Björn Poerschke

b.poerschke@student.rug.nl

Student Nr.: 3768392

Address: Otto-Fischer-Straße 6a, 50674 Köln, Germany

Master-Thesis *M. Sc. Population Studies*

Life Expectancy Inequalities between Natives and Migrants in the Netherlands – Effects of Mortality Differentials and Selection

Supervisor: Adrien Remund

University of Groningen

Faculty of Spatial Sciences

Population Research Center (PRC)

Cologne, 31st August 2019

Contents

List of Tables

List of Figures

List of Abbreviations

CBS – Centraal Bureau voor de Statistiek (Statistics Netherlands)

CVD – Cardiovascular Disease

SES – Socioeconomic Status

SRH – Self-reported Health

Abstract

The thesis aims at analyzing current differences in mortality between Dutch natives and three migrant groups and the identification of social processes which underlie these differences. The overall theoretical framework of the "healthy-migrant-paradox" is conceptualized for the Dutch case. Also, the salmon-bias, one of the mechanisms contributing to the paradox, is dealt with in particular. Mortality inequalities are analyzed with different life table techniques as well as survival analysis. The salmon bias analysis is carried out with a Cox hazard model. Additional attention is paid to the different datasets and methodological approaches used in the different components of the analysis. The results suggest that the healthy migrant effect is viable mostly for Moroccans. Additionally, Turkish individuals show mortality advantages throughout adult age-groups, but not in total life expectancy and older ages. The Surinamese population is found to be almost uniformly disadvantaged. There is also a large data effect: when changing the population from residentially restricted to non-restricted, Dutch natives enjoy the most favorable mortality patterns and highest life expectancy. The results of the salmon bias analysis suggest that moving and health are related for Turkish immigrants; however, not in the direction that the salmon bias suggests. Conclusively, migrant health in the Netherlands is less paradoxical as expected and a considerable bias due to unhealthy re-migration seems unlikely.

Keywords: Healthy Migrant Paradox, Salmon Bias, Mortality, Migration, Population Health, Life Table, Survival Analysis, Netherland

1. Introduction

Western European societies are becoming more and more multicultural, with some regions and cities where the native-born population is in the process of losing the absolute majority status such as Amsterdam (European Research Council, 2019). The same societies are becoming older on average, overall life expectancy has increased in the Netherlands, even throughout the more recent past (Mackenbach et al., 2011). The monitoring of life expectancy, health and the respective differences between population subgroups are therefore an important building block in understanding and developing societies. Monitoring life expectancy developments, whether this might be of a specific group or the whole population, is important to identify societal problems as well as progress. Should it be the case that a specific group is ahead of another or lagging behind, then scientific findings could hint on social issues that should be tackled. But it could also be implied that the social situation is improving. Moreover, identifying factors that de- or increase mortality are vital in successfully ensuring population health. Migrant groups are a special case in this strain of research.

> "Refugee and migrant health is a highly complex topic and research findings often cannot be generalized to wider refugee and migrant populations in a country, in a region or globally. The effects of the migratory process, social determinants of health and the risks and exposures in the origin, transit and destination environments interact with biological and social factors to create different health outcomes." (WHO, 2018, p. 11)

Life expectancy measures increasingly play a role in retirement policymaking all over Europe. In the Dutch case, statistically calculated mortality patterns are used as an indicator for the universal retirement age. When citizens are allowed to start receiving pensions is going to be based on life expectancy from 2022 on (Statistics Netherlands (CBS), 2014). Demographic measurements are therefore not only the reflection of accumulated individual experience of mortality, they are also used with real consequences for the individual. To evaluate policies like this, continuous research is essential. Additionally, finding out about mechanisms that influence the values is an important part of evaluating the societal value and the implications for the individual of such policies. The case of the Netherlands is therefore especially interesting. The country has historically been a place with ethnically multifaceted populations. Similar to the rest of the western world, immigrants exhibit systematically higher proportions of economically deprived individuals (Statistics Netherlands (CBS), 2016).

A lot of past research has pointed out that mortality of migrant groups is not following the established links of socioeconomic positioning, health and mortality. Either health or only mortality of migrants proved to be better than those of natives, which always had advantages regarding socioeconomic resources. Subsequently, scholars have called these contradicting findings the "healthy migrant paradox". Thus, the factors underlying this finding has been studied throughout a lot of different cases and for over thirty years (s. Markides and Coreil, 1986). However, it remains up for debate what the main drivers of mortality differentials are. Also, it is important to examine every case individually, findings from the past in other European countries cannot be generalized for the Dutch case which has a specific history of migration and legislative framework. Until this day even findings within the EU are differing from each other and so are the interpretations. It still remains relevant to look into the situation itself and take further steps in finding out about the reasons for the paradox.

The aim of this thesis is therefore to shed light on current developments in mortality differentials between migrants and natives in the Netherlands. Ultimately, its aim is to identify a possible healthy migrant paradox and look into the sociodemographic mechanisms that are the determinants of variation in mortality. This is done to broaden the knowledge about the relatively understudied field of native-migrant mortality differentials in the Netherlands, with focusing on three specific groups of migrants. This is possible due to the detailed microdata which underlies the analyses carried out in this paper. This data was made available by the 'Centraal Bureau voor de Statistiek (CBS)' or 'Statistics Netherlands' due to a Master-Thesis-Internship (from here on referred to as *CBS*). Whereas the aggregated data can be accessed by anyone, the micro-data is not publicly available. It can only be processed with permission and within the CBS infrastructure. With micro-data from multiple state registers such as residential, migration and demographic basis, a more nuanced view, as compared to classic aggregated period data, on mortality differences is possible. Lastly, data from the country's health services is combined with the demographic registers and used to conduct in-depth analyses on out-migration-patterns to review the salmon bias hypothesis. Latter is one of the expected mechanisms behind the healthy migrant paradox. It refers to out-migration in case of worsening health or higher ages.

In practice this will be done with analyses of Turkish, Moroccan and Surinamese populations and their mortality patterns. Native Dutch individuals will work as a reference point, as a kind of 'national average' from which migrant groups are thought to deviate from. More precisely, period life tables for multiple timeframes, age-specific mortality rates and cause-specific mortality will be examined in the first part. Inductive analyses will be used to find out whether the differences in life expectancy can be traced back to composition or selection effects. To operationalize this, a survival model will be run on the basis of individual register data. First, to test mortality within a data-framework that surpasses single year cross-sections and residential status. Second, to examine selection processes on the basis of health and emigration from the Netherlands. This innovational research design and the variety of data are unique in the Dutch case and offer the possibility to uncover differentials and better understand the determinants behind it. Moreover, the data allows to reach a new level of refinement regarding the analysis of life expectancy. In the past this has mostly been done on a macro level by lifetable or mortality ratio analyses. Especially selection effects where therefore hard to identify.

The thesis will start with a comprehensive recap of the theoretic background and major empiric findings on the topic. Next, the theoretical framework will then be translated to fit the specific case and population at hand. Both theory, literature and the specific circumstances of this case will be used to design an empirically analyzable concept. Factors and relationships will be operationalized in order to fit into the frame of statistical analysis. Respectively, the data sources will be described in depth. As mentioned, the model which this paper uses to empirically test the theoretical concept involves multiple data sources: samples from health surveys, pension data registers and population registers. Also, the overall operationalization of the concept and its consequences will be laid out. More precisely, the populations and the variables with their respective empiric measurements. This part will also include a discussion about overall data quality and the advantages and limitations of doing statistical analyses with the demographic registry data at hand. Then, the methods through which analysis is carried out are described. The main ones being life table analysis, the respective age- and cause-specific measurements and inductive history event models. Following up on that, the results of the statistical analyses are laid out in form of tables and visualizations and described in relation to the concept. Next, those results are interpreted in regard to the theoretical framework and used to give an answer to the research question. Further, the implications of the empiric findings are discussed and critically put into perspective. Lastly, the whole paper is recapitulated with focus on the goals, scientific value and general limitations of the study design. Which is then followed up by the substantial implications for further actions on a scientific and societal level.

2. Theoretic Background and Empirical Evidence

Socioeconomic status, health and mortality have been largely recognized for being interlinked, with research on socially-varying mortality dating as far back as the 1800's (Elo, 2009). Until today, life expectancy is found to be unequally distributed among different levels of education, income and overall socioeconomic status, in a variety of cases throughout western countries (e.g. Currie and Schwandt, 2016, Krokstad et al., 2002). Even though convergence is still not the case, most socioeconomic groups have improved in their mortality over the years, most prominently in life expectancy at birth, however at different levels (e.g. Currie and Schwandt, 2016, Bengtsson and van Poppel, 2011). The mechanisms behind this linkage seem to be multidimensional as the correlation of SES and health hold in a variety of different frameworks. This includes both cultural circumstances worldwide and a large array of different diseases. Lower SES always indicates a disadvantage in both health and mortality. Due to this long-standing stability, the relation of SES to health has been termed a 'fundamental cause'. The differentials in class related health have been traced back to underlying factors such as childhood effects, health care utilization, residential context and biology (Elo, 2009). Most importantly psychosocial strain connected to the SES has been proven to lead to physical and mental illness. For example, low reward jobs (low salaries, societal approval and without employee-friendly structure/climate) are also falling under this category of psychosocial health risks. Thus, predominantly affecting people with disadvantaged socioeconomic positions (Dragano & Wahrendorf, 2016). This in turn has effects on overall mortality rates, either directly or through wear and tear which leads to long term risks and shortens the lifespan.

These class-related differences in mortality, health and their persistence across contexts are important to note in this paper. In western European societies large parts of socioeconomically deprived groups in society are made up by migrants. Unemployment, lower education and lower income are all factors in which individuals born outside the host-countries are overrepresented (EUROSTAT, 2019a). Therefore, they are the ones that are thought to be facing these adversities in health and mortality. Additionally, psychosocial strain specifically targets migrant groups in terms of racist discrimination, something that has been measured to negatively impact both mental and physical health (e.g. Ikram et al., 2015, Pascoe and Richman, 2009, Williams, 1999). Also, cultural and language barriers might inhibit effective health-service utilization; a mechanism which led for example to lower rates of vaccination or use of medical specialists amongst migrants (e.g. Glaesmer et al., 2011, Lampert et al., 2005). All these stressors have the potential to either actively, through risk of injury or violence, or passively, via stress, to negatively affect one's health. These migrant-specific issues have been reported to increase health-risk and morbidity for migrants in western host-countries (Razum, Zeeb, Akgün, & Yilmaz, 1998).

Because of these social and health adversities but also because of growing societal importance of migrant groups, research on health differences between population subgroups has been a scholarly issue for a large array of scientific fields. Despite the implications of their specific socioeconomic stance in society and higher health risk as compared to non-migrants, epidemiologic and demographic research has not always delivered the intuitive results when it came to mortality. In the northern American case, even for health. For example, in the 1960's it was discovered, that Mexican-Americans had better psychological health than other socioeconomically deprived groups (Karno & Edgerton, 1969). Later, mortality patterns were discovered to show similar results. These findings replicated manifold and have been termed "Healthy Migrant Paradox". The term dates back to the late 1980's when a paper (Markides & Coreil, 1986) summarized findings that showed that the average health of Hispanics was similar to non-Hispanic whites on a number of physical health indicators (such as: infant-mortality, life expectancy, functional health and major causes of death). The paradoxicality of this phenomenon stems from the fact that Hispanic health was good, despite their deprivation in regard to socio-economic factors. The latter were rather similar to black Americans, which faced both disadvantages in health and socio-economic status, a finding in line with what was already known about the SES-health-relationship. Thus, Hispanic health was unexpectedly good, and the finding was termed "Hispanic Epidemiologic Paradox". In these early studies, the health of Hispanics was found to be similar or close to native health but still slightly disadvantaged (Markides & Coreil, 1986). In more recent examples this paradox 'grew' to an extent where studies reported mortality rates for Hispanics that were even lower than the ones of white Americans (e.g. Abraído-Lanza et al., 1999, Xu et al., 2018, Diaz et al., 2016). However, this phenomenon is not solely found in northern America. Research on the healthy migrant also has been conducted throughout Europe with similar results. In Belgium, Moroccans and Turkish have been found to enjoy a mortality advantage for almost all causes of death. Except diabetes, which seems to affect almost all non-western migrants highly (Reus-Pons, Vandenheede, Janssen, & Kibele, 2016). In the city of Amsterdam studies comparing different ethnic groups showed that native Dutch people had a mortality disadvantage compared to multiple migrant groups. The highest life expectancy values were measured for individuals with a Mediterranean background; that is, mostly individuals with an Arab or north African descent (Uitenbroek and Verhoeff, 2002, Uitenbroek, 2015). Overall, this advantage of Mediterranean migrant groups is continuously reported in case studies throughout Europe (e.g. Razum et al., 1998, Norredam et al., 2015, Gruer et al., 2016, Guillot et al., 2018).

The first approaches in explaining this finding ranged from genetic heritage, cultural practices such as early fertility and healthy lifestyles, social support through families and selective migration. These effects are thought to be further enhanced by the better developed health services in the host-countries. More precisely, health advantages are made possible by an improvement in health services, decreasing the high risk of infectious disease in developing origin-countries (Markides & Coreil, 1986). However, it was also already stated that migrant populations might face an increased risk for certain causes of death through acculturation in the host country, such as picking up white American dietary and alcohol consumption choices (Markides & Coreil, 1986). First, the “Healthy Migrant Effect”, is concerned with positive selection at the moment of immigration. Migration is thought to be a strenuous experience which brings about certain costs for the individual that migrates because abstract and physical hurdles have to be overcome. Depending on the context, migrants have to transcend geographic distance, legal boundaries, cultural and language barriers in order to settle successfully in another country. In more extreme contexts like refugee migration there are also manifold direct physical dangers. Thus, vulnerable individuals such as older or unhealthy ones, generally those groups with unfavorable mortality patterns, are discouraged to migrate and remain in their origin-country (Abraído-Lanza et al., 1999; Kohls, 2015; Razum et al., 1998). Also, migration is highly affected by economic disparages, with richer countries pulling individuals into the economy and unfavorable situations such as mass-unemployment push out individuals. Thus, younger individuals who face better chances in the host-countries labor-market are additionally motivated while older or unhealthy migrants face even higher costs. This also regards collective household decisions in which the more capable individuals are sent away in order to generate revenue for themselves and the stayers (Marmot, Adelstein, & Bulusu, 1984; Stark & Bloom, 1985). In some cases, these specific people are also even legally favored to move. For example, language courses offered to young and educated migrants and laws specifically designed to attracted qualified personnel from outside the EU. The most relevant example being young workers from the Mediterranean that were specifically recruited to counter shortages of manpower in booming western European countries of the 1950’s (Lafleur & Stanek, 2017). Thus, overall disadvantageous circumstances which set costs high or simply make it impossible for anyone to legally migrate, build the framework in which only a specific type of individual moves abroad in the first place. Combined with incentives for abled or educated persons this builds the first social mechanism behind the healthy migrant paradox.

The other side of this effect is the so called “*Salmon Bias*” or “*unhealthy re-migration effect*”, which refers to the return of migrants to their home country in case of declining health or severe illness (Razum, 2006). As the health situation worsens, individuals are increasingly motivated to return to their place of birth. This is due to social ties and emotional ties to the home-country as well as a lack of integration in the host-society which works as a constant stress-factor. If severe illness sets in, these push- and pull-factors lead individuals to leave the host country. Even in the absence of illness this effect is also viable for older age-groups. Thus, the vulnerable people leave and with them an increased risk of dying in their home country, whereas the healthy individuals stay in the host country. Therefore, the staying migrant group has selectively advantaged mortality rates as leavers are not in the picture anymore. This produces the said bias which eventually leads to the impression that migrants have advantages in mortality (Wallace, 2019). As death is likely to happen at ‘home’ the leavers take up the metaphorical journey of the salmon, a fish-species which also returns to the place of birth in order to die. Analytically, this bias stems from individuals with high mortality risks being removed from the exposure; thus, increasing the number of person years lived throughout each age interval. This is specifically viable for the older ages as higher mortality risks and worse health due to degenerative disease are increasing (Guillot et al., 2018). Both above described mechanisms lead to healthy individuals entering and staying in a western host-society while unhealthy ones are prone to re-migrate. Evidence on this phenomenon is mixed. Some studies find evidence for the mortality-remigration relationship (Palloni & Arias, 2004), others no evidence at all (Abraído-Lanza et al., 1999) or even the opposite effect (Norredam et al., 2015; Puschmann, Donrovich, & Matthijs, 2017). Again others find mixed results depending on the used health indicators (Diaz et al., 2016) or results suggesting that unhealthy remigration happens but at scales that are too small to affect macro mortality patterns and thus do not explain the mortality advantage (Turra & Elo, 2008; Wallace & Kulu, 2018). Whereas the healthy migrant paradox seems to be rather stable throughout different contexts and study designs, the salmon bias is much more debatable. European studies finding factual evidence on the mechanism are notably missing as compared to results from the US. However, the latter case has also been studies much more extensively.

Thirdly, above mentioned lifestyles are among the most prominent explanations for migrant mortality advantages. More precisely, cultural customs which includes healthier foods or cultural refusal of alcohol consumption. These have been found to be positively associated with better health and found to be more popular within migrant populations (Abraído-Lanza, Chao, & Flórez, 2005; Akresh, 2007). Also, smoking has been found to be lower for specific migrant groups. Both after they have entered the country and even before the move into the other country, migrants were the ones less likely to smoke as compared to their peers in both locations (Riosmena, Kuhn, & Jochem, 2017). Therefore, this regards the population of migrants during their stay in the host country and even selection before the move happens. Lastly, even genetic factors can play a role in migrant health. Hispanics have been argued to be naturally aging slower than their Caucasian counterparts (Horvath et al., 2016). In contrast, south Asian migrants seem to be genetically predisposed towards a higher risk in CVD and diabetes (Gupta, de Belder, & Hughes, 1995). As many Surinamese migrants are of south Asian descent, this phenomenon has also been found to be the case in the Netherlands (Statistics Netherlands (CBS), 2017b). An important distinction between the European and the American case also needs to be noted. That is, despite of a mortality advantage for migrants, European cases have shown morbidity and health disadvantages in the same groups. In the US, Hispanic migrants both dimensions are better. Thus, one could say that the situation is even “more paradoxical” in Europe. Lifestyle and health behaviors might therefore be less relevant.

In more recent times, throughout the 2000’s, data from Germany showed that throughout adulthood, from 20 to 60 years of age, migrants showed more favorable age-specific mortality risks. However, for ages above 60 the mortality patterns of migrants were clearly disadvantaged compared to German natives (Kohls, 2015). The healthy migrant effect applies to the majority of migrants in younger adult ages as they arrive in the host country due to mainly economic or educational reasons. The older age groups on the contrary, have arrived throughout the 1950's to 70's and mostly spent their productive ages in Germany, where they were affected by more strenuous jobs and lesser socioeconomic resources. Thus, their mortality patterns are disadvantaged (Kohls, 2015). Similar results of a decline of health outcomes over the life-course within the host country were reported especially in European cases. Healthy migrant patterns are mostly found in younger and adult age groups while with older ages health outcomes are worse than those of natives (Guillot et al., 2018; Loi & Hale, 2019; Norredam et al., 2015; Reus Pons, de Valk, & Janssen, 2018). A recent study examining data from the UK, US and France (Guillot et al., 2018) suggests that there is generally a u-shaped pattern of mortality inequality between migrants and natives. This refers to migrant mortality being higher or at the same level as the native one for very low and high age groups, whereas within the middle and adult ages migrants exhibit migrant mortality advantages. Next to the low SES life course, explanations for this phenomenon include acculturation (Abraído-Lanza et al., 2005). This refers to adaptation of cultural practices, such as health behaviors which reduces the initial advantage that migrants had by sticking to their original lifestyles when they arrived in the host country. Prominent examples are picking up higher alcohol and tobacco consumptions, and a shift to more industrially processed diets (Abraído-Lanza et al., 2005).

