Erasmus University Rotterdam
Erasmus School of Economics

Bachelor Thesis Economie & Bedrijfseconomie

Racial differences in two-year mortality rates in the US for the years 1996, 2006 and 2016

Name student: Bob Melzer
Student ID number: 483791

Supervisor: Teresa Marreiros Bago d'Uva
Second assessor: Elisa de Weerd

Date final version: 27-09-2021

Abstract

The aim of this research is to find out if there is an association between the two-year mortality and ethnicity for the years 1996, 2006 and 2016. There is evidence that the mortality rates for different racial groups follow a different trend, which suggests that there are racial differences. The analysis that is going to be used in this paper is a multiple linear regression model with robust standard errors. In the baseline model there is evidence for racial inequality in the two-year mortality rate. Nonetheless in the final model there is no evidence for racial inequalities.

Introduction

In relatively recent history the mortality rate around the world changed significantly. Cutler, Deaton & Lleras-Muney (2006) stated in their paper that the drop in mortality rate can be attributed to three phases. The first phase was from the middle of the eighteenth century till the middle of the nineteenth century. Improved nutrition and economic growth helped to reduce mortality rates. Public health measures started to play a role as well, but not as much as the former two. The second phase was from the closing decades of the nineteenth century into the beginning of the twentieth century. Now public health measures were the leading cause for the improvement in mortality rates. The last phase started around the 1930s and is still going on till this day. This is the era of big medicine where vaccinations and antibiotics were discovered. It is remarkable to see how much can change in roughly 250 years. An example also given by Cutler, Deaton & Llera-Muney (2006) where the decrease in mortality rate is visible, is that in the US and other high-income countries decline in mortality rates caused life expectancy at birth to grow by nearly 30 years through the twentieth century. The decrease in mortality rate ensures that people mostly live better, are healthier and live longer than their ancestors.

As mortality rates around the world were declining in the twentieth century, a different trend occurred in the US. Since the beginning of the twenty-first century the all-cause mortality of middle-aged white men and women is increasing, particularly among those with a high-school degree or less (Case & Deaton 2015). All other racial and ethnic groups still see their mortality rates fall. The increased all-cause mortality is a reversal in the progress made earlier in the US. There is no other first world country in the world with the same reversal. When looking at Europe, mortality rates for those with low levels of educational attainment are still falling (Case & Deaton 2017). And it is not only the all-cause mortality rates of middle-aged white men and women in the US that differ. Fuchs (2016) found that in the same period the increase of black life expectancy at birth was double the increase of white life expectancy at birth in the US. There is strong evidence that the mortality rates for different racial groups are evolving in different directions.

In the US the mortality rate for the ages 10 till 24 was 60.1 per 100000 in 2005 (Mulye et al 2009). This is a low mortality rate compared to mortality rate for older people. Woolf and Schoonmaker (2019) documented that the mortality rate for people in midlife is 384.2 per 100000 in 2014. Midlife in this research spans from 25 years old till 64 years old. This interval

is much bigger than the interval in the paper of Mulye. A direct comparison because of this is difficult. However, the mortality rate for people in midlife is more than six times higher than that from teens/adolescence. At least it shows that mortality rate is higher for older people. Wang et al (2017) researched the long-term trends in global mortality. They did not look at the mortality rates per age group, but at the overall mortality within the age group globally. They present their findings in Figure 1.

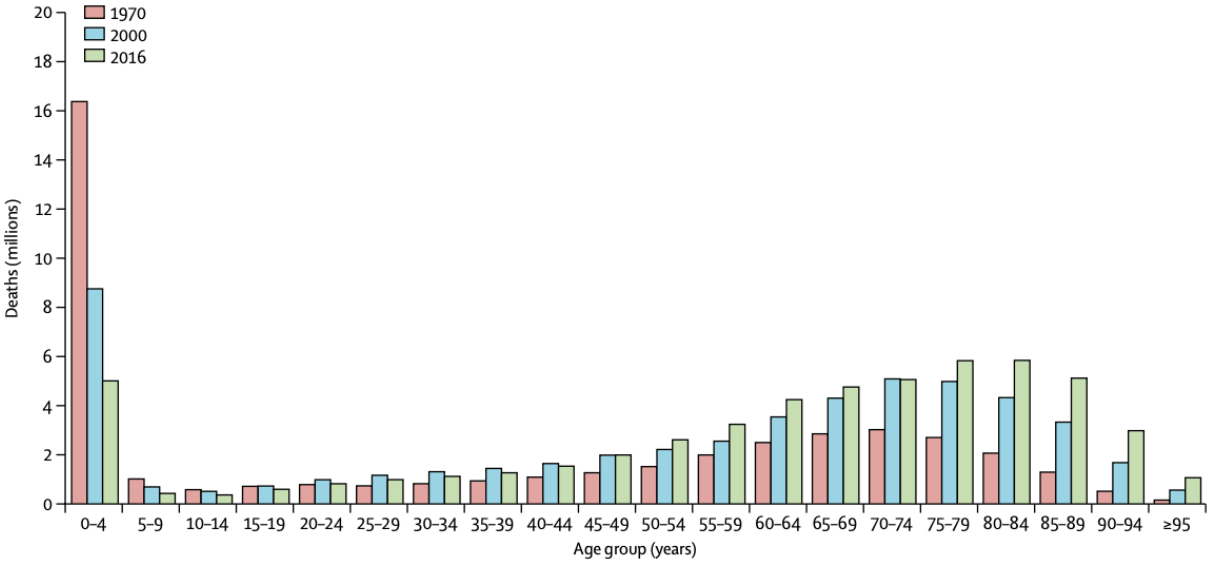

Figure 1: global deaths for all age groups for the years 1970, 2000 and 2016 from the paper from Wang et al (2017)

