? Exploring the Ages of Least Vulnerability. *Demography*, 54(3), 1097–1118. https://doi.org/10.1007/s13524-017-0569-z

Fenelon, A. (2013). Geographic Divergence in Mortality in the United States. *Population and Development Review*, 39(4), 611–634. https://doi.org/10.1111/j.1728-4457.2013.00630.x

Frenk, J., J.L., B., Stern, C., Frejka, T., & Lozano, R. (1991). Elements for a theory of the health transition. *Health transition review*, 21–38.

Government of the District of Columbia. (2013). *District of Columbia Community Health Needs Assessment*.

Government of the District of Columbia department of health. (n.d.). *90/90/90/50 Plan ending the HIV Epidemic in the District of Columbia by 2020*. https://doh.dc.gov/sites/default/files/dc/sites/doh/page_content/attachments/DC%2090-90-90-50%20Plan%20-%20FINAL.pdf

Harper, S., Riddell, C. A., & King, N. B. (2021). Declining Life Expectancy in the United States: Missing the Trees for the Forest. *Annual Review of Public Health*, 381–403. https://doi.org/10.1146/annurev-publhealth-082619-104231

Harris, K. M., Majmundar, M. K., Becker, T., Sciences, E. N. A., Medicine, N. A. S. E., Education, D. B. S. S., Statistics, C. N., Population, C., & Disparities, C. R. M. M. R. S. (2021). *High and Rising Mortality Rates Among Working-Age Adults*. Amsterdam University Press.

Hazen, H., & Anthamatten, P. (2020). *An Introduction to the Geography of Health*. 2nd Edition. Routledge.

Hinde, A. (1998). *Demographic Methods*. Taylor & Francis.

Horiuchi, S., Wilmoth, J. R., & Pletcher, S. D. (2008). A decomposition method based on a model of continuous change. *Demography*, 45(4), 785–801. https://doi.org/10.1353/dem.0.0033

Hua, J., Ning, P., Cheng, P., Rao, Z., He, J., Xiao, W., Li, L., Fu, Y., Li, R., Li, J., Wang, W., Schwebel, D. C., & Hu, G. (2021). Coding Quality of Deaths and its Impact on Elderly Unintentional Fall Mortality Data from 1990 to 2019: A Retrospective Analysis of the WHO Mortality Database. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3907205

Human Mortality Database. (n.d.). *U.S.A.* www.mortality.org. Retrieved on May 7, 2022 from https://www.mortality.org/cgi-bin/hmd/country.php?cntr=USA&level=1

Jackson, J. (2014). The Consequences of Gentrification for Racial Change in Washington, DC. *Housing Policy Debate*, 25(2), 353–373. https://doi.org/10.1080/10511482.2014.921221

Kret, J., Gebeyehu, S., & Diggs-Outlaw, R. (2014). *The Burden of Cardiovascular Disease in the District of Columbia*. Government of the District of Columbia department of health. https://doh.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/DC%20CVD%20REPORT_Final%20v1.pdf

Land, K. C., Guralnik, J. M., & Blazer, D. G. (1994). Estimating Increment-Decrement Life Tables with Multiple Covariates from Panel Data: The Case of Active Life Expectancy. *Demography*, 31(2), 297–319. https://doi.org/10.2307/2061887

Leyland, A. H., & Groenewegen, P. P. (2020). *Multilevel Modelling for Public Health and Health Services Research*. Amsterdam University Press. https://doi.org/10.1007/978-3-030-34801-4.pdf

Majewska, J. (2017). The Quality of Mortality Data. *Infromatyka i Ekonometria*, 2017(344), 76–96.