From the first studies on mortality advantages of Hispanics compared to white Americans, it was hypothesized that the paradoxically favorable situation was due to data artifacts, such as underreporting (Markides & Coreil, 1986). The data that was used to derive interpretations of advantageous health of immigrants came from health services, for example reporting the number of psycho-therapy patients (Karno & Edgerton, 1969). Thus, lower health service utilization might as well work as a mechanism behind the ethnic differences. Moreover, later studies discussed the various possibilities in which shortcomings of the data might be the leading factor behind the paradox and found that demographic registration might account for certain differentials (Uitenbroek & Verhoeff, 2002). Underreporting of migration moves or deaths abroad lead to a mismatch between exposures and incidences. This kind of error is mostly affecting cross-sectional designs and period data that rely on registers of residential status. In longitudinal settings e.g. for the Dutch case, longitudinal data has shown migrant mortality to be less advantaged compared to natives (Bos, Kunst, Keij-Deerenberg, Garssen, & Mackenbach, 2004). However, other longitudinal studies have still proven a significant migrant mortality advantage (Abraído-Lanza et al., 1999; Swerdlow, 1991). Longitudinal data is also not immune against underreporting, individuals still might be administratively lost, and deaths can go without notice. Overall, this topic is hard to grasp analytically, as non-registration already implies a lack of reliable data which is hard to counteract by study design. Empirical analyses such as the one in this paper have to assume that the data is reliable in reporting.

All these mechanisms are most viable for first-generation migrants. For the second-generation the country of birth is logically already the host country; thus, 'arrival' as a selectively healthy group is impossible. The opposite might even be the case, younger individuals might start out as unhealthier through socioeconomically deprived upbringing. Socialization, through for example the host-countries school system and peer groups, is also very different. However, some case-studies have shown mortality patterns of migrants that prevailed throughout generations (e.g. Razum et al., 1998).

It remains up for debate if all these dimensions apply for migrant groups under specific circumstances. Many European migrant populations have long standing ties to their host countries including family, friends and possibly even became official citizens. For example, in the Netherlands about 25 % of migrants have been nationalized (EUROSTAT, 2019b). It is therefore questionable whether for example the social network component of the salmon bias might fully apply. Many of the western European countries also might offer better health services. In the case of the Netherlands mandatory health insurance for every legal resident is granted, which would make it feasible to stay in case of illness rather than leave. Acculturation is also debatable, migrants have been reported to still have lower rates of tobacco and alcohol consumption as natives (Kohls, 2015). Moreover, the EU migration history is shaped to a considerable degree by chain-migration (Lafleur & Stanek, 2017; Zorlu & Hartog, 2001). The supposed healthy and abled first arriving individuals might have united with their relatives in the host country which implies that also the older and less healthy might have entered. Put shortly, the migrant populations might be less selective than assumed.

3. Conceptual Model and Hypotheses

Taking into account the empiric findings, theory and the fact that most migrant groups in the Netherlands are socioeconomically deprived (Statistics Netherlands (CBS), 2019b), the Netherlands are also an interesting case to research mortality differences between natives and migrants. Large-scale migration in the Netherlands started after 1945 when decolonization and recruitment of foreign labor came into effect. The former regards mostly Indonesian and Surinamese migrants while individuals who came to the Netherlands through the latter were mostly from the Mediterranean (Zorlu & Hartog, 2001). Today Surinamese, Moroccans, Turkish and Indonesians are the biggest non-EU migrant groups in the Netherlands (Statistics Netherlands (CBS), 2019a). However, many of the Indonesian migrants are from European or Dutch origin who fled Japanese invaders or lost incentives to stay after colonization was ended (Zorlu & Hartog, 2001). The registers do not distinguish between ethnicity, just between place of birth, so Indonesians hardly qualify for any analysis of the healthy migrant paradox. Dutch demographic statistics even count Indonesians as 'western migrants' (e.g. Statistics Netherlands (CBS), 2016). The same applies for many of the southern European countries. As many of the Mediterranean countries joined the EU, the legislative framework changed gravely and does not apply to the healthy migrant paradox as well as for Moroccans and Turkish. The latter groups still face high restrictions which makes it likely only abled and healthy persons take the cost to migrate. Moreover, chain-migration following labor migration has mostly been happening for non-EU populations, while the numbers of southern European migrants in the Netherlands declined over the years (Zorlu and Hartog, 2001). Many of the large migrant groups in the Netherlands stem from EU-countries such as Germany and Poland (Statistics Netherlands (CBS), 2019a). These groups play an important role for society as well but fit poorly into the healthy migrant paradigm for similar reasons as migrants from Mediterranean EU countries. Conclusively, the origin countries chosen for this paper are Tukey, Morocco and Surinam. These three groups also still face large disparities in socioeconomic status compared to Dutch natives. Regarding education, employment, income, societal participation and even self-reported health all of the respective migrant populations exhibit disadvantages (Statistics Netherlands (CBS), 2016).

Restrictive migration legislation has inhibited a continuous inflow of new migrants from these countries. The possibility to be nationalized as a Surinamese individual has already expired in 1980, the same applies for the active recruitment of labor from the Mediterranean which stopped during the oil crisis. Labor migration from outside the EU has been banned almost entirely. Thus, a large part of what is today commonly referred to as migrant population is already native-born to the Netherlands. Especially for the younger age-cohorts these individuals are the majority (Zorlu & Hartog, 2001). Thus, healthy migrants might have entered as the first-generation of labor migrants as they were the ones that could profit most of the situation. They were also the ones the recruitment policies ultimately aimed at. The subsequent years increased cost of migration due to higher legal hurdles and thus favor healthy migrants as well. For example, younger students who are granted visas as students and enter the labor market as highly skilled and positioned personnel, which are exempted from complete labor migration bans (Zorlu and Hartog, 2001). Thus, the overall setting for a selective group to enter the Netherlands is given, at least for those groups that are further away or legally restricted to enter.

Even tough, in the 1970's economic decline and unemployment hit migrants specifically hard, many from Turkey, Morocco or Surinam did not return to their home-country (Zorlu and Hartog, 2001). On the other hand, the migration patterns of Mediterranean migrants from EU countries such as Spain, Italy and Greece differ greatly. Many of them returned to their home-countries in large numbers; however, with better prospects of coming back to the Netherlands once again (Zorlu and Hartog, 2001). Thus, restrictive migration frameworks also have the ability to inhibit circular migration as it would 'shut the door' of re-entering the host country. For migrants from Turkey and Morocco this was the case. Thus, returning to the country of birth seems to be a grave decision which requires a fundamental change in one's situation, such as the worsening of health. A share of ca. one fifth to one third of migrants from the countries of interest returned to their place of origin before the beginning of the 21st century (Bos et al., 2004). This legal framework and problems of integration into the economy might then have laid out the ground for the salmon bias to become a reality, if the better-off individuals then stayed in the Netherlands, while the worse-off left as theory suggests.

It can be argued that emigration of unhealthy migrants is counterintuitive as the health care services in the Netherlands are well-developed and health care is universal. The rational choice of individuals in a state of bad health would then be to actually remain in the host country rather than return or leave in general. However, any member of a Dutch health insurer is also entitled to receive at least a part of the health care expenses while being registered abroad (HollandZorg, 2019). As any resident staying in the Netherlands for more than 4 months is obliged to be insured under a Dutch health care scheme (CAK, 2019), it is likely that migration in terms of the salmon bias is not contradicting further medical attention in the country of origin. This is underlined by large proportions of people of Turkish and Moroccan descent, that regularly make use of health care in their country of origin (Şekercan, Snijder, Peters, & Stronks, 2018). Returning in case of sickness, primarily to reconnect with one's roots and for social support is therefore not held back due to a lack of formal care.

Additionally, other social services such as the Dutch pension system also pays recipients outside the Netherlands. Especially in those countries where considerable amounts of migrants come from, including the top five origin-countries of migrants in the Netherlands. Pension entitlement does not stop if an individual returns (Sociale Verzekeringsbank (SVB), 2019). Also, state pension funds are acquired easily by everyone legally working in the Netherlands. Even if one might not be entitled to the full amount, smaller pensions might facilitate living in countries with lower cost of living and thus make it easier for people to emigrate. Thus, even those migrants who are not directly experiencing a decline in health but are old enough to receive pension might leave as well. In those groups, the salmon bias is expected to be highest as ageing, especially for groups above pension ages, goes hand in hand with an increase in morbidity (Guillot et al., 2018).

Following the situation laid out above, both healthy migrant and salmon bias seem to be theoretically viable for migrants in the Netherlands. First-generation Surinamese, Moroccans and Turkish reflect both socioeconomic deprivation as well as importance for the Netherlands in terms of integration into society, history and labor market importance. Their long history in the Netherlands have left large parts of this group in older age groups, thus with higher morbidity and mortality risks. Additionally, a large part of adult foreign-born individuals exists to compare them with natives and research the healthy migrant effect. The timing seems well suited to research the healthy migrant paradox and negative health selection effects such as the salmon bias. For the following analyses second generation migrants are not incorporated. The definition of origin makes a salmon bias less valid for them.

Conclusively, it is hypothesized that migrants will show mortality advantages as compared to natives. Second, the mortality differentials between the groups will be partly made up by sociodemographic composition. Lastly, it is expected that worse health will lead to higher probabilities of leaving and better health to stay in the host country. Moreover, the data at hand makes it possible to distinguish between mortality in cross sectional and longitudinal formats. The former being restricted to residential status and the latter offering follow-up despite possible migration. Thus, data effects can be identified. The age structure and history of the migrant groups makes it possible that similar results as in the case of other European studies mentioned above are found. That is, longer tenure in the host country reduces any theoretical initial health advantage and leaves current first-generation migrant populations rather disadvantaged in the older age-groups. Thus, health convergence or an ‘unhealthy migrant effect’ are also likely to be found.

4. Data and Analytical Concept

To empirically test above formulated hypotheses, multiple datasets will be utilized and analyzed within different methodological frameworks. The datasets, which will be described in more detail below, used in this paper are the aggregated public data of Statistics Netherlands, the base residential register of the Netherlands and the Dutch Health Monitor of 2012.

As a first step of identifying the expected mortality inequalities between migrants and Dutch natives, a life table analysis of multiple periods is conducted. The dataset itself is comprised of aggregated death and population counts of all four subgroups separately. Combined with other registers such as labor, income and health the CBS regularly produces aggregated period data which is publicly available (CBS: Open Data/Statline). These cover the Dutch de-jure population of the single years. People moving in and out of the country are only included into the period data when they return within the next eight months (emigrants) or do not leave the country within the next four months (immigrants) (Prins, 2016).

Single measures derived from the life table and period mortality data such as age-specific death rates and causes of death are examined to make comparisons at a more refined level. Period life tables in this analysis will be referring to multiple periods at once to get a more lucid overview of recent developments. The periods measured are 1995-1999, 2000-2004, 2005-2009, 2010-2017. The latter period is a bit longer than the ones before to better compare to the following micro-analyses and avoid having a 3-year-period with too much uncertainty. Data for 2018 was not available at the time of the analysis. Thus, exposures and incidences throughout multiple years are added up and treated as one period. Taking these multi-year measures also helps to generate more cases. Especially on behalf of incidences in some of the migrant populations this has been done to reach more reliable results. Overall, the goal of this analysis is to provide a first overview of the situation, to examine whether a healthy migrant effect is the case for either Turkish, Moroccan or Surinamese individuals and how they are shaped by age-patterns and causes of death.

Secondly, a large longitudinal micro-dataset is utilized to surpass the descriptive level and refine the analysis of origin-specific mortality. This data comes from the "Digitized Municipal Personal Records Database" which is a digital registry that encompasses all legal residents and non-residents who reside in the Netherlands. Any individual officially staying in the Netherlands for at least four months needs to register. Non-residents are those who have left the Netherlands but were residing in the Netherlands before. They are kept track of through e.g. social benefit agencies or foreign authorities. The database entails detailed personal data that is kept up to date and saved. All Dutch municipalities are obliged to collect and report this data (Koninkrijksrelaties, 2010). Next to the municipalities, other institutions such as the mentioned social benefit insurers and pension funds are obliged to deliver data to keep the register up to date. Any changes only observed by one specific institution will thus still arrive in the central register and is updated everywhere. Data is saved forever, thus once somebody has entered the dataset its entries are continuously followed-up and cannot leave the system. Entries are also saved for individuals that have already died. Thus, mortality is thoroughly recorded, even historically or if somebody has already left the Netherlands. However, only if recorded. All data that can be updated is registered in longitudinal form. For example, a change of address will not lead to a renewed entry but an additional observation. The former entry will then include beginning and end of residence. This also includes moves to and from abroad, making it possible to exactly trace immigration periods. However, not all cases include the country of origin or destination. Only few variables are updated without saving the former status and some can be erased or changed upon request of the citizen itself (e.g. biological gender). Basic information on individuals ranges from birthdate to information about voting rights, connection to other citizens through family or marriage and possibly, date of death. To ensure that the entries are valid, especially regarding the public, all citizens are required to register moves and changes by law (Prins, 2016). Statistics Netherlands among other state agencies and non-governmental organizations uses it to derive all its population statistics. Moreover, it is used as a basis of sampling for population surveys. It is not freely available, any analyses based on the micro data needs permission by the ministry of security and justice; publication and sharing of micro-data is prohibited due to the privacy implications (Prins, 2016).

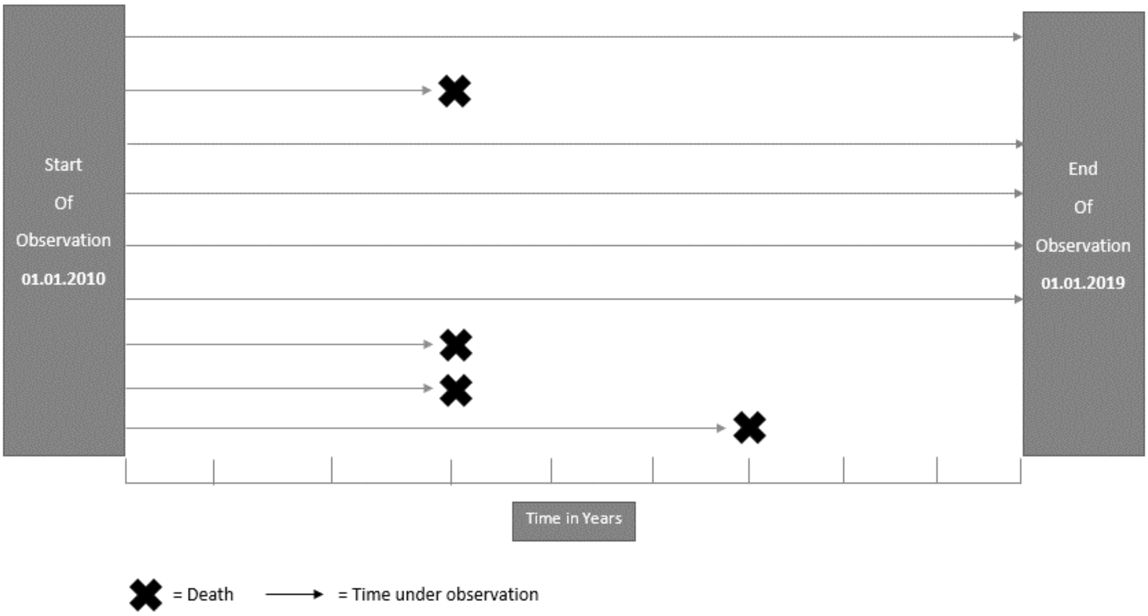
Despite the legal and administrative framework which ensures registration and data-quality, misregistration is still a shortcoming of the data. Thus, it possibly plays a role in the data analyzed in this thesis. Because migrants have a higher probability of migrating compared to Dutch natives, the problem of misregistration is expected to be larger for these three migrant groups. However, the incentive to not register migration-moves e.g. due to social benefits is minimized for the populations in question as they are also entitled to receive pension benefits in their country of birth. Misregistration of moves within the Netherlands is reportedly a problem for about 2 to 3% of the register population (Prins, 2016). However, for this analysis this issue does not play a role as domestic moves are not of any interest. Unregistered migration is also included into the datasets through assuming that people have left the Netherlands when they fail to respond to official mail or are otherwise not available. In this case individuals are assigned an estimated date of emigration to an unknown country (Prins, 2016). For scientific analyses of migration this comes quite handy, as no additional constructs have to be created in order to operationalize loss-to-follow-up. Statistical immortality due to non-reporting of deaths is however still a possibility. Despite the efforts of keeping track of migration, recent estimations suggested that about 33.000 people live abroad without having unregistered in the Netherlands (Statistics Netherlands (CBS), 2017a). Thus, complete security about the whereabouts of residents is not given. But, as the basis of analysis is a subpopulation, in the case of the salmon bias analysis even a subsample, the total number of unregistered migration moves is deemed marginal. Overall, the cooperation of municipalities, ministry of security and justice and social services lead to a high quality in data. This makes it possible that even non-residents are kept in the register, and more importantly, followed-up. Thus, the loss of data is minimized as well as the risk of non-residents becoming statistically immortal. Both of these points are of importance for the following analyses, as the event of interest is either death or emigration.

The micro-data will be used as survival data. Due to their nature as register data, a lot of the information is primarily not produced to fit directly into statistical models. To use it for the latter, extensive work on the data-structure has to be done. Some of it is undertaken by CBS's data-builders, who produce the final datasets on a variety of subjects, more specific stuff has to be done by the analyst himself (Prins, 2016). For this paper, work on the data had to be conducted in order to fit the survival structure. Ultimately, data from the basic demographic data set, which is essentially the basis of everything, is combined with several longitudinal data sets by the means of the individual identifiers. The longitudinal sets include periods of emigration, cohabitation and marital status. To reduce the number of observations, the last entries before censoring are kept while older entries are omitted. This was a necessity as the combination of both datasets would have multiplied many of the individuals and led to amounts of data not processable with the accessible means for this thesis. The resulting set is thus reduced to constant-only variables. The final survival structure is built upon the year of birth and year of death or year of censoring, in case no death occurred. Thus, the exact age is ignored, and the time dimension is measured in life-years. The final population consists of the whole population legally residing in the Netherlands between the 1st of January 2010 and 31st of December 2018. Regardless of when people have entered or left the Netherlands, the whole life course in between these 9 years is taken as the frame in which individuals are exposed to the hazard of dying. Every individual which was legally registered in the Netherlands at some point in time between these two dates and who is either a native-born Dutch, Turkish, Moroccan or Surinamese is therefore within the data. Individuals are also allowed to enter during the observation period. Thus, migrants who entered after 2010 are also included. A lot of individuals from the municipal registration are excluded because they have either left the Netherlands or died before 2010. Individuals exit either through dying or are censored on 1st of January 2019 for those who live through the whole period without an event. The analysis focusses therefore on mortality in between 2010 and 2018, similar to the life table analysis with an additional year. Due to the selective age-structure of the migrant groups, all observations below 15 (age at 2010) are omitted. Despite the theoretic completeness of the data, in practice the registers have a number of migrant individuals that were lost to the administration and have not been assigned a 'theoretic emigration date'. Their demographic characteristics are in the data, they are however without a migration history. Also, some probably have died but were not reported as it occurred outside of Dutch registration. Therefore, all cases of migrants without a migration history (missing entry- and exit-date) are omitted due to the likelihood of being lost to follow-up and the subsequent risk of statistical immortality. For example, some individuals showed peculiarly high ages – up until 126 years in 2018. Besides that, due to the missing migration history it is unknown whether they have been present in the Netherlands in between

2010 and 2019. Additionally, the omission was also carried out to reduce the number of cases and gain computability.