From the figure can be concluded that the trend for overall mortality for the ages 0 till 44 decreased, the trend for overall mortality for the ages 45 till >95 increased and that most deaths in 2016 are for people 50 years old or older. The fact that most death occurred for people 50 years and older is not shocking. Life expectancy around the world has increased. Gulland (2016) showed that life expectancy since 2000 has increased by five years. More people around the world make it to an older age, hence the higher number of deaths for the elderly. According to the United Nations, the world population older than 50 years accounts for 22.3 percent of the total world population in 2015. This makes it interesting to look further into the mortality rates for older men and women, since only a small part of the population accounts for the most deaths.

Apart from differences in mortality for racial groups it is also interesting to investigate if mortality rates within the US vary from one region to another. Mackenbach et al (1988) discovered that in The Netherlands there is important geographical variation in mortality

due to conditions amenable to medical intervention within the country. Proven regional differences in a small country, raises the question if these regional differences also exist in the US, being 250 times the size of The Netherlands.

As discussed above, mortality rates have decreased drastically since the past century. In the US however the mortality rates for Whites have increased. There is evidence that the mortality rates for different racial groups follow a different trend, which suggests that there are racial differences. The mortality rates are higher for older people. Most deaths globally are for people over 50 years, while being a small part of the population. For that reason this research is interested in the probability of dying between different racial groups for the ages around 50 and above. This is going to be examined by calculating the two-year mortality across multiple time periods. The different time periods are 1996, 2006 and 2016. The reason that this paper looks up and till 2016 is that the data used in this paper is only complete up and till 2016. The year 2016 is the closest year till now and will give the most recent representation. The 10-year gap between each point in time make the differences over time more visible.

The research question for this paper is: Are there racial differences in the two-year mortality in the US for the years 1996, 2006 and 2016? The research question can be divided into the following sub questions.

1. Are there racial differences in the two-year mortality? Case and Deaton already found that the all-cause mortality for middle-aged Whites follow a different trend than for the other racial and ethnic groups. Not only for all-cause mortality but also for life expectancy at birth do Whites and Blacks differ according to Fuchs. It will be interesting to find out if these differences occur when researching the two-year mortality for different racial groups.
2. Do these differences change over this period of time? Mortality rates and life expectancy have changed massively over the past century. In both positive and negative directions. In this paper a 20-year time period is used, which is sufficiently wide to investigate if in this period any significant differences change over time.
3. Are the racial differences explained by socio-economic factors? Differences in two-year mortality may be explained by numerous socio-economic factors. Examples of influences by socio-economic factors are that people who complete more education live longer (Meara, Richards & Cutler 2008) and that the life expectancy of the richest 1% is higher than for the

poorest 1% (Chetty et al 2016). In this paper, education, income, employment status and living area are used in the analysis to see if potential differences can be explained.
As mentioned earlier, middle-aged mortality for white men and women are increasing. This is a serious issue. It is important to see if this affects the two-year mortality and how this compares between different racial groups. By doing so possible inequalities can be revealed. This evidence can then be used to help the groups in need. To contribute to the literature this paper wants to assess if living in different regions of the US influences the two-year mortality. To my knowledge, the influence of different regions on the two-year mortality has not been researched.
The dataset that is going to be used in this research and the variables used in the analysis will be discussed next, after which the methodology containing the research set up is explained. Then the results will follow. Findings of the analysis are shown in this part. The paper ends with the conclusion and discussion.

Data

This section discusses the variables that are used in this research. The dataset used to conduct this research is the Health and Retirement Study (HRS)[1]. The HRS is a national panel survey of individuals over age 50 and their spouses. Its rich, sophisticated and its multidisciplinary data give researchers the opportunity to explore many different aspects related to population aging in the US. This is why the HRS will be used for this research. The HRS covers a large range of topics such as demographics, employment history, health and many more. The main goal is to provide panel data that enable research and analysis in support of policies. The HRS consists of seven cohorts divided over 14 waves. The seven cohorts and the corresponding waves are.

- 1992 (wave 1 initial HRS cohort) cohort contains US citizens born in 1931 to 1941
- 1993 (wave 2 AHEAD cohort) cohort contains US citizens born before 1924
- 1994 (wave 2 initial HRS cohort)
- 1995 (wave 3 AHEAD cohort)
- 1996 (wave 3 initial HRS cohort)
- 1998 (wave 4 CODA & WB cohort) cohort contains US citizens born 1924 to 1930 and 1942 to 1947 respectively
- 2000 (wave 5 CODA & WB cohort)
- 2002 (wave 6 CODA & WB cohort)
- 2004 (wave 7 EBB cohort) cohort contains US citizens born 1948 to 1953
- 2006 (wave 8 EBB cohort)
- 2008 (wave 9 EBB cohort)
- 2010 (wave 10 MBB cohort) cohort contains US citizens born 1954 to 1959
- 2012 (wave 11 MBB cohort)
- 2014 (wave 12 MBB cohort)
- 2016 (wave 13 LBB cohort) cohort contains US citizens born 1960 to 1965
- 2018 (wave 14 LBB cohort)

[1] The HRS is sponsored by the National Institute of Aging (grant number NIA U01AG009740), with additional funding from the Social Security Administration (SSA) and administered by the Institute of Social Research (ISR) and is conducted by the University of Michigan

The HRS consists of fourteen waves over sixteen survey years, starting in 1992 annually until 1995, continuing biennially from 1996 to 2018. The data that is collected from interviews is only complete up to and including 2016, which corresponds with wave 13.