Mehta, N. K., Abrams, L. R., & Myrskylä, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. *Proceedings of the National Academy of Sciences*, 117(13), 6998–7000. https://doi.org/10.1073/pnas.1920391117

Mississippi State Department of Health. (n.d.). *Burden of Cancer-Mississippi*. https://msdh.ms.gov/msdhsite/_static/resources/6191.pdf

Mokdad, A. H., Ballestros, K., Echko, M., Glenn, S., Olsen, H. E., Mullany, E., Lee, A., Khan, A. R., Ahmadi, A., Ferrari, A. J., Kasaeian, A., Werdecker, A., Carter, A., Zipkin, B., Sartorius, B., Serdar, B., Sykes, B. L., Troeger, C., Fitzmaurice, C., . . . Murray, C. J. L. (2018). The State of US Health, 1990–2016. *JAMA*, 319(14), 1444. https://doi.org/10.1001/jama.2018.0158

Montez, J. K., Beckfield, J., Cooney, J. K., Grumbach, J. M., Hayward, M. D., Koytak, H. Z., Woolf, S. H., & Zajacova, A. (2020). US State Policies, Politics, and Life Expectancy. *The Milbank Quarterly*, 98(3), 668–699. https://doi.org/10.1111/1468-0009.12469

Muennig, P. A., Reynolds, M., Fink, D. S., Zafari, Z., & Geronimus, A. T. (2018). America's Declining Well-Being, Health, and Life Expectancy: Not Just a White Problem. *American Journal of Public Health*, 108(12), 1626–1631. https://doi.org/10.2105/ajph.2018.304585

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2021). *Mortality in the United States, 2020*. Retrieved on February 16, 2022 from https://www.cdc.gov/nchs/products/databriefs/db427.htm

Murthy, P. K. (2005). A comparison of different methods for decomposition of changes in expectation of life at birth and differentials in life expectancy at birth. *Demographic Research*, 12, 141–172. https://doi.org/10.4054/demres.2005.12.7

Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. *Methodological Issues in Health and Mortality*.

Omran, A. R. (1971). The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *The Milbank Memorial Fund Quarterly*, 49(4), 509. https://doi.org/10.2307/3349375

Omran, A. R. (1998). The epidemiologic transition theory revisited thirty years later. *Medicine*, 99–119. https://apps.who.int/iris/bitstream/handle/10665/330604/WHSQ-1998-51-n2-3-4-eng.pdf

Ortiz-Ospina, E. (2017). *"Life Expectancy" – What does this actually mean?* Retrieved on May 6, from 2022 https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted

Preston, S., Heuveline, P., & Guillot, M. (2000). *Demography*. Wiley.

Rau, R., Soroko, E., Jasilionis, D., & Vaupel, J. W. (2008). Continued Reductions in Mortality at Advanced Ages. *Population and Development Review*, 34(4), 747–768. https://doi.org/10.1111/j.1728-4457.2008.00249.x

Riddell, C. A., Morrison, K. T., Kaufman, J. S., & Harper, S. (2018). Trends in the contribution of major causes of death to the black-white life expectancy gap by US state. *Health & Place*, 52, 85–100. https://doi.org/10.1016/j.healthplace.2018.04.003

Riffe, T. (2018). *Package 'DemoDecomp'*. https://cran.r-project.org/web/packages/DemoDecomp/DemoDecomp.pdf

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

Roser, M., Ortiz-Ospina, E., & Ritchie, H. (2013). *Life Expectancy*. Our World in Data. Retrieved on February 18, 2022 from https://ourworldindata.org/life-expectancy

Santosa, A., Wall, S., Fottrell, E., Högberg, U., & Byass, P. (2014). The development and experience of epidemiological transition theory over four decades: a systematic review. *Global Health Action*, 7(1), 23574. https://doi.org/10.3402/gha.v7.23574

Srikanthan, K., Feyh, A., Visweshwar, H., Shapiro, J. I., & Sodhi, K. (2016). Systematic Review of Metabolic Syndrome Biomarkers: A Panel for Early Detection, Management, and Risk Stratification in the West Virginian Population. *International Journal of Medical Sciences*, 13(1), 25–38. https://doi.org/10.7150/ijms.13800

Sudharsanan, N., Aburto, J. M., Riffe, T., & Van Raalte, A. (2022). Commentary: Large variation in the epidemiological transition across countries: is it still valuable as a mortality theory? *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyac107

Swanson, D. A., Kposowa, A., & Baker, J. (2020). Estimating the Underlying Death Rate of a Small Population: A Case Study of Counties in Kansas, Nebraska, North Dakota, and South Dakota. *Transactions of the Kansas Academy of Science*, 123(3–4). https://doi.org/10.1660/062.123.0303