Without restrictions on de-jure population, mortality can be analyzed on a detailed individual basis even surpassing national borders. Analyzing the individual lifeline-data offers a degree of immunity against mechanisms such as the salmon bias, as persons still remain under observation even though they might currently not be officially residing in the Netherlands. Thus, aggregated mortality data would not be influenced by their exposure or incidences. The final data-structure can be found in Figure 1. In it beginning and end of the observation period can be found, time is measured in years and individuals either survive and thus become right-censored on the 1st of January 2019 or die, as indicated by the '**X**'.

Figure 1: Illustration of the Data-structure for the individual Mortality Analysis



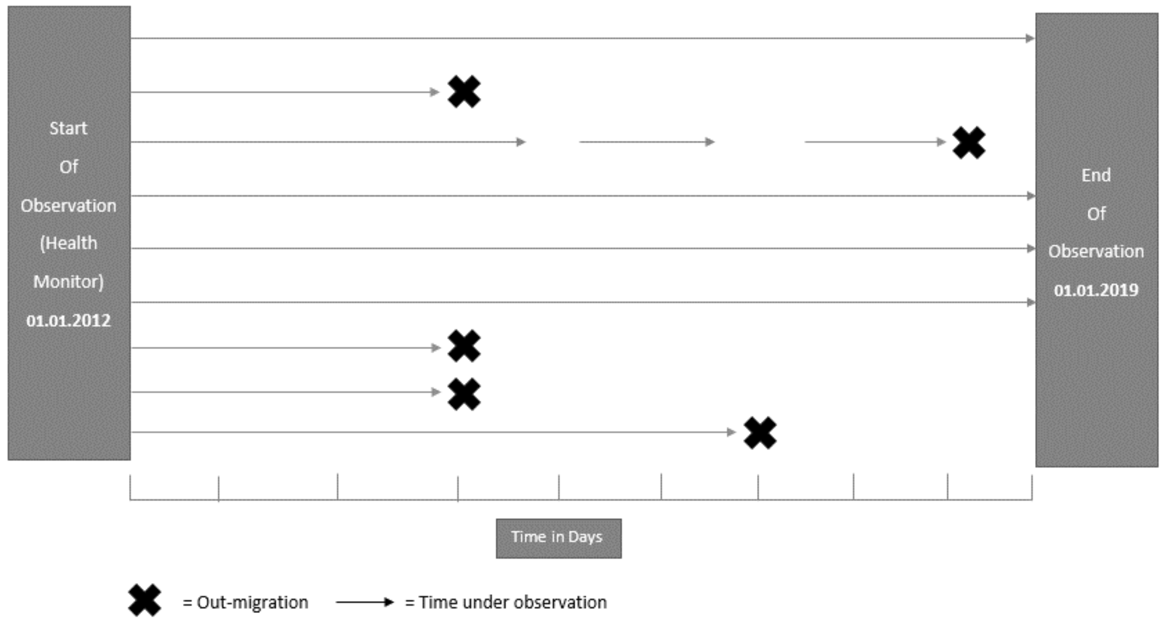
Source: Own illustration

Using this data and type of analysis offers additional insights; as mentioned, one of the possible reasons for the migrant mortality advantage are shortcomings in the data. More precisely, many datasets, especially cross-sectional ones including numbers of deaths, are based on residency in a country. As migrants are more mobile as natives, it could be that this restriction biases the results. Cross sectional life tables essentially only reflect a very specific window of time and the respective mortality. The micro-data analysis poses a more holistic approach that does not ignore large portions of time and also reports on deaths of people that might have left the Netherlands. Overall, this analysis holds the possibility of partly answering whether mortality differences really are inherent to ethnicity/origin and not due to data artifacts or composition. More precisely, migrant-native mortality differentials are analyzed on terms of overall group-differences instead of the instantaneous differences per year. Moreover, the amount of information is bigger, estimations are now based on the length of individual lives instead of rates in aggregated age groups. All in all, the micro-data builds a more detailed step into the uncovering of migrant-native mortality divergence and can also simultaneously control for demographic composition.

The third and last dataset used is a combination of the micro register and a national health survey. It holds the possibility to track individual's migration histories, determine demographic characteristics and health status. All of which are necessary to analyze the salmon bias. However, not the full register can be used as health data comes from the survey "Gezondheidsmonitor" (~ health monitor). The health monitor is a large-scale social survey which was conducted in 2012 and 2016 by CBS, the ministry of health and wellbeing and the Dutch community health services. It includes a variety of indicators on physical and mental health, as well as chronic illness and lifestyle (RIVM, 2019). Similar to the register data, health monitor microdata is not publicly available and needs special research permission. In total 700.000 Dutch residents from age 19 and older have been receiving the survey online or were reached by phone, based on a sample from the population register (Statistics Netherlands (CBS), 2015). The response rate was at about 45 to 50 % depending on the region. A complex weighing model helps to counter selectivity regarding demographic factors, but it is not deliberately usable subsamples (Buelens, Meijers, & Tennekes, 2013). As the salmon-bias analyses systematically exclude Dutch natives, no population weights were applied. The micro-data includes the same individual identifiers as the demographic registers do, this makes it possible to combine survey data with registered migration moves. For this paper, only the 2012 health-monitor data will be used as the more recent data set does not provide an analyzable amount of out-migration cases. Also, it would not serve as a follow-up to the 2012 data since it is a cross-sectional survey and not a panel.

The statistical method applied is again survival analysis; however, the parametrization is different due to the assumed underlying hazard. Also, instead of a focus on lifetime in the survival framework, the analysis is now shifted towards the migration history. Exposure is now given whenever individuals stay in the Netherlands and the event of interest (failure) is out-migration. The first point in time is set to 1st January of 2012, which is the year of the health monitoring. Thus, it is assumed that the health was monitored at exactly that date and the subsequent migration history is followed. Therefore, no additional individuals can enter the observation after it started. Individuals with circular migration patterns are identified and kept in the sample as partly censored with gaps (interval-truncation). The last exit is then assumed to be the final outmigration move. Thus, cases with multiple observations exist as opposed to the mortality dataset. This does not pose any computational problems as the merger with the health monitor only left 6.829 cases with 6.865 records from the original population. Dutch natives are now omitted, too; as their origin is contradicting the salmon bias hypotheses. Shortly put, the theoretical framework of the salmon bias does not encompass natives as they are already located in their country of birth. For all following points in time until 31st December of 2018 the health monitor variables are assumed to be constant. Substantially this implies that individuals do not change in their health. It is rather unlikely that health stays completely the same over the years. Nevertheless, the reported health can serve as an overall indicator of long-standing wellbeing and be more than only a momentary record. Time is measured on the basis of days, as posed by the registration of migration in the migration register. There are no overlaps or exclusions. The data-structure is visualized in Figure 2. Right-censored individuals are those not leaving the Netherlands throughout the whole period, failures (~out-migration) are indicated by the '**X**'. Interval censoring is illustrated by the gaps in the survival-lines.

Figure 2: Illustration of the Data-structure for the individual Salmon-Bias Analysis



Source: Own illustration

The combination of health survey data and detailed governmental registers offers the possibility to research the salmon bias on a detailed individual level. Nevertheless, strong assumptions about the continuity of health remain, and so do uncertainty about representativity and the irreversibility of migration moves. The primary aim of this model can thus be understood to deliver evidence for a possible salmon bias but is not able to deliver specific measurement of life expectancy change due to outmigration of certain individuals.

All in all, the data at hand provides possibilities to operationalize models on both the healthy migrant paradox and selection through salmon effect. Moreover, it can be tested whether differentials in mortality have something to do with the overall composition of the population. Mortality is examined from different angles such as overall life expectancy, major causes of death, age-specific mortality. The salmon bias analysis sheds light on a specific type of social selection through out-migration.

4.1. Variables and Operationalization

The dependent variables in the case of the survival analyses are failures. First, death is examined. If a respondent dies, it gets assigned a 1 for the failure variable at time-point t and subsequently leaves the observation. There are no multiple or competing events. All cases without incidence get assigned missing values for the failure variable. The same accounts for outmigration which is coded similarly.

However, as this event is not necessarily permanent, only moves without return to the Netherlands afterwards are treated as failure. Thus, outmigration incidences can only be assumed final at the end of the observation period.

As mentioned, all variables in both survival analyses are treated as constant. First, origin is taken into the model. This refers to the country of birth of the respondent and at least one parent, to come as close to actual ethnicity as possible. For example, an individual respondent counts as first-generation Turkish if they and one of their parents was born in Turkey. This automatically excludes second-generation migrants and individuals who just happened to be born abroad but are actually of Dutch origin. Subsequently, a four-category dummy variable indicates origin. Similarly, marital status is operationalized as a dummy with four categories. Individuals are either unmarried, married, separated or widowed. The raw data would allow for more specific categories such as differentiation between same sex marriages and opposite sex ones, and the differentiation of those after separation. For the sake of conciseness, they are treated as 'married' or 'divorced'. The remaining two variables, gender and cohabitation, are dichotomous. Individuals are either cohabitating with one or more persons or live in a single household. They are subsequently categorized as either 'cohabitating' or 'not cohabitating' which is oriented on the last register-entry before the respondent's censoring. The gender variable is either 'male' or 'female'.

For the salmon bias analyses, all variables except the dependent construct Y are cross-sectional, stemming from a sperate *survey*, the discussed health-monitor. Thus, also assumed to be constant. In substantial terms: what is measured, is whether the characteristics found in 2012 affect the hazard of out-migration in the subsequent 6 years. However, in order to determine the impact on out-migration hazards within a survival framework they are treated as longitudinal. The health monitor makes it possible to include several metrical confounders. First, income quintiles regard standardized household income according to the overall Dutch population's income distribution. Thus, individuals are put in one of the five categories according to which population quintile they belong. In detail that is: *1=0 up to 20% (max. €15.200) 2=20 up to 40% (max. €19.400) 3=40 up to 60% (max. €24.200) 4=60 up to 80% (max. €31.000) 5=80 up to 100% (> €31.000).* Educational level is measured in 4 categories from low to high. These are determined by the highest obtained degree and categorized by the CBS' standardized procedures of ranking different degrees from low to high (Statistics Netherlands (CBS), 2015). Age and years since migration are obtained through the demographic register, by subtracting the birthyear or the year since someone first arrived in the Netherlands from 2012.

On behalf of the aggregated life tables some additional factors are operationalized. To shed light on mortality differentials in terms of causes of death, the life tables analyze three subgroups. These are cancer and cardiovascular disease as they are the top-ranking unnatural causes of death amongst the Dutch population and have been shown to appear disproportionally amongst different ethnicities

(Statistics Netherlands (CBS), 2017b). Third, despite not being a cause itself, deaths abroad are also included in the analysis as they are interesting to examine due to their salmon bias implications.

4.2. Discussion and Limitations

The studies design is not without its limitations. In the demographic registers, education is not yet completely available for the whole population. Especially older age-cohorts have large amounts of missing values. Due to this systematic bias, education is not included into the mortality-analysis. Which is an unfortunate shortcoming regarding socioeconomic predictors of mortality. Moreover, the individual mortality data offers no reliable way to check for SES overall. Thus, it is not possible to control for it and examine the compositional effect of SES, which is presumably one of the strongest life expectancy indicators. It is reasonable to assume that migrant groups are systematically more deprived than Dutch natives due to empiric findings mentioned above. This component of the healthy migrant paradox is subsequently treated as a part of the conceptual framework, the exact influence can however not be quantified.

The bias of unreported deaths could pose a problem for all analyses of this thesis. Even though the descriptive numbers suggest that statistical immortality is unlikely, the possibility of deaths being not reported cannot be completely ruled out. The strangely old individuals were omitted but further than that no procedures could have been carried out to minimize the risk. Thus, this insecurity remains relevant to mention as it is especially likely in the case of individuals leaving the Netherlands without reporting it. Older studies often refer to the inability of registers to portray emigration, as individuals are not legally obliged to let the municipality know about it when they leave (e.g. Wallace and Kulu, 2014). However, this is not the case in the Netherlands where it is legally required, but still without sanctioning. Dutch registry data deals with these cases in a way of recording not only registered immigration and emigration cases but also the cases that have been removed administratively. This is indicated by a variable for the type of migration. Therefore, the timeframe of immigration or emigration might not be perfectly reflected but the overall assumed moves are in the data. Another shortcoming of the register is the inability to distinguish between ethnicity beyond country of birth. Especially for the racially diverse Surinamese this is a cause of inaccuracy, as it has been shown that the health outcomes of the subgroups are rather different (Statistics Netherlands (CBS), 2017b).

Also, the health monitor data has some problems. Its cases are limited, and due to the specific subsample representation might be limited as well. As mentioned, weighing is not applicable in this instance, which causes the representation-problem. However, for a social scientific sample analysis, the number of cases is still analyzable. As the Dutch are omitted, there is likely no group analytically overshadowing another.

Beyond the health monitor, continuous data on health exists also in population registers. The available data are health care expenditures per person and subdivided by type of expense. As intriguing as this variable may seem, it is not without fallacies as high expenditures are not always a reliable reflection of an individual's actual situation. A dental therapy for example might be much more cost-intensive than a treatment for adult-onset diabetes; however, the latter might impair the subjective experience of health as well as actual mortality risk much more and thus fit much better into a salmon bias. Also, regarding the salmon bias analysis; the remigration concept is only a very broad approximation of the actual phenomenon. The numbers hold very accurate information regarding time but no qualitative information about the intentions of the migrant, also the destination country is unknown in many cases. The individuals could move towards any destination which is a shortcoming as the salmon bias specifically refers to remigration. However, omitting all these cases would make estimations impossible. The consequence is that a number of incidences might have different implications as assumed by the hypotheses. Some may leave to a third country with completely different motivations. To counter this at least to some degree, circular migration is taken into account as interval truncation instead of out-migration incidences. Also, there is an infinite amount of other reasons to migrate except health, unobserved heterogeneity is therefore expected to be high. However, the aim of the salmon bias survival model is not to explain emigration perfectly but examine whether there is any influence at all. Put shortly, whether there is some evidence for health related out-migration or not. The latter factor is also not perfectly in line with the salmon-bias's theoretic framework which specifically refers to remigration into the country of birth.

Lastly, the manifold confounder variables of the survival analyses are all used as constants which limits the interpretability. For example, marital status can change, and nobody is 'born married'. That is however what the model assumes with the simplified data. Individual lifetime under the prior marital statuses is therefore ignored which might lead to overestimation of the respective coefficients.

5. Methods

Following the theoretical concept and the datasets which are accessible for this analysis, three different methodological processes are chosen to empirically evaluate the research questions. For all survival analyses Stata 14 is used, the life tables and mortality ratios are computed with Microsoft Excel and lastly the confidence intervals for mortality ratios are obtained with R.

5.1. Mortality and Life Table Analyses

The period life tables use the cross-sectional data in order to estimate the theoretic life expectancy of a synthetic cohort, given this cohort undergoes the same age-specific mortality patterns of the given period. This is mainly achieved by transforming lifetable death rates ${}_nm_x$ to age-specific risks of death of the cohort ${}_nq_x$ while assuming that ${}_nm_x$ is equal to the observed population mortality rates ${}_nM_x$. Latter is obtained by dividing the observed incidences by exposures (Preston et al., 2000, p. 42). To do transformation an average measurement of life-years lived by those dying in the age-interval ${}_na_x$ is necessary. For this analysis this factor will be approximated by *n/2*, n being the size of the age interval in years, except for the first and the last age-group. Thus, calculations are based on the assumption that individuals are dying on average half-way through the age-interval. This has been shown to be source of only minor inaccuracies (Preston et al., 2000, p. 46). These are negligible, especially as the life tables are computed on basis of the exposures and incidences of multiple years, as an average measurement of broader time frames. The youngest age-group (0-5) gets assigned a ${}_na_x$ value of 2 and the oldest age-group (95<) the same value as e_{95}. The remaining measurements (Survival probability ${}_n\boldsymbol{p}_x$, number of individuals left alive at age x $\boldsymbol{l}_x$, number of deaths between age x and x+n ${}_n\boldsymbol{d}_x$, person-years lived between age x and x+n ${}_n\boldsymbol{L}_x$, person-years lived above age x $\boldsymbol{T}_x$, life expectancy at age x $\boldsymbol{e}_x$) are calculated according to standard lifetable procedures (Preston et al., 2000, p. 45 f.). The focus in comparing mortality patterns and identifying differentials lays on age-specific mortality and life-expectancy. Life expectancies are always compared from age 20 onwards, as the lower ages are often nonsensical to compute due to zero death counts.

The age-specific mortality differences between natives and migrants are further analyzed through standardized ratios. These ratios are obtained by dividing migrant through native death rates (${}_nm_{xi}/{}_nm_{xj}$). Thus, making age-specific differences more easily comparable and age-specific mortality dis-/advantages of the migrant groups as compared to natives visible. Lastly, confidence intervals for age-specific death rates are calculated using a Poisson probability distribution with a Cornish-Fisher expansion (n. A., 2019). The confidence intervals for age-specific life expectancy measurements are calculated with Monte Carlo Simulations (s. Andreev and Shkolnikov, 2010).

As mentioned in chapter 4.1. three major causes of death are analyzed to provide additional insights and a first glance at selection underlying mortality. Therefore, associated single decrement life tables are constructed for each population and cause of death. These delete cause-specific deaths to obtain theoretic values of life expectancy; thus, estimating how life expectancy in each age group would look like if cause of death *i* had been cured completely. Due to the interdependence of life expectancy values and mortality rates of all age-groups it is not enough to simply subtract the numbers of cause-specific deaths (Preston et al., 2000, p. 80). This paper's analysis uses Chiang's (1968) method which assumes "[…] that the force of decrement function from cause *i* is proportional to the force of decrement function from all causes combined in the age interval *x* to *x+n*." (Preston et al., 2000, p. 82). Attached to the standard life tables are the proportion of cause-specific deaths to all deaths R_{-i} which is used to derive the cause-deleted survival probability ${}_np_{x-i}$. Moreover, measurement ${}_na_{x-i}$ is assumed to be higher than ${}_na_x$, as the cohort that misses one cause of death completely in all age groups should by definition live more person-years per period. Subsequently, ${}_na_{x-i}$ is calculated according to a quadratic distribution of deaths within the single 5-year age intervals (Preston et al., 2000, p. 82).