The dependent variable of interest is died in next wave. Died in next wave equals one if the individual died between the current and the next wave (i.e., approximately within the next two years, given that the interviews are bi-annual), and zero otherwise.

The independent variable of interest is ethnicity. Ethnicity is a categorical variable with three categories and describes the race of the respondent. The value one is equal to White/Caucasian, which is also the reference category in the regression. The value two holds Black/African American and the value three is Other. The category other consists of the other racial groups in the US such as Hispanics, Native Americans and more.

In this part, the control variables that are used in the regression are debated starting with the first control variables that were inserted in the baseline model. These are age, a second order polynomial for age and gender. Age is a continuous variable ranging from 21 to 107 where age is the age of the respondent in a given wave. Gender is a categorical variable with two categories, where one is male and two is female and male is the reference category in the regression.

Now the control variables that are used to measure the differences due to socio-economic factors are explained, beginning with education. Education is a categorical variable with five categories. The first category is left high school, this is also the reference category. The second category to the fifth category respectively are GED graduate, high school graduate, some college and college and above. The next control variables clarified are total household income and number of people living in the household. Total household income is a continuous variable and gives the total household income in dollars. It captures the total annual income in dollars with the average being $62978.74. Number of people living in the household indicates how many people are living in the household in a given wave. It is a continuous variable with a minimum of 1 and maximum of 19. Employment status is the following control variable. Employment status is a categorical variable with seven categories ranging from 1 up to 7. The categories are works full time, works part time, unemployed, partly retired, retired and disabled in that order form 1 up to 7. The reference category is works full time. The last variable is region. Region explains in what region a respondent lives. It is a categorical variable with five categories where one is equal to Northeast, two is equal

to Midwest, three is equal to South, four is equal to West and five is equal to other. The reference category is Northeast.

Table 1 shows the descriptive statistics and Table 2 the proportion of people who died in the next wave separated per racial group. Notable descriptive statistics show that there are more men than women in the dataset, more Whites/Caucasians than Blacks/African Americans and Other and the overall two-year mortality is 5.47 percent points. In Table 2 the overall two-year mortality for White/Caucasian is 6.15 percent point, for Black/African American is 5.86 percent point and for Other is 2.86 percent point.

Table 1: descriptive statistics

	N	**Mean**	**Std Dev**	**Min**	**Max**
Died in next wave	69,822	.0554	.2287	0	1
Ethnicity	126,360	1.3573	1.3573	1	3
Age	57,372	66.6475	11.4073	21	107
Gender	126,696	1.5607	.4963	1	2
Education	126,633	3.0497	1.4329	1	5
Total household income	57,372	62978.74	209689.7	0	$2.54 \cdot 10^7$
Number of people living in the household	57,372	2.2928	1.2723	1	16
Labor force	57,372	3.8762	1.9813	1	7
Region	57,337	2.6462	.9650	1	5
Wave	126,699	8	4.0311	3	13

Table 2: proportion of the people who died separated for the different racial groups

	Ethnicity		
	White/Caucasian	Black/African American	Other
Percentage of people who died	0.0615	0.0586	0.0286

Methodology

The analysis that is going to be used in this paper is the regression analysis. More specifically a multiple linear regression model with robust standard errors. A regression investigates if there is a correlation between the dependent variable and the independent variable of interest, conditional on the control variables included. Furthermore, this research uses heteroskedasticity robust standard errors.

The aim of this research is to find out if there is an association between the two-year mortality and ethnicity for the years 1996, 2006 and 2016. In this case the dependent variable is the two-year mortality, and the independent variable is ethnicity. As usual in regression models for health and mortality, all regressions will control for age, age^2 and gender. This research will also do that, and this will be the baseline model. Age and gender are strong demographic control variables and are therefore useful to start the model with. Age^2 is used so that it allows to better predict the effect of age, because age can have a non-linear relationship with the two-year mortality. The formula of the baseline model will then be: two-year mortality = $\alpha + \beta_1$*ethnicity + β_2*age + β_3*age^2 + β_4*gender + ε. This model will give the differences in the two-year mortality for ethnicity, controlling for age and gender. Going forward the models are estimated sequentially to investigate to what extent the differences in the baseline are explained by the additional sets of control variables. The control variables used in the regression are education, total household income, number of people living in the household, employment status and region described in the data section above. The second model is the model where the variable education is added. The formula for this model is: two-year mortality = $\alpha + \beta_1$*ethnicity + β_2*age + β_3*age^2 + β_4*gender + β_5*education + ε. People who complete more years of education tend to live longer. Meara, Richards and Cutler (2008) discovered that between 1990 and 2000 life expectancy grew for high-education groups with 1.6 years but remained unchanged for the low-education group. Also the proportion of people who graduate from college is declining, especially among minority students (Strage 1999), providing evidence for racial differences in education. This is why education may explain racial differences in mortality. The third model adds the variables total household income and number of people living in the household to the regression, with the formula for this model: two-year mortality = $\alpha + \beta_1$*ethnicity + β_2*age + β_3*age^2 + β_4*gender + β_5*education + β_6*total household income + β_7*Number of people living in the household + ε. People whose income is higher generally live longer. Chetty et al