Tomblin, E. R., & Lewis, M. J. (2011). *Obesity in West Virginia*. Department of Health and Human Resources. https://www.wvdhhr.org/bph/hsc/pubs/other/obesityreport2011/obesityreport2011.pdf

U.S. Department of Health and Human Services. (2004). *Vital Statistics of United States*. https://wonder.cdc.gov/wonder/help/CMF/TechnicalAppendix1999.pdf

Vallin, J., & Meslé, F. (2004). Convergences and divergences in mortality. *Demographic Research, Special 2*, 11–44. https://doi.org/10.4054/demres.2004.s2.2

Vallin, J., & Meslé, F. (2008). Minimum Mortality: A Predictor of Future Progress? *Population*, 63(4), 557–590. https://doi.org/10.3917/pope.804.0557

Vermont Department of Health. (2021). *2025 Vermont Cancer Plan*. https://www.healthvermont.gov/sites/default/files/documents/pdf/hpdp_CompCancer_2025-CancerPlan-Web-12.22-FullReport.pdf

Wang, H., Schumacher, A. E., Levitz, C. E., Mokdad, A. H., & Murray, C. J. (2013). Left behind: widening disparities for males and females in US county life expectancy, 1985–2010. *Population Health Metrics*, 11(1). https://doi.org/10.1186/1478-7954-11-8

Whitmore Schanzenbach, D., Nunn, R., & Bauer, L. (2016). The Changing Landscape of American Life Expectancy. *The Hamilton project*, 1–20.

WHO. (2020). *Data Quality*. Retrieved on May 3, 2022 from https://platform.who.int/mortality/about/data-quality

WHO. (1979). *Medical Certification of Cause of Death*. https://apps.who.int/iris/bitstream/handle/10665/40557/9241560622.pdf

WHO. (2022). *WHO Mortality Database*. Retrieved on May 3, from 2022 https://www.who.int/data/data-collection-tools/who-mortality-database

Williamson, M. R., & Ahmed, R. M. (2021). Cancer Mortality and Research Outcomes in a Rural State. *WMJ*, 120–126. https://wmjonline.org/wp-content/uploads/2021/120/2/120.pdf

Woolf, S. H., & Aron, L. (2018). Failing health of the United States. https://doi.org/10.1136/bmj.k496

Woolf, S. H., Schoomaker, H., Hill, L., & Orndahl, C. M. (2019). The Social Determinants of Health and the Decline in U.S. Life Expectancy: Implications for Appalachia. *Journal of Appalachian Health*, 6–14. https://doi.org/10.13023/jah.0101.02

Woolf, S. H., & Schoomaker, H. (2019). Life Expectancy and Mortality Rates in the United States, 1959–2017. *JAMA*, 322 (20), 1996. https://doi.org/10.1001/jama.2019.16932

Wu, Y., Braun, K., Onaka, A. T., Horiuchi, B. Y., Tottori, C. J., & Wilkens, L. (2017). Life Expectancies in Hawai'i: A Multi-ethnic Analysis of 2010 Life Tables. *Hawaii J Med Public Health*, 2017, 9–14. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5226016/

Xu, W., Topping, M., & Fletcher, J. (2021). State of birth and cardiovascular disease mortality: Multilevel analyses of the National Longitudinal Mortality Study. *SSM - Population Health*, 15, 100875. https://doi.org/10.1016/j.ssmph.2021.100875

Yorgason, J. B., Draper, T. W., Bronson, H., Nielson, M., Babcock, K., Jones, K., Hill, M. S., & Howard, M. (2018). Biological, Psychological, and Social Predictors of Longevity Among Utah Centenarians. *The International Journal of Aging and Human Development*, 87(3), 225–243. https://doi.org/10.1177/0091415018757211

Zarulli, V., Kashnitsky, I., & Vaupel, J. W. (2021). Death rates at specific life stages mold the sex gap in life expectancy. *Proceedings of the National Academy of Sciences*, 118(20). https://doi.org/10.1073/pnas.2010588118

Appendix 1 Proportion minimum and maximum suppressed cells

Table 4: Proportion minimum and maximum suppressed cells for every age groups in a year for females