Moreover, to uncover which impact the different causes of death have on the life expectancy differentials between migrants and natives, the life tables are decomposed by age and causes following Arriaga's (1984) method. In this procedure, the effect of age- and cause-specific excess mortality is decomposed into absolute differences in life expectancy between two groups. Thus, it can be traced back which factors are underlying mortality differentials and quantify the effects precisely. Arriaga's method can be applied to a standard life table when the numbers for specific causes of death are known. First, the age-decomposition is carried out and thus the total effect ${}_n\Delta_x$ of mortality differentials in each age group on the overall differential in life expectancy of two populations(Preston et al., 2000, p. 64 f.). The total effect in this analysis is calculated as:

$$ {}_n\Delta_x = \frac{l_x^1}{l_{20}^1} * \left(\frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^2} \right) + \frac{T_{x+n}^2}{l_{20}^1} \left(\frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2} \right) $$

The first half of the equation before the '+' calculates the *direct effect* and the latter half the *indirect and interaction effect* (Preston et al., 2000, p. 64). The superscripts 1 and 2 refer to the populations that are being compared, in this case 1 always refers to Dutch natives and 2 to the respective migrant group. The measurement l_{20} is chosen in order to interpret the resulting values referring to life expectancy at age 20, similar as in the other calculations.

The same process is repeated with cause-specific mortality. To do so, additional proportions of cause-specific deaths to the total incidence of death per age group ${}_nR_{xi}$ have to be computed. Additionally, age-specific all-cause mortality rates ${}_nm_x$ are included into the computation. The resulting cause-specific total effect ${}_n\Delta_{xi}$ can be calculated using:

$$
{}_n\Delta_x^i = {}_n\Delta_x * \frac{{}_nR_x^{i,2} * {}_nm_x^2 - {}_nR_x^{i,1} * {}_nm_x^1}{{}_nm_x^2 - {}_nm_x^1},
$$

Superscripts *1* and *2* are again referring to the same groups while *i* stands for the cause of death (Preston et al., 2000, p. 84). Ultimately, this method shows how single age-groups and specific causes of death contribute to the difference in overall life expectancy between two populations. In this analysis the focus lies on overall causes. The single measurements of the effects can be interpreted as change of years of life expectancy attributable to a cause.

5.2. Survival Analysis

The following chapters go into detail about the two different kinds of survival analyses that are used to empirically measure the research questions of this thesis. Both rely on similar data structures, as laid out in chapter 4, but are different in their assumptions and underlying micro-data. Both are proportional hazard models; they differ however in their parametrization.

5.2.1. Healthy Migrant – Parametric Gompertz Regression Model

The Gompertz survival model is used for the mortality analysis according to the given data and the assumptions about the underlying hazard. It belongs to the parametric family of survival models. These models base their estimations on the information contained within the survival duration and individual characteristics of each individual over the whole observation period. They require a specific parametrization for the hazard function along which the estimates are then fitted. This makes them more efficient but also less flexible than semi- or non-parametric models. The parametrizations are chosen due to the distribution of hazard which is most likely found in reality. For example, when it is reasonable to assume a constant hazard throughout the whole observation, an exponential model would be the best fit. As mortality is the focus of this analysis, a Gompertz model is chosen as it reflects the asymmetrical distribution of human mortality best and has widely been used to analyze it (Cleves, Gould, Gould, Gutierrez, & Marchenko, 2008, p. 231). The hazard and survival function for the Gompertz model are

$$
h(t) = \lambda \exp(\gamma t)
$$

$$
S(t) = \exp\{-\lambda\gamma^{-1}(e^{\gamma t} - 1)\}
$$

λ is the ancillary parameter estimated to determine the overall shape of the function. Gamma determines the slope in which the hazard develops over time. All values above 0 imply monotonic exponential increase over time, below 0 monotonic decrease and 0 itself implies a constant $h(t)=\lambda$. The regression estimates of any parametric model can be interpreted as a standard hazard ratio (Cleves et al., 2008, p. 231 f.). A test to determine whether the chosen parametrization is adequate can be done by applying a cox model and comparing the estimates. If both models yield approximately the same results the chosen constraints can be accepted as realistic. All in all, parametric models are used for analyses of data and respective distributions about whose shape the researcher is relatively sure of. Such as in the case of mortality, which is universally distributed to increase monotonically over an asymmetrical distribution (Cleves et al., 2008, p. 231 f.). The model fit of the Gompertz regression is evaluated through estimating the same combination of covariates within different parametrizations. The resulting logged predicted individual survival times are correlated with the observed logged survival. Ultimately, this test serves to find out which hazard restriction offers the best predictive power. More precisely, which model has the best fit (Cleves et al., 2008, p. 285 f.).

Estimates from some survival regression models can be transformed into an 'accelerated failure time' logic, translating the individual hazards into estimated survival time. This way, independent variables can be quantified on how they alter survival time. For analyses of mortality on the basis of life years this could serve as a way of interpreting gains and losses in life expectancy. However, for the Gompertz distribution this transformation is not possible (Cleves et al., 2008). However, there exists another process in which individual hazard estimates from a survival regression, including Gompertz estimates, are translated directly into measures of life expectancy. This is possible due to the relationship of the conditional survival function and expected survival time (Moser, Clough-Gorr, & Zwahlen, 2015). Life expectancy ***E*** is then derived from the similar parameters, which are also used to estimate the baseline hazard of the Gompertz function. The full form is:

$$\mathbb{E}(T|\mathbf{X}) = \frac{1}{\gamma}\exp\frac{\alpha}{\gamma}\int_{\alpha/\gamma}^{\infty} x^{-1}\exp(-x)dx$$

, (Moser et al., 2015)

This procedure has been carried out and found useful to examine differences in life expectancy due to different factors in a variety of studies (e.g. Missov and Lenart, 2011; Moser et al., 2015; Robertson et al., 2013). For the transformation undertaken in this paper, the formula is broken down. This is done to produce life tables, based on mortality rates ${}_nm_x$ that are derived from the regressions baseline shape, location and scale parameters. The formula is written

$${}_nm_x = \alpha\ exp(\gamma x)\ \beta_j$$

Subsequently ${}_nq_x$, l_x, ${}_nd_x$, ${}_nL_x$ and T_x are computed using the formulas found in Appendix I.

5.2.2. Salmon Bias – Semiparametric Cox Model

The method of choice for analyzing the salmon bias is a cox proportional hazard model, which is a semi-parametric survival model. Semiparametric survival models "compare subjects at the times when failures happen to occur" (Cleves et al., 2008, p. 231). Thus, time varying independent variables are estimated less efficiently than in parametric models, which include all information regardless of failure occurrences. However, as all predictor variables used in this analysis are constant, stemming from a cross-sectional survey, this is not a limitation to be concerned about. As focus lies on failures, the time variable in Cox regressions is not of specific interest. Subsequently there is no intercept or slope parameter and thus no estimations of survival time (Cleves et al., 2008, p. 130 f.). This is unproblematic as the theory behind the salmon bias does not specify any timeframes when it comes to outmigration due to ill health, there is no "duration hypothesis". The main interest lays on whether there is any influence of health on out-migration at all to identify empiric evidence that the salmon bias is a valid concept.

The baseline hazard is not restricted to any shape. However, the coefficients are assumed to affect the baseline in multiplicative and proportional manner (Proportionality Assumption). This means that the baseline hazard $h_0(t)$ can be left completely un-estimated, but its shape is restricted to be a multiplicative variation according to one's combination of confounding variables (Cleves et al., 2008, p. 129 f.). The missing shape-restriction is important as there is no reason to assume any specific form of hazard distribution in between the start and end of the observation (as e.g. compared to mortality by years of age which has a rather clear risk development conditional on time). The interest lies more on the overall influence of health or other morbidity factors on the risk of out-migrating, as stated by the theory. The underlying time variable are single days between beginning and end of the observation period. Hazards are estimated in the following form:

$$h(t|x_j) = h_0(t) exp(x_j \beta_x)$$ (Cleves et al., 2008, p. 130).

Regression diagnostics regarding the proportionality assumptions that underlies the Cox model is carried out through a Schoenfeld residuals test. This method is a Chi²-Test of the model's residuals which checks for relationships between the residuals of the single covariates, the overall model and the time variable t. The respective slope is assumed to be zero, if this is not the case the proportionality assumption is violated and thus the fit of the model is to be reevaluated (Cleves et al., 2008, p. 206 f.).

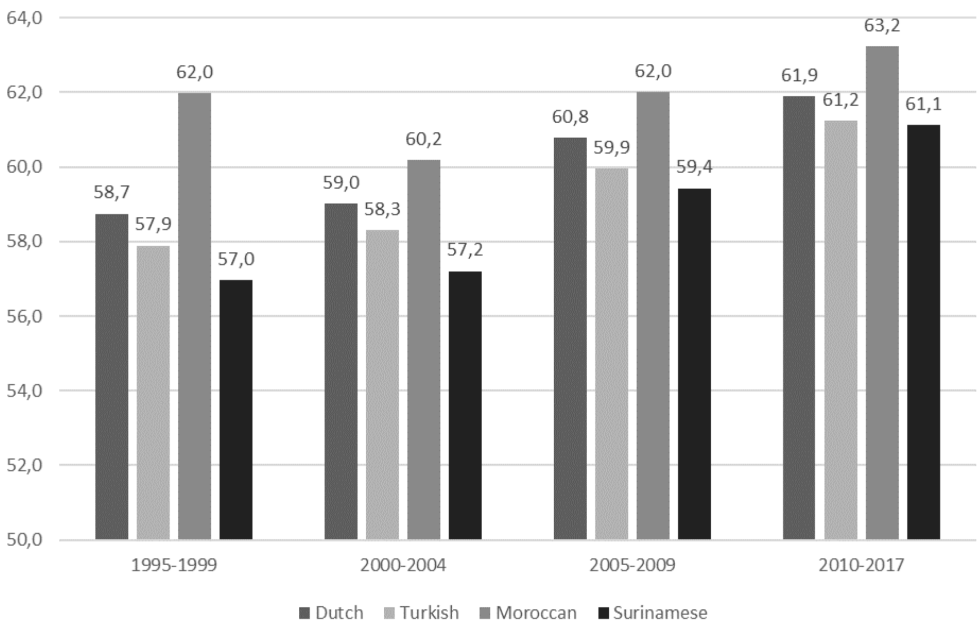
6. Analysis Results

6.1. Macro Data – Life Table Analyses

First, life expectancies of the different ethnic groups are examined. Subsequently all other substantially interesting measures from the life tables are displayed. Life expectancy measures at age 20 throughout different periods of time derived from the appropriate life tables (s. Appendix II) are shown in Figure 3. As mentioned, life expectancy at birth is ignored due to the zero death counts and miniscule exposures at younger ages for the migrant groups. At least from age 20 and onwards sufficient numbers of cases are reached, indicated by the relatively narrow confidence intervals (s. Appendix III). Because of that and the more intuitive interpretation, as compared to for example life expectancy at 15, this specific life expectancy measurement is chosen.

From 1995 onwards, life expectancy has increased for all groups, except for Moroccans who experienced a loss from the first to the second period. It is unlikely that some macro-social event caused this drop. It could possibly be the case that the for example the data collection improved in between the two centuries which eliminated some biases. This would also explain the large advantage that Moroccan showed in the 1990’s. The largest growth happens between the early and late 2000s for all groups.

Figure 3: Life Expectancy at Age 20 – Development from 1995 to 2017

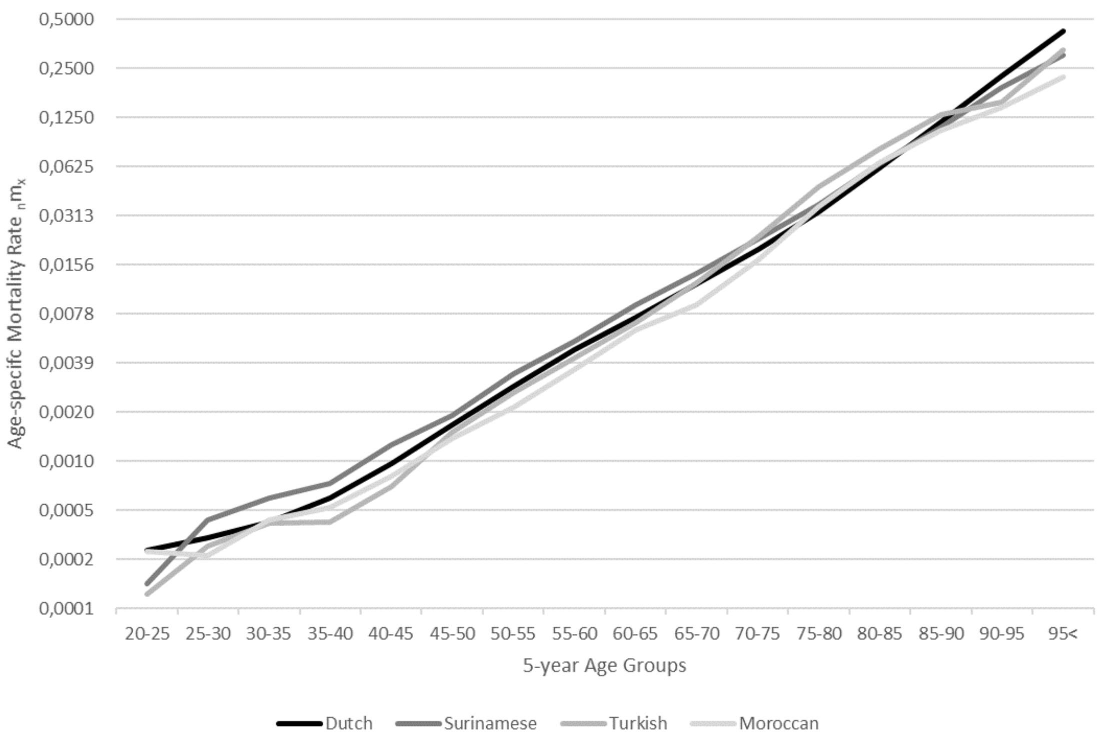


Source: CBS, own illustration

The Moroccans exhibit a very high value already in the late 1990’s and are improving the least in absolute terms, for about 1,2 additional years of life expectancy. The highest absolute growth is found among Surinamese, who gain over 4 years of life expectancy since the first observation period (1995-1999). Second in terms of absolute growth is the Turkish population, followed by the Dutch. In terms of overall ranking however, the Moroccan population always scores highest. Also, the general ranking stays the same over the periods. Moroccans are always on top, followed up by the Dutch, then Turkish and Surinamese seem to have the least favorable values. However, judging from both growths it seems as if there is a slight development towards convergence of all groups. The groups with worse values increase more than those with already high ones. If the rate at which the groups experience gains in life expectancy remains similar in the future, then convergence could be the case.

Figure 4 shows the age-specific death rates of the different life tables in log-linear scale, indicating the frequency of death within the single age-groups. Seemingly, there are no major differences except Turks having higher rates than the other groups in between 70 and 85 years. Moroccans and Surinamese seem to have the most favorable patterns.

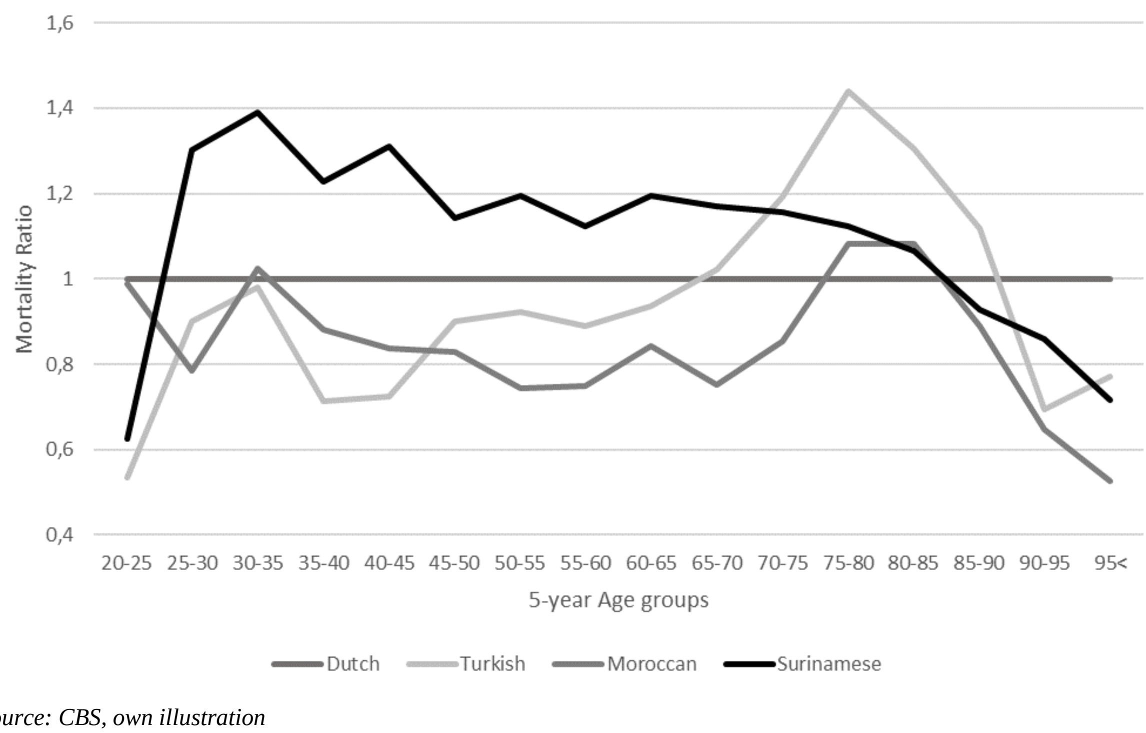
Figure 4: Age-specific Death Rates $_nm_x$ in log-linear scale, 2010-2017



Source: CBS, own illustration

To make differences in age-specific mortality more obvious, the age specific rates of the migrants are divided by the native mortality rate. This results in mortality ratios, as in the migrant-native-ratio for each migrant group (s. Figure 5). The Surinamese are mostly in a disadvantaged position while Moroccans and Turkish groups are in an advantaged stance up until the age of 70. The only exception being the age-group '30-35'. Turks show excess mortality compared to natives between 70 and 90 years of age. This spike can also be found for Moroccans but is much smaller and only ranges from 75 to 85. The Surinamese mortality-curve slowly converges with the one of the Dutch natives over time and then advances for the last three age-groups. For the latter, all migrant groups show a mortality-advantage, but the absolute number of exposure and incidences in these groups are small which could imply that the differences are rather accidental. On the contrary, the confidence intervals of the age-specific death rates (s. Appendix IV) are rather narrow, even for higher ages. Thus, the differentials between the groups can be interpreted as systematical throughout all groups. A clear mortality-advantage as compared to native Dutch is therefore only visible for Moroccans. The Surinamese population is overall disadvantaged, except for the very high ages. Turks show a mixed picture with advantages in adult ages; but have the highest mortality rates for groups above 70 years. Essentially, the Turkish population's advantages throughout the adult ages do not make up for the large disadvantages in between 65 and 90. These gains and losses in terms of life expectancy per age-group can also be found in the age-decomposition section of the tables in Appendix V. Following the theoretical conceptualization laid out above, especially the curves of the Moroccan and Turkish population suggest a positive health selection of adults entering the Netherlands. The convergence and subsequent mortality disadvantage can be seen as an acculturation effect. Most migrants arrive during adult ages, the older individuals have thus mostly spent considerable time in the host-country. Thus, the discussed effects of adopting western lifestyles and low SES can work as an explanation for the excess mortality for older Turks and Moroccans. A first hint for a salmon-mechanism are the mortality advantages in the last three age-categories. The sudden drop in migrant mortality as compared to the years before could indicate an out-migration pattern for increasingly morbid individuals. Interestingly, similar results were found among Moroccans in France and Pakistanis in the United Kingdom (Guillot et al., 2018).