(2016) found out that the gap in life expectancy for the richest 1% and poorest 1% in the US is 14.6 years for men and 10.1 for women. There is not just an association between differences in income and life expectancy but also between differences in income and ethnicity. Akee, Jones & Porter (2019) found that in the US there is a rigid income structure, with mainly Whites and Asians positioned at the top and Blacks, American Indians, and Hispanics confined to the bottom. Number of people living is the household controls for how the money is divided in a household. Larger households need to spread their income over more people than smaller households. High total household income will have a different effect for big households than for small households. The fourth model adds the variable employment status to the regression. The formula for this model is: two-year mortality = α + β_1*ethnicity + β_2*age + β_3*age^2 + β_4*gender + β_5*education + β_6*total household income + β_7*Number of people living in the household + β_8*employment status + ε. Employed individuals tend to live longer than unemployed individuals. White men in the labor force live on average about twelve more years than those not in the labor force and white women on average nine more years (Rogot, Sorlie & Johnson 1992). It is not just that employed individuals live longer. Minorities do less well than equally qualified non minorities on such employment outcomes as representation in higher-level occupations, wages, returns on investment in educational credentials, and rates of job dismissal (Bendick Jr, Jackson & Reinoso 1994). This paper also wants to investigate if different regions in the US differ in the two-year mortality. In order to do so the variable region is added to the regression. The fifth and final model than looks as follows: two-year mortality = α + β_1*ethnicity + β_2*age + β_3*age^2 + β_4*gender + β_5*education + β_6*total household income + β_7*Number of people living in the household + β_8*employment status + β_9*region + ε. The software package that is going to be used to run the analysis is Stata.

Results

The findings gathered in this research are displayed below and discussed about in this section. As explained above, the model is built up by adding control variables step-by-step. For this reason, each model will be reviewed separately and at the end compared to one another.

The results of the baseline model can be found in Table 3. The two-year mortality is significantly different for Black/African American individuals compared to White/Caucasian individuals across all years. Other is never significantly different compared to White/Caucasian. This indicates that, for example in 1996, a person whose ethnicity is Black/African American has a 1.80 percent point higher two-year mortality compared to someone whose ethnicity is White/Caucasian. Age and age^2 both control for age in this regression. They are both significant in all years. Here age has a negative coefficient and age^2 has a positive coefficient, which means that as people get older the effect of age increases. Gender is the last variable in the regression. Just like age, gender is significant in all years. Women in 1996 are 2.35 percent point less likely than men to die in the next two years. Looking at the different years the coefficient of Black/African American decreases over the years. Additionally all years have the same significant variables and the values of the coefficients are mostly the same.

Table 3: linear regression for the two-year mortality as a function of ethnicity for the baseline model

	Wave 3 (1996)	**Wave 8 (2006)**	**Wave 13 (2016)**
Ethnicity (ref. category White/Caucasian)			
- Black/African American	.0180**	.0159**	.0114***
	(.0054)	(.0052)	(.0035)
- Other	.0024	-.0033	-.0000
	(.0091)	(.0063)	(.0035)
Age	-.0245***	-.0309***	-.0251***
	(.0020)	(.0019)	(.0017)
Age^2	.0002***	.0003***	.0003***
	(.0000)	(.0000)	(.0000)
Gender (ref. category male)			
- female	-.0235***	-.0309***	-.0167***
	(.0037)	(.0019)	(.0030)

Constant	.6920	.9108	.7282
R^2	.0899	.0957	.0900
N	17,975	18,467	20,835

Note. Standard errors are in parentheses; *p<0.05, **p<0.01, ***p<0.001

In the second model, the variable education is added. The results are visible in Table 4. Ethnicity is significantly different in 1996 and 2016 and insignificantly different in 2006 for Black/African American and insignificantly different across all years for Other compared to White/Caucasian. This implies that for 1996 and 2006 a respondent who is Black/African American has a higher two-year mortality than White/Caucasian. By including education, the significance of ethnicity drops compared to model 1 and the values of the coefficients are lower. Education shows a trend that the more one completes advanced education the lower the two-year mortality. All years show the same trend. Only GED graduate is insignificant, all the rest is significant. This could explain why the coefficients and the significance are lower for Black/African American, since part of the difference is now captured by education. Again the coefficients of Black/African American are getting smaller over time and 1996 and 2016 are similar to 2006.

Table 4: linear regression for the two-year mortality as a function of ethnicity, with control variable education

	Wave 3 (1996)	**Wave 8 (2006)**	**Wave 13 (2016)**
Ethnicity (ref. category White/Caucasian)			
- Black/African American	.0141*	.0097	.0084*
	(.0055)	(.0052)	(.0035)
- Other	-.0007	-.0111	-.0047
	(.0092)	(.0065)	(.0036)
Age	-.0245***	-.0311***	-.0250***
	(.0020)	(.0019)	(.0017)
Age^2	.0002***	.0002***	.0002***
	(.0000)	(.0000)	(.0000)
Gender (ref. category male)			
- female	-.0243***	-.0174***	-.0172***
	(.0038)	(.0036)	(.0030)
Education (ref. category lt. high school)			
- ged graduate	-.0071	-.0155	-.0055
	(.0090)	(.0093)	(.0079)

-	high school graduate	-.0102* (.0049)	-.0260*** (.0056)	-.0208*** (.0052)
-	some college	-.0171** (.0054)	-.0320*** (.0057)	-.0150** (.0051)
-	college graduate	-.0246*** (.0054)	-.0346*** (.0057)	-.0307*** (.0050)
Constant		.7090	.9499	.7454
R^2		.0909	.0983	.0922
N		17,974	18,463	20,830