	<1	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
1999	Min 0.5 Max 1	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.45 Max 1	Min 0.25 Max 1	Min 0.25 Max 0.9	Min 0.3 Max 0.85	Min 0.25 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.75
2000	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.55 Max 1	Min 0.45 Max 1	Min 0,3 Max 1	Min 0,25 Max 0,85	Min 0,25 Max 0.85	Min 0.25 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.8
2001	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.6 Max 1	Min 0.45 Max 1	Min 0.3 Max 1	Min 0.25 Max 0.9	Min 0.25 Max 0.85	Min 0.2 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.70
2002	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.35 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.85	Min 0.25 Max 0.80	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.7
2003	Min 0.55 Max 1	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.35 Max 0.95	Min 0.25 Max 0.95	Min 0.2 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.80	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.65
2004	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.4 Max 1	Min 0.25 Max 1	Min 0.25 Max 0.9	Min 0.2 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.65	Min 0.25 Max 0.7
2005	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.35 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.95	Min 0.2 Max 0.75	Min 0.25 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.65	Min 0.25 Max 0.65
2006	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.4 Max 1	Min 0.25 Max 1	Min 0.2 Max 0.9	Min 0.25 Max 0.85	Min 0.25 Max 0.85	Min 0.25 Max 0.75	Min 0.25 Max 0.6	Min 0.25 Max 0.65
2007	Min 0.45 Max 1	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.4 Max 1	Min 0.3 Max 0.95	Min 0.25 Max 0.9	Min 0.2 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.55	Min 0.25 Max 0.6
2008	Min 0.5 Max 1	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.4 Max 0.95	Min 0.25 Max 1	Min 0.25 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.60	Min 0.25 Max 0.60
2009	Min 0.55 Max 1	Min 0.7 Max 1	Min 0.6 Max 1	Min 0.45 Max 0.9	Min 0.3 Max 1	Min 0.25 Max 0.95	Min 0.2 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.6	Min 0.25 Max 0.6

2010	Min 0.55 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.4 Max 1	Min 0.3 Max 1	Min 0.25 Max 1	Min 0.25 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.6	Min 0.25 Max 0.55
2011	Min 0.55 Max 1	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.4 Max 1	Min 0.3 Max 0.95	Min 0.25 Max 1	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.55
2012	Min 0.55 Max 1	Min 0.7 Max 1	Min 0.7 Max 1	Min 0.4 Max 1	Min 0.3 Max 0.95	Min 0.25 Max 0.9	Min 0.2 Max 0.85	Min 0.25 Max 0.75	Min 0.25 Max 0.65	Min 0.25 Max 0.6	Min 0.25 Max 0.6
2013	Min 0.6 Max 1	Min 0.7 Max 1	Min 0.7 Max 1	Min 0.4 Max 1	Min 0.25 Max 1	Min 0.25 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.5
2014	Min 0.65 Max 1	Min 0.75 Max 1	Min 0.7 Max 1	Min 0.45 Max 1	Min 0.25 Max 0.95	Min 0.25 Max 1	Min 0.2 Max 0.8	Min 0..25 Max 0.6	Min 0.25 Max 0.7	Min 0.25 Max 0.6	Min 0.25 Max 0.55
2015	Min 0.6 Max 1	Min 0.7 Max 1	Min 0.7 Max 1	Min 0.4 Max 1	Min 0.25 Max 0.95	Min 0.25 Max 0.9	Min 0.2 Max 0.85	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.55	Min 0.25 Max 0.5
2016	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.4 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.9	Min 0.2 Max 0.85	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.65	Min 0.25 Max 0.5
2017	Min 0.55 Max 1	Min 0.75 Max 1	Min 0.65 Max 1	Min 0.4 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.55	Min 0.25 Max 0.5
2018	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.45 Max 1	Min 0.25 Max 0.95	Min 0.25 Max 0.9	Min 0.25 Max 0.85	Min 0.25 Max 0.65	Min 0.25 Max 0.6	Min 0.25 Max 0.6	Min 0.25 Max 0.55
2019	Min 0.55 Max 1	Min 0.7 Max 1	Min 0.65 Max 1	Min 0.5 Max 1	Min 0.25 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.7	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.5
2020	Min 0.6 Max 1	Min 0.8 Max 1	Min 0.7 Max 1	Min 0.4 Max 1	Min 0.2 Max 0.95	Min 0.2 Max 0.9	Min 0.2 Max 0.85	Min 0.2 Max 0.65	Min 0.2 Max 0.55	Min 0.2 Max 0.45	Min 0.2 Max 0.45