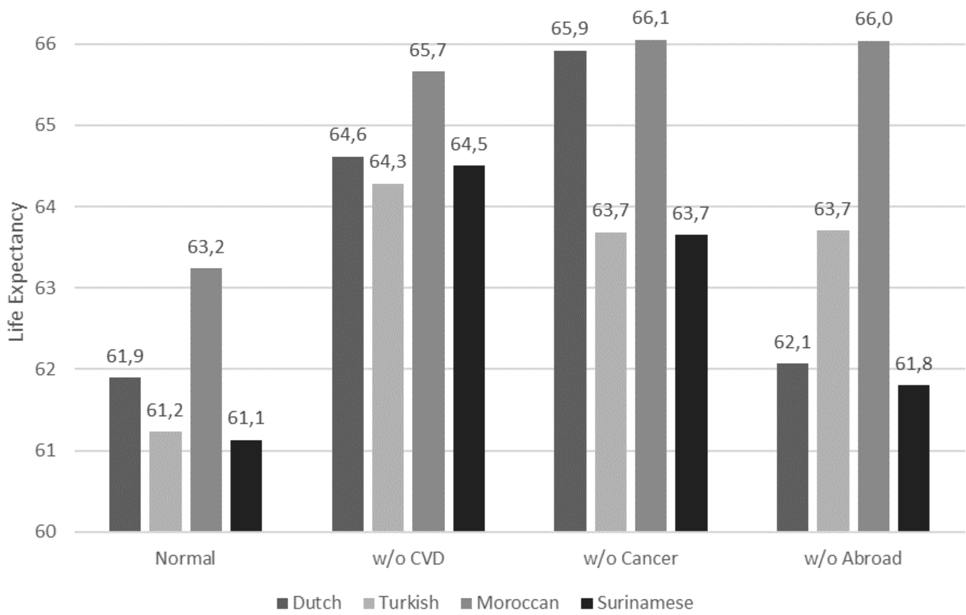
Figure 5: Relative age-specific Mortality Ratios, 2010-2017



Source: CBS, own illustration

Figure 6 shows the life expectancy measures when the major causes of deaths and deaths abroad are eliminated from the life tables. First, cardiovascular diseases seem to affect all groups to a significant amount. Curing cancer would result in lower improvements for the Turkish and Surinamese population. These two groups seem to be more affected by deaths through cardiovascular disease. Moroccans and native Dutch profit highly with about 3 and 4 years of gained life expectancy when cancer is eliminated. Interestingly, potential gains from deleting deaths abroad can be found mostly for the Moroccans and Turkish. For them the gains would be almost as high as for the cancelling out deaths by cancer. The Surinamese and Dutch seem not to be very highly affected by deaths abroad and only show low increases.

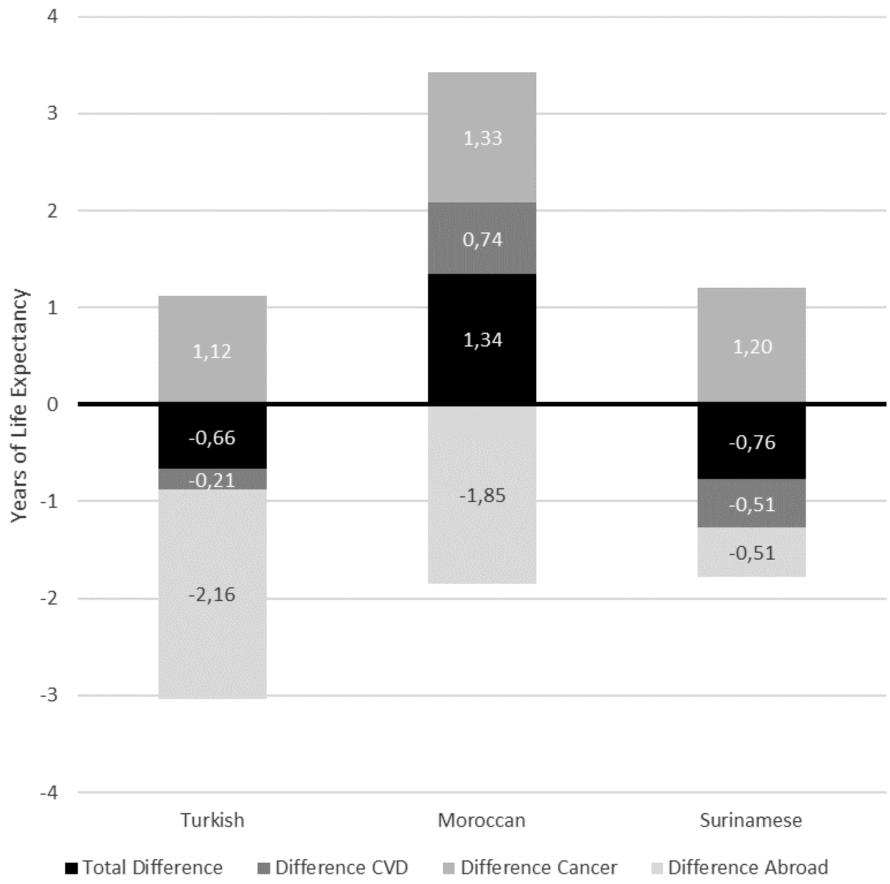
Figure 6: Cause-deleted Life Expectancies at age 20, 2010-2017



Source: CBS, own illustration

When these respective differences are decomposed (Figure 7) it can be seen that all groups have a mortality disadvantage as compared to Dutch due to deaths abroad. These are specifically high for Turks (2,16) and Moroccans (1,85). However, the latter group still retains an overall advantage regarding the total difference in life expectancy and an advantage for both CVD and cancer. Turks are less affected by Cancer as compared to the natives but have a disadvantage on behalf of CVD. About 0,5 years of life expectancy of the disadvantage that the Surinamese have as compared to is attributable to deaths abroad. Similar to Turks, they also have a disadvantage considering CVD. All migrant groups have an advantage in cancer-related mortality of above 1 year.

Figure 7: Decomposition of Differences in Life Expectancy at age 20 by Cause, 2010-2017



Source: CBS, own illustration

It has to be kept in mind that for deaths abroad no specific causes are reported. As shown, the number of deaths abroad can be quite significant for some groups and lead to large increases in life expectancy if analytically eliminated. This could mean that the results of the migrant groups regarding the other causes of death are skewed. They might for example profit more from CVD or cancer-relief than these figures might suggest. On the other hand, the low increases of native life expectancy when deaths abroad were deleted could be a result of more thorough reporting on the causes of death for this group. Moreover, the fact that these deaths are reported in the period life tables means that by definition they cannot be the equated to a salmon bias. The question of which effects are behind possible migrant mortality advantages can therefore not be answered through this part of the analysis. However, this high incidence on officially reported deaths abroad delivers a first hint on a salmon bias that underlies the initial mortality advantage that the Moroccan population apparently has. Moroccans, Turkish and to a smaller degree Surinamese that have deregistered might die abroad and thus cause the salmon bias. Even though the people who have permanently emigrated are not analytically covered in these figures, the large losses in life expectancy due to deaths abroad for Turkish and Moroccans might indicate that there is a possible migration-related selection which leads to a higher initial life expectancy for migrant groups. Put shortly, it seems to be a common occurrence for migrant populations that deaths happen abroad; thus, this can also apply to individuals that have officially left the Netherlands and thus its vital statistics.

6.2. Micro Data – Gompertz Regression

In the survival analysis, the final number of subjects under observation is at *11.354.073* with a total of *1.114.996* failures (i.e. deaths). This number of people consists mainly of native Dutch. They make up 95,2 % of the cases which is 10.893.041. Of all individuals, about 210.000 (1,9 %) are from Turkish first-generation background and both Moroccans and Surinamese are at approximately 170.000 (1,5 %). Migrant numbers in demographic statistics might vary a bit due to different definition of migrants. Since this study focusses on first generation migrants, the numbers may seem reduced as some reports on migrants in the Netherlands also include second generation individuals. The average age at last observation is 57,3 years, with a minimum at 16 years and maximum at 114 years. The confounders will be introduced into the model in a stepwise procedure, creating in total four estimations.

Contrary to the results of the life table calculations, the estimated Gompertz models (s. Table 1) indicate that the instantaneous hazard of dying is the lowest for Dutch natives[1]. This finding is consistent throughout all models except when all covariates are included. Moreover, the life table results for both micro and aggregated data are contradicted. The full model shows that Turks have a hazard disadvantage of about 36,7 % and the Surinamese at about 9,4 % in comparison to the natives. For Moroccans there is a minuscule hazard difference of about 0,2 % below the Dutch. Unsurprisingly, females compared to males have a lower mortality hazard of about 28,7 %. Cohabitation leads to a lower hazard ratio. Individuals that live together with at least one other person have a 10,6 % lower instantaneous risk of deaths than people in single households. Compared to singles, married and widowed people have lower mortality risks. Especially the difference of single and widowed individuals is high, latter hazard is about 48,7 % percent lower. Separated people face a hazard disadvantage of about 6,9%.

The intercept is at almost zero for every model, which is normal concerning the nature of the Gompertz distribution and the distribution of mortality in the data. More precisely, the youngest age-groups are at age 15 at the beginning of the observation and therefore show mortality hazards of almost 0; it is only for older ages that mortality drastically increases until reaching 1. Gompertz models also always give out Gamma-values that hold information about the shape of the distribution. In this case it is above zero across all models, thus the curve rises monotonically as implied by the position of Gamma in the Gompertz hazard function (s. Chapter 5.2.1.). More precisely, it can be interpreted as the percentage-based increase per year of life. Thus, in all models the risk of dying increases about 13 to 14 % each year, starting at almost 0 as indicated by the intercept. With introduction of the different covariates,

[1] Constructing life tables on the basis of the micro data yielded questionable results and were not included into the results-section of this thesis due to possible errors in the data-extraction.

some groups are identified to have favorable demographic compositions and others do not. When sex is introduced into the model, the hazard ratio of Surinamese rises while Moroccan and Turkish ratios decrease. This is due to the higher share of females found in the Surinamese individuals and the male dominated populations of Turks and Moroccan. This shows that some of the mortality disadvantage of both latter groups is actually due to a gender-effect. Also, the natives lose a lot of their initial mortality advantage when more covariates are introduced, indicated by the lower hazards of the origin estimates except for Surinamese. This shows that much of the 'original' disparage between natives and migrants is due to the favorable social structure of the Dutch regarding marital status (s. change between Model 3 and 4). This is especially important for the Surinamese population.

Table 1: Results Gompertz Regressions

Gompertz regression – Model Comparison
No. of subjects = 11354073
No. of observations = 11354073
No. of failures = 1131600

_t	Model 1	Model 2	Model 3	Model 4
	Haz. Ratio	**Haz. Ratio**	**Haz. Ratio**	**Haz. Ratio**
Origin (Ref.: Dutch)				
Turkish	1,491***	1,448***	1,439***	1,367***
Std. Err.	0,01688	0,0164	0,0163	0,0155
Moroccan	1,173***	1,083***	1,052***	0,998
Std. Err.	0,01553	0,0143	0,0139	0,0132
Surinamese	1,275***	1,301***	1,337***	1,094***
Std. Err.	0,01290	0,0132	0,0135	0,0111
Sex (Ref.: Male)				
Female	-	0,634***	0,677***	0,713***
Std. Err.		0,0012	0,0014	0,0015
Cohabitation (Ref.: No)				
Yes	-	-	1,228***	0,894***
Std. Err.			0,0027	0,0037
Marital Status (Ref.: Single)				
married	-	-	-	0,894***
Std. Err.				0,0059
separated	-	-	-	1,069***
Std. Err.				0,0055
widowed	-	-	-	0,513***
Std. Err.				0,0025
Intercept	*5,76E-07*	*5,98E-07*	*4,39E-07*	*5,13E-07*
Std. Err.	*3,49E-09*	*3,65E-09*	*3,65E-09*	*3,83E-09*
Gamma	*0,1342436*	*0,1370879*	*0,1395441*	*0,1433329*
Std. Err.	*0,0000757*	*0,0000777*	*0,0000843*	*0,0000843*

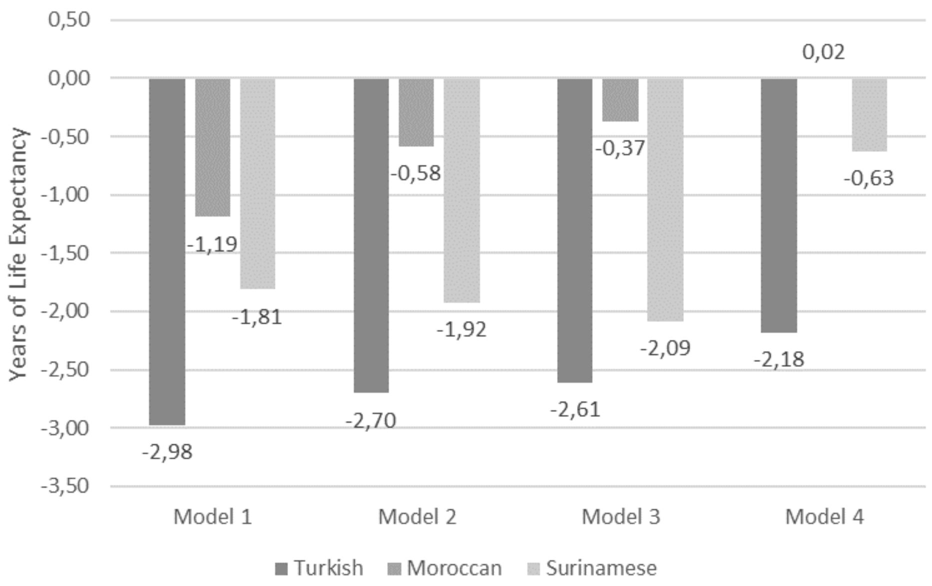
* p<0,10, ** p<0,05, *** p<0.001

Source: CBS, own calculations

As the whole population is analyzed instead of a sample, the significance levels are of much interest. However, it is interesting to see that for the last model, the Moroccan mortality hazard is so similar to the one of the natives that even for these large numbers of cases the regression estimates do not yield significance. Thus, it can be stated that the difference between the two is so miniscule that it could as well be at random, when inferred to a superordinate population. Also, this is surprising as large sample sizes are likely to yield significance even for very small differences.

As discussed above (s. chapter 5.2.1.), the regression estimates can also be used to be translated into life tables using the intercept, gamma and the age specific hazards. Thus, regression results become directly comparable to those of the life tables. The direct comparison is only applicable to Model 1, as the earlier life table calculations had no compositional factors involved. However, for analyzing compositional effect it is interesting to illustrate the changes in relative life expectancy differences for each demographic factor that is controlled for, starting at model 2. This is laid out in Figure 8 which shows the relative differences of life expectancy according to the origin estimates of each model. Most interestingly, Moroccans catch up completely with Dutch natives and even take over very slightly. Also, Surinamese come closer to native life expectancy as well as the Turks. However, latter group still remains at a clear disadvantage even when all confounders are included into the model.

Figure 8: Gompertz Estimates transformed into Life Expectancy (at age 20) Differences

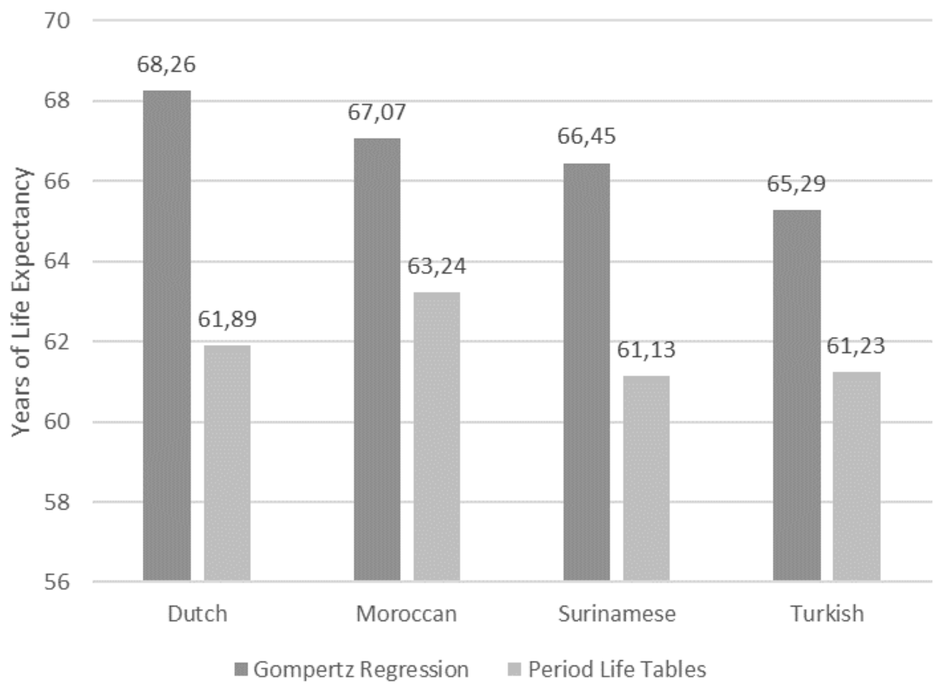


Source: CBS, own illustration

Through the transformation, interpretation can be done in additive terms. This offers a more intuitive interpretation regarding the combination of confounding variables (Models 2-4). Displayed in Appendix VI are the relative gains and losses implied by the single hazard estimates. Being female implies gains in life expectancy of 2,4 to 3,3 years depending on the model while holding all other covariates constant. Small gains in life expectancy can also be noted for married and cohabitating individuals (both: +0,8).

The biggest gain is noted for widowed individuals with 4,7 years. Life expectancy losses for migrant groups are uniformly found throughout all models, the magnitude changes. The ‘theoretic individuals’ with the longest estimated life expectancy at age 20 are thus Dutch or Moroccan females, that are cohabiting and widowed (64+0,0+2,4+0,8+4,7 = ***71,8***). These implications are much less obvious in the original regression table as relations are not distributed linearly, but through the Gompertz curve that follows the factor gamma.