Note. Standard errors are in parentheses; *p<0.05, **p<0.01, ***p<0.001

The third model adds total household income and number of people living in the household as control variables into the regression. Table 5 holds the results for model three. Ethnicity is significantly different for Black/African American in 1996 and 2016 compared to White/Caucasian, whereas Other is insignificantly different across all years compared to White/Caucasian. Meaning that being Black/African American gives you a higher two-year mortality. Total household income is significant in 1996 and 2016 and insignificant in 2006. For example, 2016 presents that for every dollar a household income earns the two-year mortality decreases with $1.94 \bullet 10^{-8}$ percent point. Number of people living in the household is insignificant in 1996 and 2006 and is significant in 2016. The effect of number of people living in the household is different over time. In 1996 and 2006 it is positive and in 2016 negative. In 2016 for each extra person living in the household, the two-year mortality decreases with 0.30 percent point. Compared to the second model there is not much of a change in the third model. The coefficient of Black/African American is significant in the same years. Adding total household income and number of people living in the household does not change the model much. It appears that total household income and number of people living in the household have no to little effect on the two-year mortality for different racial groups. Models 2 and 3 look rather the same. Most variables have the same coefficients in both models and 1996 and 2016 are similar compared to 2006. The coefficient of Black/African American decreases with each year, which the previous models also show.

Table 5: linear regression for the two-year mortality as a function of ethnicity, with control variables education, total household income and number of people living in the household

	Wave 3 (1996)	**Wave 8 (2006)**	**Wave 13 (2016)**
Ethnicity (ref. category White/Caucasian)			
- Black/African American	.0128*	.0093	.0078*
	(.0055)	(.0053)	(.0035)
- Other	-.0023	-.0120	-.0035
	(.0093)	(.0065)	(.0037)
Age	-.0245***	-.0310***	-.0253***
	(.0020)	(.0019)	(.0017)
Age^2	.0002***	.0003***	.0002***
	(.0000)	(.0000)	(.0000)
Gender (ref. category male)			
- female	-.0244***	-.0171***	-.0179***
	(.0038)	(.0036)	(.0030)
Education (ref. category lt. high school)			
- ged graduate	-.0063	-.0150	-.0066
	(.0090)	(.0093)	(.0079)
- high school graduate	-.0091	-.0254***	-.02140***
	(.0049)	(.0056)	(.0053)
- some college	-.0152**	-.0312553***	-.0155027**
	(.0054844)	(.0057364)	(.0051872)
- college graduate	-.0210***	-.0336***	-.0301***
	(.0057)	(.0058)	(.0052)
Total household income	$-5.06 \bullet 10^{-8}$**	$-4.49 \bullet 10^{-9}$	$-1.94 \bullet 10^{-8}$**
	$(2.09 \bullet 10^{-8})$	$(2.52 \bullet 10^{-9})$	$(6.83 \bullet 10^{-9})$
Number of people living in the household	.0020	.0021	-.0030**
	(.0014)	(.0017)	(.0010)
Constant	.7037	.9385	.7680
R^2	.0912	.0984	.0927
N	17,974	18,463	20,830

Note. Standard errors are in parentheses; *p<0.05, **p<0.01, ***p<0.001

Employment status is added to the regression in the fourth model. See Table 6 for the findings of this model. The coefficients are insignificant for Black/African American across all years compared to White/Caucasian. This is the same for Other. Employment status in 1996 and 2006 are mostly similar. Only retired and disabled are significantly different, meaning that being retired or disabled increases the two-year mortality with 2.27 percent points and

7.01 percent points respectively compared to working full time in 2006. In 2016 the coefficients of partly retired, retired, disabled and not in the labor force are all significant. Working part time decreases the two-year mortality whereas the rest increases the two-year mortality compared to working full time. By adding employment status the difference between Black/African American compared to White/Caucasian is no longer significant. This is in contrast with the previous models. The inclusion of employment status possibly captures part of the effect from the two-year mortality for different racial groups, therefore explaining part of the racial differences. Furthermore, 2016 explains the most, since it has clearly the most significant variables. The years 1996 and 2016 still look more alike than 2006. For the first time there is no clear trend over time for the coefficient for Black/African American.

Table 6: linear regression for the two-year mortality as a function of ethnicity, with control variables education, total household income, number of people living in the household and employment status

	Wave 3 (1996)	**Wave 8 (2006)**	**Wave 13 (2016)**
Ethnicity (ref. category White/Caucasian)			
- black/African American	.0102 (.0055)	.0063 (.0052)	.0067 (.0036)
- other	-.0044 (.0093)	-.0132 (.0065)	-.0040 (.0037)
Age	-.0257*** (.0020)	-.0312*** (.0019)	-.0251*** (.0016)
Age^2	.0002*** (.0000)	.0003*** (.0000)	.0002*** (.0000)
Gender (ref. category male)			
- female	-.0254*** (.0040)	-.0170*** (.0037)	-.0200*** (.0031)
Education (ref. category lt. high school)			
- ged graduate	-.0045 (.0090)	-.0147 (.0093)	-.0049 (.0079)
- high school graduate	-.0059 (.0050)	-.0226*** (.0057)	-.0184** (.0053)
- some college	-.0117* (.0055)	-.0274*** (.0058)	-.0115* (.0052)