Source: Own table based on CDC wonder database

Table 5: Proportion minimum and maximum suppressed cells for every age groups in a year for males

	<1	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
1999	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.55 Max 1	Min 0.40 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.85	Min 0.3 Max 0.8	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.85
2000	Min 0.45 Max 0.95	Min 0.55 Max 1	Min 0.55 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.85	Min 0.3 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.85
2001	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.55 Max 1	Min 0.4 Max 1	Min 0.30 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.85	Min 0.25 Max 0.70	Min 0.25 Max 0.75	Min 0.25 Max 0.70	Min 0.25 Max 0.85
2002	Min 0.4 Max 1	Min 0.65 Max 1	Min 0.55 Max 1	Min 0.4 Max 0.95	Min 0.35 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.75	Min 0.25 Max 0.70	Min 0.25 Max 0.85
2003	Min 0.50 Max 0.95	Min 0.55 Max 1	Min 0.65 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.70	Min 0.25 Max 0.70	Min 0.25 Max 0.65	Min 0.25 Max 0.85
2004	Min 0.45 Max 0.95	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.4 Max 0.95	Min 0.35 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.8
2005	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.5 Max 1	Min 0.4 Max 0.95	Min 0.35 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.70	Min 0.25 Max 0.65	Min 0.25 Max 0.75
2006	Min 0.4 Max 1	Min 0.65 Max 1	Min 0.6 Max 1	Min 0.4 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.6	Min 0.25 Max 0.55
2007	Min 0.45 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.4 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.75	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.6	Min 0.25 Max 0.65
2008	Min 0.45 Max 1	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.4 Max 0.95	Min 0.35 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.7	Min 0.25 Max 0.55	Min 0.25 Max 0.75
2009	Min 0.55 Max 1	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.55	Min 0.25 Max 0.75

2010	Min 0.5 Max 1	Min 0.7 Max 1	Min 0.65 Max 1	Min 0.45 Max 0.95	Min 0.35 Max 0.95	Min 0.3 Max 0.85	Min 0.25 Max 0.85	Min 0.25 Max 0.65	Min 0.25 Max 0.65	Min 0.25 Max 0.65	Min 0.25 Max 0.7
2011	Min 0.45 Max 1	Min 0.6 Max 1	Min 0.6 Max 1	Min 0.45 Max 0.95	Min 0.35 Max 0.95	Min 0.3 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.7	Min 0.25 Max 0.65	Min 0.25 Max 0.65	Min 0.25 Max 0.7
2012	Min 0.45 Max 1	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.75	Min 0.25 Max 0.7	Min 0.25 Max 0.6	Min 0.25 Max 0.6	Min 0.25 Max 0.75
2013	Min 0.5 Max 1	Min 0.7 Max 1	Min 0.6 Max 1	Min 0.4 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.75	Min 0.25 Max 0.6	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.7
2014	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.65	Min 0.25 Max 0.65	Min 0.25 Max 0.55	Min 0.25 Max 0.65
2015	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.6 Max 1	Min 0.45 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.75	Min 0.25 Max 0.65	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.6
2016	Min 0.55 Max 1	Min 0.65 Max 1	Min 0.6 Max 1	Min 0.4 Max 0.95	Min 0.3 Max 0.95	Min 0.3 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.6	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.7
2017	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.6 Max 1	Min 0.45 Max 0.95	Min 0.35 Max 0.95	Min 0.35 Max 0.9	Min 0.25 Max 0.8	Min 0.25 Max 0.6	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.55
2018	Min 0.55 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.5 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.9	Min 0.25 Max 0.85	Min 0.25 Max 0.65	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.55
2019	Min 0.55 Max 0.95	Min 0.6 Max 1	Min 0.65 Max 1	Min 0.4 Max 0.95	Min 0.3 Max 0.95	Min 0.25 Max 0.85	Min 0.25 Max 0.8	Min 0.25 Max 0.6	Min 0.25 Max 0.55	Min 0.25 Max 0.6	Min 0.25 Max 0.65
2020	Min 0.5 Max 1	Min 0.65 Max 1	Min 0.65 Max 1	Min 0.35 Max 0.95	Min 0.25 Max 0.95	Min 0.25 Max 0.85	Min 0.2 Max 0.75	Min 0.2 Max 0.6	Min 0.2 Max 0.5	Min 0.2 Max 0.5	Min 0.2 Max 0.6