Figure 9: Comparison of Life Expectancy Measurements at age 20



Source: CBS, own illustration

The values of life expectancy from model 1 are displayed in comparison to the values that resulted from the life table analyses in Figure 9. As mentioned, the life table analysis does not take into account any covariates; model 1, which does not control for any covariates either, is therefore the most accurate point of comparison. Estimation on the basis of the non-restricted micro data within a survival model leads to gains in life expectancy for all groups. All populations exhibit higher values of life expectancy in the survival case. As the regressions have shown, the mortality advantage that Moroccans exhibited in the cross-sectional life tables disappears when micro data is analyzed. Same accounts for the small advantage that the Turkish population had compared to the Surinamese. Thus, both the nature of the data as well as the estimation method influence the results. First, the proportionality assumption of the Gompertz model as well as the omission of the individuals under 15 might have pushed the survival times upwards. Whereas for life expectancy through life tables, ignoring individuals from certain age groups does not influence the overall calculations. The restriction to proportionality might lead to life expectancy estimates that ignore the large fluctuations of age-specific mortality differences (s. Figure

5), which is not the case for life table estimations. Third, registration of deaths abroad might be more problematic than assumed.

The fit of the full Gompertz model is tested by correlating the predictions of logged survival-time that it provides to those of other parametrizations (Weibull, Exponential) and the observed logged survival time. This is done with a simple bivariate Pearson's correlation estimates between these variables. The results (s. Table 2) show that neither model predicts individual survival time for those individuals that died within the period specifically good. The best one is the Gompertz model, closely followed by the Weibull distribution. Both models are almost the same in their predictions as indicated by their high correlation (0,995). The exponential model does not fit at all which is unsurprising considering that the observed risks of death over the ages are far from constant. The Gompertz distribution can thus be considered the best option. Moreover, estimation of the models within a Cox regression yielded almost identical parameters – the chosen parametrization of the baseline hazard is therefore deemed fitting. The overall lack in predictive power could also be the reason that the transformed life expectancy is much higher than the life table measurements.

Table 2: Correlations between observed and predicted Survival

Pearson's R - Correlations	Log observed Survival	Log predicted Survival (Gompertz)	Log predicted Survival (Weibull)	Log predicted Survival (Exponential)
Log observed Survival	1,0000	-	-	-
Log predicted Survival (Gompertz)	0,4231	1,0000	-	-
Log predicted Survival (Weibull)	0,4186	0,9952	1,0000	-
Log predicted Survival (Exponential)	-0,4605	-0,6928	-0,6599	1,0000

Source: CBS, own calculations

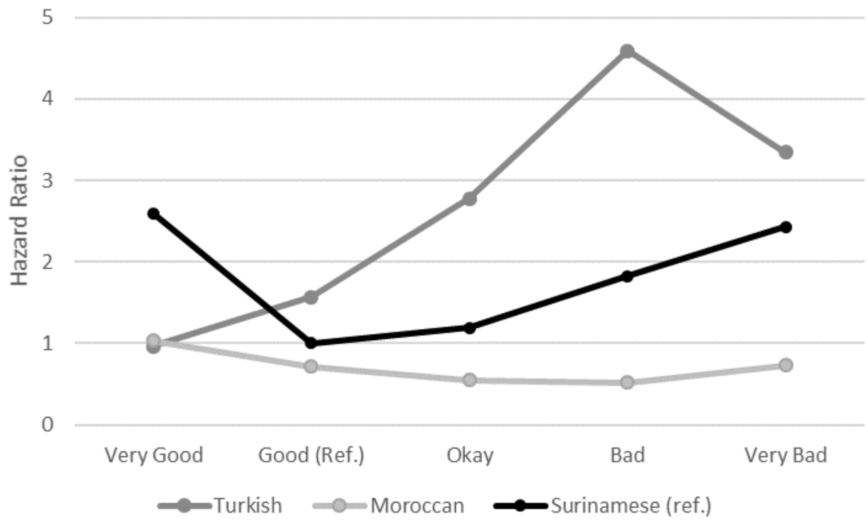
6.3. Health Monitor – Cox Hazard Model

Most respondents in the health monitor sample are from Surinam, closely followed up by Turkish, and Moroccans with the lowest number of people. This ranking changes for the cases that out-migrate. By far the most leavers are Turkish. Almost 60 % of all departures concern Turks, followed by the Surinamese with 27 % and the Moroccans with 14 %. Over the years, all groups experience their highest occurrences of out-migration in the years 2014 and 2015. After that, overall migration slows down with the exception of a few small increases. The total number of failures is at 279 out-migration occurrences. Overall, 16.703.270 days at risk are under observation. The maximum number of observations per respondent is 2, indicating no cases characterized by circular migration that could bias the results. 36 subjects have a gap, amounting to 16.179 days lost to the observation. The two main independent variables, that is self-reported health and number of chronic diseases, are tested within two different models each. The first model tests the general combination of covariates and includes either self-reported health or number of chronic impairments. The second model then tests the respective interactions between origin and the two main independent variables. The self-reported health model analyzes 6127 respondents in total, which is a slight reduction to the migrant subsample of the health monitor. Similarly, the number of failures is reduced. This is because the Health Monitor variables are subject to missing values in contrast to the register variables. Nevertheless, both health indicators utilized in these models have the most valid cases out of the health indicators of the health monitor. Due to preliminary diagnostics, including checks for the proportionality of the variables, both main predictors are treated as categorical.

It appears that self-reported health has a significant effect on the hazard of out-migration, but only for some of the categories. In model 1, worse categories of self-reported health lead to higher hazards of out-migration, as compared to 'good' health (s. Table 3). The highest effect is reported for individuals reporting 'Bad' health, as they experience over two times higher risks of leaving the Netherlands than those with good health. The other significant categories are 'Very Bad' and 'Okay'. Thus, the effect of self-reported health on the hazard of remigration is not linear, as the second worst category of health scores higher hazards than the worst. Turkish respondents are much more likely to out-migrate in comparison to Surinamese (HR: 1,649). Moroccans however are less likely (HR: 0,481). Both effects are significant at the 95%-level. The effect of sex and years since migration are stable throughout both models, for both effect size and level of significance. Females are about 35 % less likely to leave the country as compared to males while holding all other variables constant. Every additional year spent in the Netherlands reduces the hazard of leaving by more than 5 %. The other included control variables yield no significant effects at all. Income, age and education seem not to affect out-migration at all. These effects are subject to large differences when an interaction term is included in the model (s. Table 3, Model 2). Both the origin as well as the health effects, except that for 'Very Good' health, lose their significance completely. Only the interaction term of Turkish origin and 'Very Good' health is highly significant. As opposed to model 1, now better health indicates higher out-migration hazards. The interaction of origin and health only systematically affects out-migration for Turkish respondents. More precisely, the said interaction for Turkish and self-reported health now lowers risks, thus reducing the impact of the health category itself. However, the total effect for Turkish individuals with very good health still indicates a high probability of out-migration. The notion that worse health making individuals more likely to leave is completely contradicted. Very bad and bad health also show high hazard ratios, which would hint on a u-shaped effect of both very healthy and unhealthy individuals leaving more likely. Unfortunately, those effects do not reach significance and can therefore not be interpreted as systematic. Moreover, the suggestion that health and migration play a role, mostly for Moroccans made by the cause-deleted life tables above can also not be confirmed.

By calculating the respective cumulative effects for each migrant group Figure 9 displays an intuitive interpretation of the impact of the interaction term for Moroccan and Turkish respondents in reference to Surinamese individuals. The Surinamese hazard ratio consists only of the net effect of the health indicators due to it being the reference-group for origin estimates and thus interactions as well. At the 'Very Good' level, Turkish and Moroccans do not differ from each other, Surinamese show the highest risk. Moroccans have subsequently lower probabilities of leaving the Netherlands for any level of health. The opposite is true for Turkish individuals who are much more at risk of out-migration, with the highest levels for 'Bad' health. However, except for Turks and 'Very Good' health these effects are not significant due to their underlying regression estimates.

Figure 10: Hazard Ratios by Origin and Health Status including the Interaction-Term



Source: CBS, own illustration

Table 4 shows the same models but with the number of chronic impairments as the main predictor. Estimating the effect of chronic impairments on the hazard of out-migration leads to a model with a further reduced number of overall cases and failures. This is due to the missing values that some respondents exhibit for this health indicator. The result pattern of this analysis is similar to the SRH analysis. When not including interaction terms (s. model 1), a higher number of chronic impairments is positively correlated with a higher probability of failure. Having four diseases indicates a higher instantaneous hazard of failure as indicated by the estimates of the single categories for less than four diseases. However, the effect is again non-linear. For example, having no chronic impairment stands for a 45 % lower risk while having two stands for a 54 % lower risk. More precisely, in reference to having four or more chronic impairments, those with two chronic impairments have the lowest risk of leaving the Netherlands. Similar to the self-reported health analysis, the Turkish origin estimate indicates higher migration-probability (HR: 1,8) and Moroccan variable a lower one (HR: 0,46). Turkish individuals are more prone towards out-migration throughout all models. Also, in these two models, sex and years since migration are stable and significant.

Table 3: Results Cox Regression – Self-reported Health

Cox regression

No. Of observations = 6161

No. of subjects = 6127

No. of failures = 218

Time at risk = 15026069

_t	Model 1	Model 2
	Haz. Ratio	**Haz. Ratio**
Self-reported Health (Ref. Good)		
Very Good	1,319	2,598**
Std. Err.	0,313	0,931
Okay	1,447**	1,191
Std. Err.	0,254	0,407
Bad	2,279***	1,826
Std. Err.	0,475	0,821
Very Bad	1,996*	2,429
Std. Err.	0,814	1,821
Origin (Ref. Surinamese):		
Turkish	1,649**	1,569
Std. Err.	0,285	0,450
Moroccan	0,481**	0,713
Std. Err.	0,114	0,260
*Interaction - Origin*SRH:*		
Very Good * Turkish	-	0,236**
Std. Err.		0,134
Very Good * Moroccan	-	0,560
Std. Err.		0,356
Okay * Turkish	-	1,487
Std. Err.		0,595
Okay * Moroccan	-	0,649
Std. Err.		0,375
Bad * Turkish	-	1,601
Std. Err.		0,804
Bad * Moroccan	-	0,398
Std. Err.		0,310
Very Bad * Turkish	-	0,878
Std. Err.		0,800
Very Bad * Moroccan	-	0,420
Std. Err.		0,536
Female	0,649**	0,637**
Std. Err.	0,090	0,088
Income-quintiles	0,980	0,978
Std. Err.	0,056	0,056
Age	0,995	0,997
Std. Err.	0,007	0,007
Years since migration	0,947***	0,946***
Std. Err.	0,006	0,006
Education	1,022	1,014
Std. Err.	0,074	0,074

* p<0,10, ** p<0,05, *** p<0.00

Source: CBS, own calculations

Table 4: Results Cox Regression – Chronic Impairments

Cox regression
No. of observations = 5.452
No. of subjects = 5.422
No. of failures = 191
Time at risk = 13.290.857

_t	Model 1 **Haz. Ratio**	Model 2 **Haz. Ratio**
Chronic Impairments (Ref. ***Four*** *or more)*		
No Chronic Impairment	0,547**	2,511*
Std. Err.	0,119	1,324
One Chronic Impairment	0,820	3,711**
Std. Err.	0,163	1,861
Two Chronic Impairments	0,460**	1,281
Std. Err.	0,125	0,813
Three Chronic Impairments	0,607*	2,213
Std. Err.	0,175	1,401
Origin (Ref. Surinamese):		
Turkish	1,798**	7,381***
Std. Err.	0,331	3,497
Moroccan	0,462**	1,211
Std. Err.	0,121	0,739
*Interaction - Origin*Chronic Illness:*		
No Chronic Illness * Turkish	-	0,133***
Std. Err.		0,077
No Chronic Illness * Moroccan	-	0,312
Std. Err.		0,237
One Chronic Illness * Turkish	-	0,138***
Std. Err.		0,077
One Chronic Illness * Moroccan	-	0,225*
Std. Err.		0,176
Two Chronic Illnesses * Turkish	-	0,273*
Std. Err.		0,197
Two Chronic Illnesses * Moroccan	-	0,589
Std. Err.		0,559
Three Chronic Illnesses * Turkish	-	0,194*
Std. Err.		0,143
Three Chronic Illnesses * Moroccan	-	0,342
Std. Err.		0,354
Female	0,620***	0,596***
Std. Err.	0,093	0,089
Income-quintiles	0,961	0,952
Std. Err.	0,058	0,058
Age	1,001	1,001
Std. Err.	0,007	0,007
Years since migration	0,947***	0,946***
Std. Err.	0,007	0,007
Education	1,019	1,016
Std. Err.	0,079	0,080

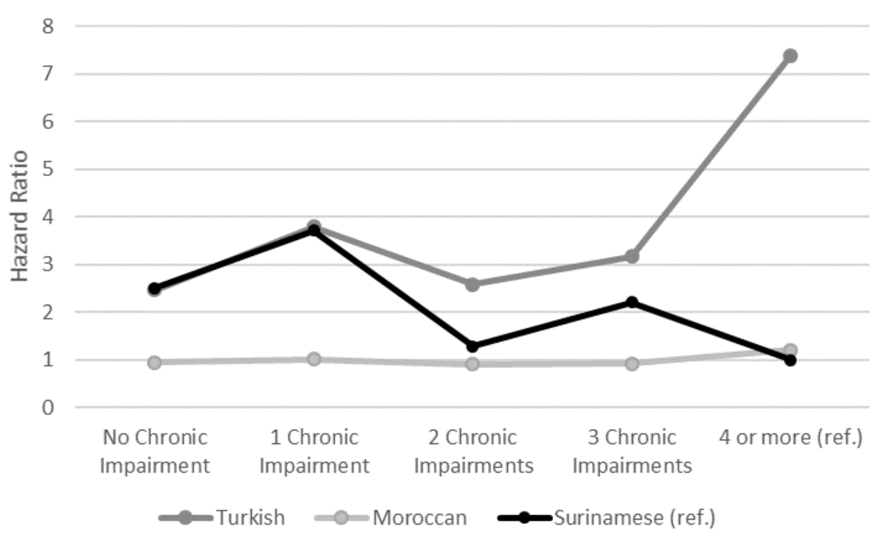
* p<0,10, ** p<0,05, *** p<0.00

Source: CBS, own calculations

The inclusion of the interaction term once again causes a huge change of the health indicators and the Turkish origin dummy-variable. What stands out is the extremely high hazard ratio for Turkish origin. Also, all impairment-dummies now indicate higher risks as compared to their reference point. All the interaction terms reach significance for Turkish individuals. They are all well below one; however, the large net-effects for origin and no or one chronic impairment show that out-migration is highest for Turkish individuals with one or no chronic impairment. Moroccans and Surinamese are again much less likely to re-migrate. Again, the results are quite contradictory to the assumptions of the theory, already in model 1 the categories of the health indicator are seemingly mixed up against their numerical order. Also, Turkish respondents seem to be by far the most prone to out-migrate. Interestingly, years since migration and gender have a consistent and significant effect throughout all models. Females are always much less likely to leave the Netherlands, as well as those individuals who have already been in the Netherlands for longer periods of time. The typical 'leaver' is thus a Turkish male with a short stay in the Netherlands, either none or one chronic impairment or with 'very good' health (s. Table 3).

Figure 11 displays again a comprehensive overview of the interaction effect and the resulting total effects for each group. Reference in this case are again Surinamese individuals, their line represents the net-effects of the chronic illness dummies. Moroccans differ almost not at all, again indicating low out-migration risk due to health for this group. The outstanding estimate for Turkish origin is balanced out by the interaction terms to a certain degree. However, they are more likely to out-migrate for almost all levels of chronic impairments with the highest net-effect for Turks with one impairment. Generally, one chronic disease is associated most highly with out-migration hazards. Interestingly for Turks there is almost no difference between having no or two illnesses. The high net-effect of the same groups also implies an outstanding hazard ratio for '4 or more' chronic impairments.

Figure 11: Hazard Ratios by Origin and Chronic Impairments including the Interaction-Term



Source: CBS, own illustration

For an evaluation of model-fit, Schoenfeld residuals are analyzed with a Chi^2-test (Table 5). The tests account for the proportionality of the single covariates as well as the full model. For self-reported health both models do reach significance for different levels. This indicates that the proportionality hypothesis has to be rejected for both of them. Testing the models that estimate the effect of chronic diseases leads to different implications. For both models, non-proportionality can be rejected, the model seems to fit the data. However, without the most promising confidence.

Table 5: Test of proportional Hazards Assumption

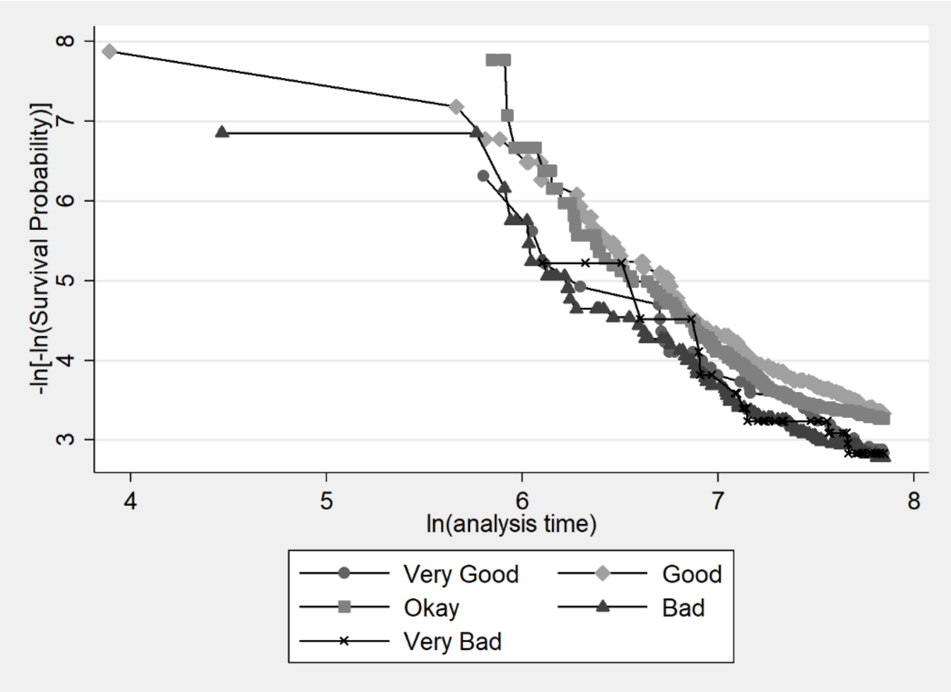
	Chi^2	**df**	**$P>Chi^2$**
Self-reported Health – Model 1	19,23	11	0,057
Self-reported Health – Model 2	39,15	19	0,004
Chronic Impairments – Model 1	15,67	11	0,154
Chronic Impairments – Model 2	24,34	19	0,150

Source: CBS, own calculations

As the health indicators, which are of focal interest in this analysis, yielded questionable results regarding the proportionality of the different levels, further graphic evaluations are carried out. Even though the variables were already treated as categorical, this is interesting as to examine where the peculiar results stem from. This analysis is done in form of log-log plots of survival probability and analysis time for the different levels of both health indicators. If the proportionality assumption was true, one would expect the different levels to show similar developments regarding survival probability over time. Thus, more or less parallel lines. Both self-reported health as well as the number of chronic diseases do not show these as can be seen in Figure 12 and Figure 13. There are a lot of intersections of the different curves, some of the levels cannot be distinguished from one another (e.g. ‘okay’ and ‘good’ self-reported health). Moreover, theoretically as assumed by the salmon bias hypothesis, there should be a continuous gap from good to bad health regarding the risk of out-migration. Put shortly, the worse the health, the higher the out-migration probability should be. However, the levels are rather mixed in their survival probabilities. People with ‘good’ or ‘okay’ health always show higher probabilities of survival than people with ‘very good’ health. However, worse health statuses imply the worst survival probabilities which is in line with the expectations. A similar picture is shown for the number of chronic diseases. The lines are overlapping each other, sometimes across multiple levels. Also, the logical ranking in survival probability is not given and even more mixed up than for self-reported health. Having 1 chronic disease implies the same survival probability as having ‘4 or more’. Medium levels seem to have the biggest probabilities of survival and individuals with 0 chronic impairments are somewhere in between. The rankings are shown to change over time, especially ‘4 or more’ chronic impairments change from highest to lowest survival over time. This initial advantage might however be due to random error within the first few failure occurrences.