- college graduate	-.0165** (.0057)	-.0278*** (.0059)	-.0240*** (.0052)
Total household income	$-2.71 \bullet 10^{-8}$ $(2.00 \bullet 10^{-8})$	$-2.44 \bullet 10^{-9}$ $(1.81 \bullet 10^{-9})$	$-1.27 \bullet 10^{-8}$* $(5.60 \bullet 10^{-9})$
Number of people living in the household	.0020 (.0014)	.0026 (.0017)	-.0029** (.0010)
Employment status (ref. category works ft.)			
- works pt	-.0014 (.0038)	-.0001 (.0042)	.0045 (.0032)
- unemployed	.0065 (.0098)	.0228 (.0131)	.0077 (.0065)
- partly retired	-.0070 (.0055)	-.0071 (.0048)	-.0140** (.0042)
- retired	.0227*** (.0041)	.0189*** (.0038)	.0176*** (.0035)
- disabled	.0701*** (.0118)	.0945*** (.0147)	.0430*** (.0088176)
- not in lbrf	.0111 (.0061)	.0009 (.0065)	.0207** (.0075)
Constant	.7409	.9399	.7579
R^2	.0945	.1030	.0952
N	17,974	18,463	20,830

Note. Standard errors are in parentheses; *p<0.05, **p<0.01, ***p<0.001

In Table 7 the fifth and final model is shown. This is the regression with all the control variables included. The two-year mortality for Black/African American compared to White/Caucasian is insignificantly different for all years. For Other the difference is insignificantly compared to White/Caucasian for 1996 and 2016 and significantly for 2006. The latest addition is the variable region. This variable displays if different regions in the US have different two-year mortality. The variable region is insignificant across all years for all categories. This reveals that there are no significant differences in the two-year mortality between regions. The fifth model and the fourth model are the same apart from the category Other from the variable ethnicity in 2006, which switches to significant. Subsequently the fifth and fourth model are the same for all years. From this it can be concluded that region has no effect on the two-year mortality.

Table 7: linear regression for the two-year mortality as a function of ethnicity, with control all control variables

	Wave 3 (1996)	**Wave 8 (2006)**	**Wave 13 (2016)**
Ethnicity (ref. category White/Caucasian			
- Black/African American	.0093 (.0056)	.0057 (.0053)	.0060 (.0037)
- Other	-.0040 (.0093)	-.0132* (.0067)	-.0025 (.0037)
Age	-.0256*** (.0020)	-.0314*** (.0020)	-.0251*** (.0017)
Age^2	.0002*** (.0000)	.0003*** (.0000)	.0002*** (.0000)
Gender (ref. category male)			
- female	-.0254*** (.0040)	-.0170*** (.0037)	-.0200*** (.0031)
Education (ref. category lt. high school)			
- ged	-.0046 (.0090)	-.0148 (.0093)	-.0053 (.0080)
- high school graduate	-.0058 (.0050)	-.0222*** (.0057)	-.0190*** (.0054)
- some college	-.0114* (.0056)	-.0272*** (.0058)	-.0115* (.0053)
- college graduate	-.0163** (.0057)	-.0275*** (.0059)	-.0239*** (.0052)
Total household income	$-2.64 \bullet 10^{-8}$ $(2.00 \bullet 10^{-8})$	$-2.44 \bullet 10^{-9}$ $(1.76 \bullet 10^{-9})$	$-1.23 \bullet 10^{-8}$* $(5.51 \bullet 10^{-9})$
Number of people living in the household	.0020 (.0014)	.0026 (.0017)	-.0028** (.0010)
Employment status (ref. category work ft.)			
- works pt	-.0013 (.0038)	.0004 (.0042)	.0047 (.0032)
- unemployed	.0068 (.0098)	.0227 (.0132)	.0080 (.0065)
- partly retired	-.0067 (.0055)	-.0072 (.0048)	-.0143** (.0042)
- retired	.0230*** (.0041)	.0189*** (.0038)	.0174*** (.0035)
- disabled	.0690*** (.0118)	.0942*** (.0147)	.0432*** (.0088)

- not in lbrf	.0112	.0005	.0210**
	(.0061)	(.0065)	(.0075)
Region (ref. category northeast)			
- Midwest	-.0018	.0061	.0044
	(.0057)	(.0056)	(.0053)
- South	.0010	.0090	.0008
	(.0052)	(.0052)	(.0045)
- West	-.0041	.0061	-.0058
	(.0061)	(.0059)	(.0048)
- other	-.0182	.0172	.0105
	(.0103)	(.0399)	(.0411)
Constant	.7378	.9394	.7568
R^2	.0942	.1032	.0955
N	17,972	18,450	20,811

Note. Standard errors are in parentheses; *p<0.05, **p<0.01, ***p<0.001

Conclusion

This research investigates the two-year mortality for different racial groups. By using multiple regressions across multiple years, an attempt has been made to find out if there are any significant differences.

Comparing the different ethnicity categories only Black/African American is significantly different from White/Caucasian. Mainly for the years 1996 and 2016. 2006 only had two significant categories, namely in the baseline model Black/African American and in the final model Other. There is evidence that Black/African Americans have a higher two-year mortality compared to White/Caucasian in the baseline model and the models controlling for education, total household income, number of people living in the household. This evidence disappears when employment status is included in the regression. The inclusion of employment status explains part of the racial difference in the two-year mortality. Age and age^2 are significant in all models and across all years. They show that as people get older the effect of age on the two-year mortality increases, which makes sense since more people die of old age than of young age. Gender is also significant in all models and across all years. Females have a lower two-year mortality compared to men. According to Seifarth, Mcgowan & Milne (2012), women exhibit greater life expectancy than men in almost every country. Their findings also hold for this research. Education gives a clear and significant trend for all years in the models. It is clear that the more education is completed the lower the two-year mortality. This was also discovered by Meara, Richards & Cutler (2008). Total household income and number of people living in the household are mostly insignificant, apart from 2016 for the models controlling for total household income, number of people living in the household, employment status and region. Total household income could be insignificant because of collinearity with education. Woodhall (1987) states that it is a well-established fact that educated workers earn higher wages or salaries than those who are illiterate or those who have completed less education or have lower educational qualifications. However, in 2016 it is significant giving evidence that higher total household income decreases the two-year mortality. In their study, Backlund, Sorlie & Johnson (1996) showed that the income-mortality gradient is much smaller at high income levels than at low-income levels in the working age and elderly population for men and women both before and after adjustments for other socioeconomic variables. This is in line with the findings in 2016. Employment status only explains the differences between working full time versus retired