Source: Own table based on CDC wonder database.

Appendix 2 Regression output suppressed cells

Table 6: Regression output

Deviance residuals				
Min	1Q	Median	3Q	Max
-4.1021	-0.3611	0.0205	0.3656	3.4315
Coefficients	**Estimate standard error**	**Error**	**z value**	**Pr(>|z|)**
Intercept	6.67E+00		0.048	0.961
Ten.Year.Age.Groups <1 year	1.49E+01		0.108	0.914
Ten.Year.Age.Groups 1-4 years	1.71E+01		0.124	0.901
Ten.Year.Age.Groups 15-24 years	1.52E+01		0.11	0.912
Ten.Year.Age.Groups 25-34 years	1.41E+01		0.102	0.918
Ten.Year.Age.Groups 35-44 years	1.31E+01		0.095	0.924
Ten.Year.Age.Groups 45-54 years	1.23E+01		0.09	0.929
Ten.Year.Age.Groups 5-14 years	1.65E+01		0.12	0.905
Ten.Year.Age.Groups 55-64 years	1.18E+01		0.086	0.931
Ten.Year.Age.Groups 65-74 years	1.16E+01		0.084	0.933
Ten.Year.Age.Groups 75-84 years	1.12E+01		0.082	0.935
Ten.Year.Age.Groups 85+ years	1.12E-01		0.081	0.935
ICD.Chapter.Code C00-D48	2.14E+01		-49.714	<2e-16***
ICD.Chapter.Code D50-D89	-1.31E+00		80.127	<2e-16***
ICD.Chapter.Code E00-E88	-2.66E-01		-10.456	<2e-16***
ICD.Chapter.Code F01-F99	5.59E-01		22.339	<2e-16***
ICD.Chapter.Code G00-G98	-1.79E-01		-7.04	1.92e-12***
ICD.Chapter.Code H00-H57	1.95E+01		0.599	0.549
ICD.Chapter.Code H60-H93	1.95E+01		0.599	0.549
ICD.Chapter.Code I00-I99	-1.09E+00		-41.794	<2e-16***
ICD.Chapter.Code J00-J98	-3.49E-01		-13.702	<2e-16***
ICD.Chapter.Code K00-K92	-2.48E-01		-9.754	<2e-16***
ICD.Chapter.Code L00-L98	3.22E+00		113.962	<2e-16***
ICD.Chapter.Code M00-M99	1.72E+00		68.773	<2e-16***
ICD.Chapter.Code N00-N98	8.39E-01		33.647	<2e-16***
ICD.Chapter.Code O00-O99	5.17E+00		109.15	<2e-16***
ICD.Chapter.Code P00-P96	3.68E+00		120.286	<2e-16***
ICD.Chapter.Code Q00-Q99	2.54E+00		97.234	<2e-16***
ICD.Chapter.Code R00-R99	5.78E-01		23.109	<2e-16***
ICD.Chapter.Code U00-U99	5.00E+00		112.47	<2e-16***
ICD.Chapter.Code V01-Y89	-3.24E+00		-103.977	<2e-16***
years	-1.01E-02	7.15E-04	-14.13	<2e-16***
sexmale	-1.53E-01	9.07E-03	-16.872	<2e-16***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Null deviance: 749110 on 567731 degrees of freedom				
Residual deviance: 311761 on 567697 degrees of freedom				

Source: Own table based on CDC wonder and WHO Mortality database.

Appendix 3 Map of the US states

Figure 10: Map of the US states

Source: (OntheWorldMap, 2020).