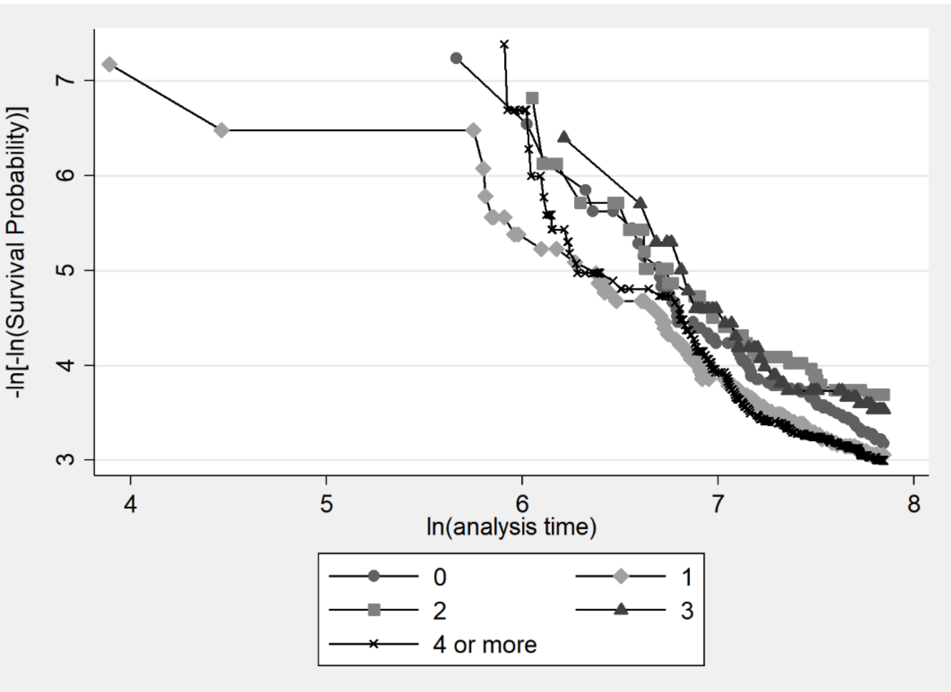
All of the above, reinforces the choice to include the health variables into the models as category-dummies as interpretation in continuous terms would be impossible and the proportionality assumption gravely violated. It has to be kept in mind that these graphs do not represent any inductive hypotheses tests, randomness cannot be ruled out. However, the regression estimates and these graphs jointly show that out-migration risks are not distributed randomly throughout different states of health, just not as it was expected.

Figure 12: Log-Log Plot self-reported Health



Source: CBS, own illustration

Figure 13: Log-Log Plot Number of chronic Impairments



Source: CBS, own illustration

Overall, the time variable shows that after a certain period of time (ca. ln(t) = 6,5), the different levels of the variables are converging towards each other. This implies that the cross-sectional time measurement of health is not fit to cover such large observational periods. Health statuses might change over time and this might have reflected upon this analysis and graphs. For out-migration a couple years after 2012, the health monitor indicators are likely to decrease in their predictive power. Leaving the Netherlands is most likely for one chronic impairment and bad health in the beginning. Additionally, very good health also has a rather low survival curve. This implies that either healthy or unhealthy individuals move more frequently which suits the theoretic considerations that either those who can afford the personal cost move frequently and those who are in bad shape leave the host country. However, as the regression showed, many of these estimates are not significant. Substantial interpretation on behalf of them is therefore purely speculative. Additionally, it is questionable to assume that more chronic impairments necessarily hint on worse well-being. From the large array of health issues that would qualify for the definition of chronic a single impairment could be much worse than multiple ones. For example, heavy diabetes could have a much higher impact on subjective well-being than chronic joint pain combined with chronic headaches. Figure 13 is another indicator for that.

The salmon bias seems most likely for the Turkish population. As shown throughout the analysis, the Turkish population fares rather worse in mortality and their disadvantage in life expectancy as compared to Dutch natives was highly attributable to deaths abroad (s. Figure 7). However, Turks were not advantaged in the first place, thus a salmon bias would imply even less favorable mortality patterns for this group. The ones that were also thought to be affected by health-migration-patterns, namely Moroccans, seem to be rather uninfluenced by health. Same applies to the Surinamese population.

The model fits paint a rather unideal picture of the computations. It seems as if the health indicators are rather unqualified for restricted proportional analyses. Moreover, the large timeframe seems to blur the results. Overall, these models are far away of showing ideal fits, the estimations are quite not as expected and the problem of representativity also still remains. Any specific interpretations regarding the salmon bias are therefore rather contestable. However, this analysis shows that bad health alone unlikely leads to a systematic out-migration, especially for Moroccans who were thought to be affected by any possible salmon bias.

7. Discussion and Conclusion

This thesis confronted the question whether a healthy migrant paradox is at hand in the Netherlands and what possible mechanisms are behind it. Three migrant groups, namely Turkish, Moroccans and Surinamese were analyzed in comparison to the native Dutch population. The migrants' socioeconomic status in the host-society was identified as disadvantaged, any mortality advantage could thus be deemed paradoxical. Different datasets and methods were utilized to shed light on the research questions, both life tables with aggregated data as well as survival analysis with individual data were carried out. Moreover, the aggregated data was always based on the de-jure residential population of single years, while the individual data surpassed residential status. The Dutch data and its quality made it possible to analyze both individual followed-up data as well as deaths abroad, which is an advantage as compared to other European register data. Especially the Turkish and Moroccans deaths abroad where influential, as shown by the life expectancy decomposition. This absence of register-bias might be one of the reasons why the Mediterranean groups have no or only low mortality advantages, as compared to results from other European studies.

Only Moroccans had higher life expectancy as a result of the life table analyses and thus illustrate the healthy migrant paradox. Moreover, Turks had mortality advantages throughout most adult age groups as compared to natives. This was however lost in older ages, when they showed excess mortality. Surinamese where in a mostly disadvantaged state, throughout all analyses. All groups have made gains of life expectancy over the last decades and showed a tendency to convergence. Cause-deleted analyses showed that especially Turkish and Moroccan populations showed large groups of people dying abroad, which lead to the conclusion that they might be susceptible to a salmon bias.

This picture changed with the analysis of individual register data which is not restricted to residential status in the Netherlands. The respective life tables reported an even larger advantage of the Moroccans whereas the survival analysis with the same data reported the Dutch to have the lowest mortality hazard. Transformed, this also resulted in higher life expectancies for all groups. The mortality analysis within the Gompertz model also proved a strong compositional effect of life expectancy. The more confounders were controlled for, the more the hazards converged. Unfortunately, education or any other SES indicators were not available for this analysis. But it is likely that the remaining differentials would have dwindled further if additional strong control variables had been introduced into the model. This is specifically important since Dutch natives enjoy systematic advantages regarding SES, as was laid out in the theoretical concept. The major implication is therefore that the differences in mortality might be largely made up by sociodemographic differences instead of any major biases in registration or cultural and lifestyle factors. Thus, if the social realities of the populations analyzed in this thesis would further align, one might expect convergence of mortality patterns.

Peculiarities like the high survival times of widowed individuals and the overall high transformed life expectancy might be due to their high starting age. Most widowed people stem from older age-groups; thus, individuals must already have reached a quite high age to qualify for that group which might skew the survival time to the right. As all variables are constant, widowed individuals enter the observation with a lead in age. Same applies to the overall population, in which individuals under 15 were deleted. Similar biases have been found with the life expectancy of OSCAR winners and their non-OSCAR winning actor colleagues in which the longer survival of the winners was due to the fact that most of the winners were already older (Sylvestre, Huszti, & Hanley, 2006).

The analysis of the salmon bias delivered rather mixed results. Depending on the model health indicators pointed in different directions. What stood out was that out-migration hazards were by far the highest for Turkish individuals, both other groups did not seem very mobile at all. Moroccans even less than Surinamese, which was quite peculiar considering that latter group was the least affected by mortality abroad. Seemingly, Moroccans and Turks are leaving the Netherlands more often without deregistering which is why they appear in the period mortality data as dying abroad. The latter group however is also prone to officially out-migrate as shown by the salmon bias analysis. The reason behind that might be different legislative frameworks regarding migration between the two countries and the Netherlands. If it is easier to return to the host country for some foreign nationals, migration without registration is less likely. The cause for the Surinamese leaving the Netherlands less often might be because many of them were naturalized as Dutch citizens the moment they entered the Netherlands. The roots in the actual country might therefore be less strong as for the other groups. This is also an indication that it is useful to include citizenship as a confounder in future analyses. Also, Turkey has developed economically and became a more favorable destination to return to. The same could not be said about Suriname and neither Morocco.

Model fit measures showed that some of the models were not complying with the proportionality assumptions. Moreover, the log-survival graphs painted a rather mixed picture of the health indicators which does not leave room for confident substantial interpretations about a clear relationship of health and leaving the host country. However, especially for the Turkish population health seemingly plays a role for outmigration. A smaller timeframe could be the answer to this problem and may even lead to better fitting models. However, this would then need other health indicators than the health monitor because of the miniscule number of cases of out-migration which makes estimations through Cox regression difficult. Thus, Salmon bias analyses could be carried out with continuous predictors from the population registers to avoid the disturbance which was shown to be the case after a certain period of observation-time. This would also provide more cases and avoid representativity issues.

Conclusively, the research questions can be answered to some degree. The situation of life expectancy and mortality patterns in the Netherlands has been examined in depth and shown interesting results. A composition effect regarding sociodemographic factors was identified to be a large driver of mortality differences which answers partly the question about mechanisms behind mortality differentials. However, selection through the salmon bias still remains up for debate. Moreover, the situation regarding life expectancy is less paradoxical to begin with and the results change gravely depending on which methodological lens is applied. There are large differences between results from different models and different data sets. It remains questionable from which side life expectancy should be monitored. Whether one would like to look at the life expectancy of a specific year in a specific place or on overall differences between ethnicities. For institutions like CBS it makes sense to look at residentially restricted data as their main concern is the situation within the Netherlands. For studies looking at differences between ethnic groups, the individual data which is not restricted to the Netherlands offers additional insights. The only group that is paradoxically healthy are the Moroccans; however, only in the life table analyses. Both Surinamese and Turkish are disadvantaged. One mechanism behind mortality differences is definitely composition in terms of sociodemographic variables. With a more holistic model, this finding might be further reinforced. Hints on a salmon bias were found, as well as contradicting measurements. All in all, the empiric results do not invite to make definite inferences about the existence of a salmon bias, much less about the impact on life expectancy.

The expected migrant mortality disadvantage in older ages was also found as hypothesized. It might be the case that a healthy migrant effect would have been observable during the years in which the migrant groups first arrived and thereafter. A sign for that is the mortality advantages that prevailed for migrants throughout most of the adult-age groups (s. Figure 5). In the recent past the Netherlands have seen more and more migrants entering older age-groups with higher mortality risks (Garssen & van der Meulen, 2008). This could have lowered the overall advantage that migrants might have had in the years before. Since large parts of the migrant groups have already spent a lot of time in the Netherlands, their health behaviors might have been converged with the ones of the Dutch (s. acculturation). Together with systematic deprivation in socioeconomic resources this might have led to a loss of any possible mortality advantage. Thus, the Netherlands is a similar case like the European studies mentioned in chapter 2, that found out about health convergence and an ‘unhealthy migrant effect’ for older age groups. As this group makes up a large part of the analyzed populations it is unsurprising that the overall results in mortality hazards are lower than the ones for native Dutch. Moroccans are still advantaged in the life tables; however, in older ages they showed a similar pattern.

For future research it is important to include further variables such as occupation and education, which are strong indicators for SES-mortality differentials (Bos et al., 2004; Meara, Richards, & Cutler, 2008). This might result in further convergence of mortality risks and thus estimated life expectancies, adding to the fact that inequalities are likely due to the social composition of populations. An interesting follow-up is also applying the same research design but including second generation migrants. They are an important group for the Netherlands, especially regarding the younger age-cohorts of individuals considered ‘non-native’. This group could also serve as a basis for evaluating acculturation effects; thus, whether or not further convergence with natives happens and whether the absence of a personal migration-history proves to be a positive influence on health and mortality. Another possibility for future research is to use the survival framework to test other periods of time to find out about past developments in the Netherlands. Moreover, a Gompertz regression which is used on the residentially restricted data to compare the results to the aggregated period life tables is an interesting way to further compare the influence of methods on the life expectancy estimates. Also, in the case that other states offer similar registers as the Netherlands does, this research design could be carried over to other cases in order to find further information on why the survival model finds such differing results compared to the life tables.

The fact that the Netherlands are basing retirement age on life expectancy measurements is still questionable when considering the significant differences that are still at hand. When fact-based policies enforce these measurements as binding regulations for the individual it can be argued that members of specific groups are treated unfairly. Socioeconomically deprived populations, that are probably even receiving less overall pension due to the inability to invest in retirement plans, will have to work longer periods of time than their groups life expectancy indicates. Until further convergence of life expectancy is still overdue this policy will always favor specific groups. Morbidity and health patterns that have their roots in the overall life course are probably hard to counter-act. A practical implication of these analyses is thus to strive for societal equity in life standard from early on to prevent such mortality pitfalls for some groups of society. Rather than health-programs targeting immediate problems, mortality differentials are best counteracted with effective social policies targeting youth and adult-life.

References

Abraído-Lanza, A. F., Chao, M. T., & Flórez, K. R. (2005). Do healthy behaviors decline with greater acculturation? Implications for the Latino mortality paradox. *Social Science & Medicine (1982)*, *61*(6), 1243–1255. https://doi.org/10.1016/j.socscimed.2005.01.016

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

Akresh, I. R. (2007). Dietary assimilation and health among hispanic immigrants to the United States. *Journal of Health and Social Behavior*, *48*(4), 404–417. https://doi.org/10.1177/002214650704800405

Andreev, E., & Shkolnikov, V. (2010). *MPIDR - Spreadsheet for calculation of confidence limits for any life table or healthy-life table quantity* (Technical Report No. 5; p. 4). Retrieved from Max Planck Institute for Demographic Research website: https://www.demogr.mpg.de/en/projects_publications/publications_1904/mpidr_technical_reports/spreadsheet_for_calculation_of_confidence_limits_for_any_life_table_or_healthy_life_table_quantity_3853.htm

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

Bengtsson, T., & van Poppel, F. (2011). Socioeconomic inequalities in death from past to present: An introduction. *Explorations in Economic History*, *48*(3), 343–356. https://doi.org/10.1016/j.eeh.2011.05.004

Bos, V., Kunst, A. E., Keij-Deerenberg, I. M., Garssen, J., & Mackenbach, J. P. (2004). Ethnic inequalities in age- and cause-specific mortality in The Netherlands. *International Journal of Epidemiology*, *33*(5), 1112–1119. https://doi.org/10.1093/ije/dyh189

Buelens, B., Meijers, R., & Tennekes, M. (2013). *Weging Gezondsheidsmonitor 2012* [Codebook]. Statistics Netherlands (CBS).

CAK. (2019). CAK - Regeling onverzekerden. Retrieved July 17, 2019, from https://www.hetcak.nl/regelingen/regeling-onverzekerden

Chiang, C. L. (1968). *Introduction to stochastic processes in biostatistics*. New York: Wiley.

Cleves, M., Gould, W., Gould, W. W., Gutierrez, R., & Marchenko, Y. (2008). *An Introduction to Survival Analysis Using Stata, Second Edition*. Stata Press.

Currie, J., & Schwandt, H. (2016). Mortality Inequality: The Good News from a County-Level Approach. *Journal of Economic Perspectives*, *30*(2), 29–52. https://doi.org/10.1257/jep.30.2.29

Diaz, C. J., Koning, S. M., & Martinez-Donate, A. P. (2016). Moving Beyond Salmon Bias: Mexican Return Migration and Health Selection. *Demography*, *53*(6), 2005–2030. https://doi.org/10.1007/s13524-016-0526-2

Dragano, N., & Wahrendorf, M. (2016). A Social Inequalities Perspective on Effort-Reward Imbalance at Work. In J. Siegrist & M. Wahrendorf (Eds.), *Work Stress and Health in a Globalized Economy: The Model of Effort-Reward Imbalance* (pp. 67–85). https://doi.org/10.1007/978-3-319-32937-6_4

Elo, I. T. (2009). Social Class Differentials in Health and Mortality: Patterns and Explanations in Comparative Perspective. *Annual Review of Sociology*, *35*(1), 553–572. https://doi.org/10.1146/annurev-soc-070308-115929

European Research Council. (2019). Becoming a Minority—Amsterdam. Retrieved August 12, 2019, from BAM - Becoming a Minority website: https://bamproject.eu/nl/steden/amsterdam

EUROSTAT. (2019a, January). Migrant integration statistics—At risk of poverty and social exclusion—Statistics Explained. Retrieved August 13, 2019, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migrant_integration_statistics_-_at_risk_of_poverty_and_social_exclusion#Material_deprivation

EUROSTAT. (2019b, May 8). Migration and migrant population statistics—Statistics Explained. Retrieved July 19, 2019, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics

Garssen, J., & van der Meulen, A. (2008). *Overlijdensrisico's naar herkomstgroep: Daling en afnemende verschillen*. Retrieved from https://www.cbs.nl/nl-nl/achtergrond/2008/02/overlijdensrisico-s-naar-herkomstgroep-daling-en-afnemende-verschillen

Glaesmer, H., Wittig, U., Braehler, E., Martin, A., Mewes, R., & Rief, W. (2011). Health care utilization among first and second generation immigrants and native-born Germans: A population-based study in Germany. *International Journal of Public Health*, *56*(5), 541–548. https://doi.org/10.1007/s00038-010-0205-9

Gruer, L., Cézard, G., Clark, E., Douglas, A., Steiner, M., Millard, A., … Bhopal, R. (2016). Life expectancy of different ethnic groups using death records linked to population census data for 4.62 million people in Scotland. *J Epidemiol Community Health*, *70*(12), 1251–1254. https://doi.org/10.1136/jech-2016-207426

Guillot, M., Khlat, M., Elo, I., Solignac, M., & Wallace, M. (2018). Understanding age variations in the migrant mortality advantage: An international comparative perspective. *PLOS ONE*, *13*(6), e0199669. https://doi.org/10.1371/journal.pone.0199669

Gupta, S., de Belder, A., & Hughes, L. O. (1995). Avoiding premature coronary deaths in Asians in Britain. *BMJ : British Medical Journal*, *311*(7012), 1035–1036.