and working full time versus disabled. Both retired and disabled have a higher two-year mortality. For retired the explanation is obvious. The dataset contains mostly people of old age, 50 years and older. Retirement usually comes when older and older people have a higher chance of dying. Hence the higher two-year mortality for retired. It is not specified what kind of disabilities the variable disabled include. But an explanation for the fact that the category disabled showed a higher two-year mortality could be mobility disability. Gill et al (2006) conclude that old age is associated with greater likelihood of transition to stages of greater disability and lower likelihood of regaining independent mobility. Since the dataset contains mostly people of old age, this could explain why disabled have a higher two-year mortality. Region is never significant across all years. There are no differences between regions and the two-year mortality.

Looking back at the first sub questions: Are there racial differences in the two-year mortality? The answer to that question is positive. There is evidence that different ethnicities have different two-year mortality. Mainly in the baseline model there is significant evidence that Black/African American have a higher two-year mortality compared to White/Caucasian. This evidence is furthermore sustained in the models controlling for education, total household income and number of people living in the household.

For the second sub question: Do the differences change over this period of time? This paper finds evidence that the difference in the two-year mortality for Black/African American compared to White/Caucasian declines over time. This evidence is found in the baseline model and the models controlling for education, total household income and number of people living in the household. This is in line with previous literature. On the one hand with Fuchs (2016). He discovered that in the beginning of the 21st century Black life expectancy at birth had double the increase compared to White life expectancy at birth. The fact that life expectancy at birth increases more for Blacks could account for the decrease in the two-year mortality compared to Whites. On the other hand with Case & Deaton (2015) they found that in the same time period as Fuchs the all-cause mortality rates for Whites increased, whereas all other racial groups saw their all-cause mortality rates decrease. The decrease in the difference in the two-year mortality between Black/African American and White/Caucasian could be because of the increase in all-cause mortality for Whites and the declining all-cause mortality for Blacks.

As for the third sub question: Are the racial differences explained by socio-economic factors? The differences in the two-year mortality are partly explained by socio-economic factors. It is clear that education and employment status explain part of the difference, whereas region has no effect and income has little to no effect on the two-year mortality and therefore do not explain the difference. To come back to the research question “Are there racial differences in the two-year mortality in the US for the years 1996, 2006 and 2016?”. There is evidence that there could be a racial inequality in the two-year mortality. In the baseline model and the models controlling for education, total household income and number of people living in the household Black/African American have a higher two-year mortality and it declines over time. Nonetheless controlling additionally for employment status and region there is no evidence for racial inequalities.

Discussion

The discussion reviews the limitations of this research and suggestions for further research. Also the internal validity and the external validity will be discussed. One of the limitations of this research is that is does not look separately at the racial group Hispanic/Latino. According to the US Census the Hispanic/Latino group is the second largest racial group in the US. For further research it would be interesting to see how this racial group compares to White/Caucasian and Black/African American. Following Case & Deaton (2015) I expect the same results for Hispanic/Latino compared to White/Caucasian. Expected is that Hispanic/Latino will have a higher two-year mortality, but the differences will get smaller over time. Another limitation is that there is no explanation for the difference between 1996 and 2016 compared to 2006. It is clearly visible that 1996 and 2016 are similar and that 2006 behaves differently comparing Other to White/Caucasian. There must be an explanation why this has happened. An idea for further research could be to compare 2004 to 2006 or to compare 2006 to 2008 to see if those years are more identical or if in 2006 something else is responsible for this difference. Another idea for further research is to look at the possible differences for younger people. This paper looks at people over 50 years old. Yet Case & Deaton (2015) found that the difference in all-cause mortality for Whites increased for people in middle-age. It would be interesting to conduct a research on younger people and see if there are racial differences in their two-year mortality. The internal validity of this paper is sufficient. The HRS is a large and random sample of the general population in the US and in this research control variables are used to account for confounding variables. The external validity is strong as the HRS contains a great number of respondents from different regions, gender, ethnicity and so on. It gives a realistic representation regarding the general population of the US.

Bibliography

Akee, R., Jones, M. R., & Porter, S. R. (2019). Race Matters: Income Shares, Income Inequality, and Income Mobility for All U.S. Races. *Demography*, *56*(3), 999–1021. https://doi.org/10.1007/s13524-019-00773-7

Backlund, E., Sorlie, P. D., & Johnson, N. J. (1996). The shape of the relationship between income and mortality in the United States. *Annals of Epidemiology*, *6*(1), 12–20. https://doi.org/10.1016/1047-2797(95)00090-9

Bendick, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring Employment Discrimination through Controlled Experiments. *The Review of Black Political Economy*, *23*(1), 25–48. https://doi.org/10.1007/bf02895739

Case, A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences*, *112*(49), 15078–15083. https://doi.org/10.1073/pnas.1518393112

Case, A., & Deaton, A. (2017). Mortality and morbidity in the 21st century. *Brookings Papers on Economic Activity*, 397.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001–2014. *JAMA*, *315*(16), 1750. https://doi.org/10.1001/jama.2016.4226

Cutler, D., Deaton, A., & Lleras-muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives*, *20*(3), 97–120.

Fuchs, V. R. (2016). Black Gains in Life Expectancy. *JAMA*, *316*(18), 1869. https://doi.org/10.1001/jama.2016.14398

Gill, T. M., Allore, H. G., Hardy, S. E., & Guo, Z. (2006). The Dynamic Nature of Mobility Disability in Older Persons. *Journal of the American Geriatrics Society*, *54*(2), 248–254. https://doi.org/10.1111/j.1532-5415.2005.00586.x

Gulland, A. (2016). Global life expectancy increases by five years. *BMJ*, i2883. https://doi.org/10.1136/bmj.i2883

Mackenbach, J. P., Kunst, A. E., Looman, C. W., Habbema, J. D., & van der Maas, P. J. (1988). Regional differences in mortality from conditions amenable to medical intervention in The Netherlands: a comparison of four time periods. *Journal of Epidemiology & Community Health*, *42*(4), 325–332. https://doi.org/10.1136/jech.42.4.325

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000. *Health Affairs*, *27*(2), 350–360. https://doi.org/10.1377/hlthaff.27.2.350

Muyle, T. P., Park, M. J., Nelson, C. D., Adams, S. H., Irwin Jr, C. E., & Brindis, C. D. (2009). Trends in adolescent and young adult health in the United States. *Journal of Adolescent Health*, *45*(1), 8–24.

Rogot, E., Sorlie, P. D., & Johnson, N. J. (1992). Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public Health Reports*, *107*(4), 457–461.

Seifarth, J. E., McGowan, C. L., & Milne, K. J. (2012). Sex and Life Expectancy. *Gender Medicine*, *9*(6), 390–401. https://doi.org/10.1016/j.genm.2012.10.001

Strage, A. A. (1999). Social and academic integration and college success: Similarities and differences as a function of ethnicity and family educational background. *College Student Journal*, *33*(2), 198.

U.S. Census Bureau QuickFacts: United States. (2019). Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/US/PST045219

Wang, H., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., & Abd-allah, F. (2017). Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1084–1150.

Woodhall, M. (1987). Earnings and education. *Economics of Education*, 209–217. https://doi.org/10.1016/B978-0-08-033379-3.50041-3

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

World Population Prospects - Population Division - United Nations. (2019). United Nations. https://population.un.org/wpp/DataQuery/

Appendix

Stata do-file

```
set maxvar 100000
use "/Users/bobmelzer/Documents/randhrs1992_2018v1_STATA/randhrs1992_2018v1.dta"

* keep only the variables I am interested in
keep hhidpn r1agey_e r2agey_e r3agey_e r4agey_e r5agey_e r6agey_e r7agey_e r8agey_e
r9agey_e r10agey_e r11agey_e r12agey_e r13agey_e r14agey_e ragender raracem r1cenreg
r2cenreg r3cenreg r4cenreg r5cenreg r6cenreg r7cenreg r8cenreg r9cenreg r10cenreg
r11cenreg r12cenreg r13cenreg r14cenreg raedyrs h1itot h2itot h3itot h4itot h5itot h6itot
h7itot h8itot h9itot h10itot h11itot h12itot h13itot h14itot r1lbrf r2lbrf r3lbrf r4lbrf r5lbrf
r6lbrf r7lbrf r8lbrf r9lbrf r10lbrf r11lbrf r12lbrf r13lbrf r14lbrf r1iwstat r2iwstat r3iwstat
r4iwstat r5iwstat r6iwstat r7iwstat r8iwstat r9iwstat r10iwstat r11iwstat r12iwstat r13iwstat
r14iwstat h1hhres h2hhres h3hhres h4hhres h5hhres h6hhres h7hhres h8hhres h9hhres
h10hhres h11hhres h12hhres h13hhres h14hhres raeduc

* reform from wide format to long format
reshape long r@agey_e r@cenreg h@itot r@lbrf r@iwstat h@hhres, i(hhidpn) j(wave)

* generating new variable for people that died in each wave
gen died = 1 if (riwstat==5)
replace died = 0 if (riwstat==1)|(riwstat==4)

* make from the panel data a time series so it can be used to include leads
tsset hhidpn wave

* generating new variable for people that died in next wave
gen died_nextw = f.died

* regressions for first model
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender if wave==3, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender if wave==8, robust
```

```
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender if wave==13, robust

* regressions for second model
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc if wave==3, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc if wave==8, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc if wave==13, robust

* regressions for third model
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres if
wave==3, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres if
wave==8, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres if
wave==13, robust

* regressions for fourth model
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf if
wave==3, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf if
wave==8, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf if
wave==13, robust

* final model (region is added last)
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf
i.rcenreg if wave==3, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf
i.rcenreg if wave==8, robust
regress died_nextw i.raracem c.ragey_e##c.ragey_e i.ragender i.raeduc hitot hhhres i.rlbrf
i.rcenreg if wave==13, robust
```

```
* descriptive statistics
drop if wave==1
drop if wave==2
drop if wave==4
drop if wave==5
drop if wave==6
drop if wave==7
drop if wave==9
drop if wave==10
drop if wave==11
drop if wave==12
drop if wave==14
sum died_nextw raracem ragey_e ragender raeduc hitot hhhres rlbrf rcenreg wave
tab died_nextw raracem
```