HollandZorg. (2019). Healthcare covered abroad | HollandZorg health insurance. Retrieved July 17, 2019, from https://www.hollandzorg.com/dutch-healthcare-insurance/care-country-of-residence/covered-abroad

Horvath, S., Gurven, M., Levine, M. E., Trumble, B. C., Kaplan, H., Allayee, H., … Assimes, T. L. (2016). An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biology*, *17*. https://doi.org/10.1186/s13059-016-1030-0

Ikram, U. Z., Snijder, M. B., Fassaert, T. J. L., Schene, A. H., Kunst, A. E., & Stronks, K. (2015). The contribution of perceived ethnic discrimination to the prevalence of depression. *European Journal of Public Health*, *25*(2), 243–248. https://doi.org/10.1093/eurpub/cku180

Karno, M., & Edgerton, R. B. (1969). Perception of mental illness in a Mexican-American community. *Archives of General Psychiatry*, *20*(2), 233–238.

Kohls, M. (2015). Sterberisiken von Migranten. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *58*(6), 519–526. https://doi.org/10.1007/s00103-015-2140-7

Koninkrijksrelaties, M. van B. Z. en. (2010, July 1). Basisregistratie Personen (BRP)—Privacy en persoonsgegevens—Rijksoverheid.nl [Onderwerp]. Retrieved June 7, 2019, from https://www.rijksoverheid.nl/onderwerpen/privacy-en-persoonsgegevens/basisregistratie-personen-brp

Krokstad, S., Kunst, A. E., & Westin, S. (2002). Trends in health inequalities by educational level in a Norwegian total population study. *Journal of Epidemiology & Community Health*, *56*(5), 375–380. https://doi.org/10.1136/jech.56.5.375

Lafleur, J.-M., & Stanek, M. (2017). EU Migration and the Economic Crisis: Concepts and Issues. In J.-M. Lafleur & M. Stanek (Eds.), *South-North Migration of EU Citizens in Times of Crisis* (pp. 1–14). https://doi.org/10.1007/978-3-319-39763-4_1

Lampert, T., Ziese, T., Saß, A. C., & Häfelinger, M. (2005). *Armut, soziale Ungleichheit und Gesundheit: Expertise des Robert-Koch-Instituts zum 2. Armuts- und Reichtumsbericht der Bundesregierung*. Retrieved from Bundesministerium für Gesundheit und Soziale Sicherung, Robert Koch Institute website: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-316047

Loi, S., & Hale, J. M. (2019). *Migrant health convergence and the role of material deprivation*. https://doi.org/10.4054/DemRes.2019.40.32

Mackenbach, J. P., Slobbe, L., Looman, C. W. N., van der Heide, A., Polder, J., & Garssen, J. (2011). Sharp upturn of life expectancy in the Netherlands: Effect of more health care for the elderly? *European Journal of Epidemiology*, *26*(12), 903–914. https://doi.org/10.1007/s10654-011-9633-y

Markides, K. S., & Coreil, J. (1986). The health of Hispanics in the southwestern United States: An epidemiologic paradox. *Public Health Reports*, *101*(3), 253–265.

Marmot, M. G., Adelstein, A. M., & Bulusu, L. (1984). Lessons from the study of immigrant mortality. *Lancet (London, England)*, *1*(8392), 1455–1457. https://doi.org/10.1016/s0140-6736(84)91943-3

Meara, E., Richards, S., & Cutler, D. (2008). The Gap Gets Bigger: Changes in Mortality and Life expectancy by Education, 1981−2000. *Health Affairs (Project Hope)*, *27*(2), 350–360. https://doi.org/10.1377/hlthaff.27.2.350

Missov, T. I., & Lenart, A. (2011). Linking period and cohort life-expectancy linear increases in Gompertz proportional hazards models. *Demographic Research*, *S8*(19), 455–468. https://doi.org/10.4054/DemRes.2011.24.19

Moser, A., Clough-Gorr, K., & Zwahlen, M. (2015). Modeling absolute differences in life expectancy with a censored skew-normal regression approach. *PeerJ*, *3*, e1162. https://doi.org/10.7717/peerj.1162

Norredam, M., Hansen, O. H., Petersen, J. H., Kunst, A. E., Kristiansen, M., Krasnik, A., & Agyemang, C. (2015). Remigration of migrants with severe disease: Myth or reality?—a register-based cohort study. *European Journal of Public Health*, *25*(1), 84–89. https://doi.org/10.1093/eurpub/cku138

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

Pascoe, E. A., & Richman, L. S. (2009). Perceived Discrimination and Health: A Meta-Analytic Review. *Psychological Bulletin*, *135*(4), 531–554. https://doi.org/10.1037/a0016059

Preston, S., Heuveline, P., & Guillot, M. (2000). *Demography: Measuring and Modeling Population Processes*. Wiley.

Prins, K. (2016). *Population register data, basis for the Netherlands' population statistics* [Data]. Retrieved from Statistics Netherlands (CBS) website: https://www.cbs.nl/en-gb/background/2016/01/population-register-data-basis-for-the-netherlands-population-statistics

Puschmann, P., Donrovich, R., & Matthijs, K. (2017). Salmon Bias or Red Herring? *Human Nature*, *28*(4), 481–499. https://doi.org/10.1007/s12110-017-9303-1

R: The Poisson Distribution. (2019). Retrieved August 13, 2019, from https://stat.ethz.ch/R-manual/R-patched/library/stats/html/Poisson.html

Razum, O. (2006). Commentary: Of salmon and time travellers—musing on the mystery of migrant mortality. *International Journal of Epidemiology*, *35*(4), 919–921. https://doi.org/10.1093/ije/dyl143

Razum, O., Zeeb, H., Akgün, H. S., & Yilmaz, S. (1998). Low overall mortality of Turkish residents in Germany persists and extends into a second generation: Merely a healthy migrant effect?*. *Tropical Medicine & International Health*, *3*(4), 297–303. https://doi.org/10.1046/j.1365-3156.1998.00233.x

Reus Pons, M., de Valk, H., & Janssen, F. (2018). *Healthy ageing in a comparative perspective: A study of the health of older migrants and non-migrants across Europe* (Rijksuniversiteit Groningen). Retrieved from https://pure.knaw.nl/portal/en/publications/healthy-ageing-in-a-comparative-perspective-a-study-of-the-health-of-older-migrants-and-nonmigrants-across-europe(4534828b-8dfe-46ea-a8a5-18d2f50eea1a).html

Reus-Pons, M., Vandenheede, H., Janssen, F., & Kibele, E. U. B. (2016). Differences in mortality between groups of older migrants and older non-migrants in Belgium, 2001–09. *European Journal of Public Health*, *26*(6), 992–1000. https://doi.org/10.1093/eurpub/ckw076

Riosmena, F., Kuhn, R., & Jochem, W. C. (2017). Explaining the Immigrant Health Advantage: Self-selection and Protection in Health-Related Factors Among Five Major National-Origin Immigrant Groups in the United States. *Demography*, *54*(1), 175–200. https://doi.org/10.1007/s13524-016-0542-2

RIVM. (2019). Gezondheidsmonitor Volwassenen en Ouderen | Gezondheidsmonitor. Retrieved July 19, 2019, from https://www.monitorgezondheid.nl/gezondheidsmonitor-volwassenen-en-ouderen

Robertson, H. T., Campos, G. de los, & Allison, D. B. (2013). Turning the analysis of obesity–mortality associations upside down: Modeling years of life lost through conditional distributions. *Obesity*, *21*(2), 398–404. https://doi.org/10.1002/oby.20019

Şekercan, A., Snijder, M. B., Peters, R. J. G., & Stronks, K. (2018). Is healthcare consumption in the country of origin among Moroccan and Turkish migrants of older age (55+) associated with less use of care in the Netherlands? *Tijdschrift voor gerontologie en geriatrie*, *49*(6), 253–262. https://doi.org/10.1007/s12439-018-0272-6

Sociale Verzekeringsbank (SVB). (2019). Treaty countries—General—SVB [Onderwerp]. Retrieved July 17, 2019, from https://www.svb.nl/int/en/algemeen/verdragslanden.jsp

Stark, O., & Bloom, D. E. (1985). The New Economics of Labor Migration. *The American Economic Review*, *75*(2), 173–178.

Statistics Netherlands (CBS). (2014, December 19). Increasing life expectancy to push up pension age further from 2022 [Webpagina]. Retrieved August 9, 2019, from Statistics Netherlands website: https://www.cbs.nl/en-gb/news/2014/51/increasing-life-expectancy-to-push-up-pension-age-further-from-2022

Statistics Netherlands (CBS). (2015, September 28). Gezondheidsmonitor 2012: Opbouw en instructie totaalbestand. Retrieved July 16, 2019, from https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/gezondheidsmonitor-2012-opbouw-en-instructie-totaalbestand

Statistics Netherlands (CBS). (2016). *Annual Report on Integration 2016* [Statistical Report]. Retrieved from Statistics Netherlands (CBS) website: https://www.cbs.nl/en-gb/publication/2016/47/annual-report-on-integration-2016

Statistics Netherlands (CBS). (2017a, January 3). Haalbaarheidsstudie Gewoonlijk Verblijvende Bevolking [Webpagina]. Retrieved July 16, 2019, from Centraal Bureau voor de Statistiek website: https://www.cbs.nl/nl-nl/achtergrond/2017/09/haalbaarheidsstudie-gewoonlijk-verblijvende-bevolking

Statistics Netherlands (CBS). (2017b, April 7). Levensverwachting en doodsoorzaak Surinamers [Webpagina]. Retrieved July 30, 2019, from Centraal Bureau voor de Statistiek website: https://www.cbs.nl/nl-nl/achtergrond/2017/27/levensverwachting-en-doodsoorzaak-surinamers

Statistics Netherlands (CBS). (2019a, February 28). Population [Webpagina]. Retrieved July 18, 2019, from Statistics Netherlands website: https://www.cbs.nl/en-gb/background/2018/47/population

Statistics Netherlands (CBS). (2019b, February 28). Socioeconomic status [Webpagina]. Retrieved July 5, 2019, from Statistics Netherlands website: https://www.cbs.nl/en-gb/background/2018/47/socioeconomic-status

Swerdlow, A. j. (1991). Mortality and cancer incidence in Vietnamese refugees in England and Wales: A follow-up study. *International Journal of Epidemiology*, *20*(1), 13–19. https://doi.org/10.1093/ije/20.1.13

Sylvestre, M.-P., Huszti, E., & Hanley, J. A. (2006). Do OSCAR winners live longer than less successful peers? A reanalysis of the evidence. *Annals of Internal Medicine*, *145*(5), 361–363; discussion 392. https://doi.org/10.7326/0003-4819-145-5-200609050-00009

Turra, C. M., & Elo, I. T. (2008). The Impact of Salmon Bias on the Hispanic Mortality Advantage. *Population Research and Policy Review*, *27*(5), 515–530. https://doi.org/10.1007/s11113-008-9087-4

Uitenbroek, D. G. (2015). Use of the life table to compare mortality in ethnic groups in Amsterdam, the Netherlands. *BMC Public Health*, *15*(1), 825. https://doi.org/10.1186/s12889-015-2170-y

Uitenbroek, D. G., & Verhoeff, A. P. (2002). Life expectancy and mortality differences between migrant groups living in Amsterdam, the Netherlands. *Social Science & Medicine*, *54*(9), 1379–1388. https://doi.org/10.1016/S0277-9536(01)00120-4

Wallace, M. (2019, April 23). Low migrant death rates: Actual or artefactual? - Faible mortalité des immigrés: réalité ou artéfact ? Retrieved July 5, 2019, from N-IUSSP website: http://www.niussp.org/article/low-migrant-death-rates-actual-or-artefactualfaible-mortalite-des-immigres-realite-ou-artefact/

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

Wallace, M., & Kulu, H. (2018). Can the salmon bias effect explain the migrant mortality advantage in England and Wales? *Population, Space and Place*, *24*(8), e2146. https://doi.org/10.1002/psp.2146

WHO, W. H. O. R. O. for E. (2018). *Report on the health of refugees and migrants in the WHO European Region: No public health without refugees and migrant health*. Retrieved from https://apps.who.int/iris/handle/10665/311347

Williams, D. R. (1999). Race, Socioeconomic Status, and Health The Added Effects of Racism and Discrimination. *Annals of the New York Academy of Sciences*, *896*(1), 173–188. https://doi.org/10.1111/j.1749-6632.1999.tb08114.x

Xu, J., Murphy, S., Kochanek, K., Brigham, B., & Arias, E. (2018). *Deaths: Final Data for 2016* (National Vital Statistic Report No. 5). Retrieved from U.S. Department of Health and Human Services website: https://stacks.cdc.gov/view/cdc/57989

Zorlu, A., & Hartog, J. (2001). *Migration and Immigrants: The Case of the Netherlands* (Working Paper No. 01-042/3). Retrieved from Tinbergen Institute Discussion Paper website: https://www.econstor.eu/handle/10419/85952

Appendix

I. Transformed Life Table Formulas

$$q_x = \frac{2_n m_x}{2 + \ _n m_x}$$

$$l_x = l_{x-1} - \ _n d_{x-1}$$

$$_n d_x = l_x \ _n q_x$$

$$_n L_x = l_{x+1} + \frac{_n d_x}{2}$$

$$T_x = \sum_{a=x}^{\infty} \ _n L_a$$

II. Life Tables

2010-2017 Dutch

Age x	$_nN_x$	$_nD_x$	$_nm_x$	n	$_na_x$	$_nq_x$	$_np_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0-5	5.390.323	4072	0,000755	6	2,000	0,004519	0,995481	100.000	452	598.192	8.241.477	82,4
5-10	5.777.992	418	0,000072	5	2,500	0,000362	0,999638	99.548	36	497.651	7.643.284	76,8
10-15	6.162.172	495	0,000080	5	2,500	0,000402	0,999598	99.512	40	497.461	7.145.634	71,8
15-20	6.188.429	1153	0,000186	5	2,500	0,000931	0,999069	99.472	93	497.129	6.648.173	66,8
20-25	6.164.928	1714	0,000278	5	2,500	0,001389	0,998611	99.380	138	496.552	6.151.044	61,9
25-30	5.839.753	1921	0,000329	5	2,500	0,001643	0,998357	99.241	163	495.800	5.654.492	57,0
30-35	5.698.498	2366	0,000415	5	2,500	0,002074	0,997926	99.078	205	494.878	5.158.692	52,1
35-40	6.196.807	3599	0,000581	5	2,500	0,002900	0,997100	98.873	287	493.648	4.663.814	47,2
40-45	7.501.964	7062	0,000941	5	2,500	0,004696	0,995304	98.586	463	491.774	4.170.166	42,3
45-50	8.189.978	13343	0,001629	5	2,500	0,008113	0,991887	98.123	796	488.626	3.678.392	37,5
50-55	8.082.068	22693	0,002808	5	2,500	0,013941	0,986059	97.327	1.357	483.244	3.189.766	32,8
55-60	7.533.084	35640	0,004731	5	2,500	0,023379	0,976621	95.970	2.244	474.242	2.706.522	28,2
60-65	7.207.806	53557	0,007430	5	2,500	0,036475	0,963525	93.727	3.419	460.087	2.232.280	23,8
65-70	6.408.552	76324	0,011910	5	2,500	0,057827	0,942173	90.308	5.222	438.484	1.772.193	19,6
70-75	4.728.826	92276	0,019514	5	2,500	0,093029	0,906971	85.086	7.915	405.640	1.333.709	15,7
75-80	3.655.769	119953	0,032812	5	2,500	0,151622	0,848378	77.170	11.701	356.600	928.069	12,0
80-85	2.701.487	166220	0,061529	5	2,500	0,266631	0,733369	65.470	17.456	283.707	571.469	8,7
85-90	1.614.148	188466	0,116759	5	2,500	0,451889	0,548111	48.013	21.697	185.825	287.762	6,0
90-95	640.591	143433	0,223907	5	2,500	0,717758	0,282242	26.317	18.889	84.361	101.937	3,9
95<	149.173	63041	0,422603	∞	2,366	1,000000	0,000000	7.428	7.428	17576	17.576	2,4

2010-2017 Turkish

Age x	$_nN_x$	$_nD_x$	$_nm_x$	n	$_na_x$	$_nq_x$	$_np_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0-5	4.116	1	0,000243	6	2,000	0,001456	0,998544	100.000	146	599.417	8.199.003	82,0
5-10	6.027	0	0,000000	5	2,500	0,000000	1,000000	99.854	0	499.272	7.599.586	76,1
10-15	11.162	3	0,000269	5	2,500	0,001343	0,998657	99.854	134	498.937	7.100.314	71,1
15-20	20.782	2	0,000096	5	2,500	0,000481	0,999519	99.720	48	498.481	6.601.378	66,2
20-25	47.094	7	0,000149	5	2,500	0,000743	0,999257	99.672	74	498.176	6.102.896	61,2
25-30	101.286	30	0,000296	5	2,500	0,001480	0,998520	99.598	147	497.623	5.604.720	56,3
30-35	154.582	63	0,000408	5	2,500	0,002036	0,997964	99.451	202	496.748	5.107.097	51,4
35-40	215.067	89	0,000414	5	2,500	0,002067	0,997933	99.248	205	495.729	4.610.349	46,5
40-45	256.549	175	0,000682	5	2,500	0,003405	0,996595	99.043	337	494.373	4.114.620	41,5
45-50	237.852	349	0,001467	5	2,500	0,007310	0,992690	98.706	722	491.726	3.620.246	36,7
50-55	171.276	443	0,002586	5	2,500	0,012849	0,987151	97.985	1.259	486.775	3.128.520	31,9
55-60	96.806	407	0,004204	5	2,500	0,020803	0,979197	96.725	2.012	478.597	2.641.745	27,3
60-65	72.615	505	0,006954	5	2,500	0,034178	0,965822	94.713	3.237	465.474	2.163.148	22,8
65-70	65.087	791	0,012153	5	2,500	0,058973	0,941027	91.476	5.395	443.894	1.697.674	18,6
70-75	53.756	1251	0,023272	5	2,500	0,109962	0,890038	86.082	9.466	406.744	1.253.779	14,6
75-80	28.680	1355	0,047245	5	2,500	0,211273	0,788727	76.616	16.187	342.612	847.036	11,1
80-85	9.382	754	0,080367	5	2,500	0,334605	0,665395	60.429	20.220	251.595	504.423	8,3
85-90	2.139	279	0,130435	5	2,500	0,491803	0,508197	40.209	19.775	151.608	252.828	6,3
90-95	412	64	0,155340	5	2,500	0,559441	0,440559	20.434	11.432	73.592	101.220	5,0
95<	89	29	0,325843	∞	3,069	1,000000	0,000000	9.002	9.002	27628	27.628	3,1

2010-2017 Moroccan

Age x	$_nN_x$	$_nD_x$	$_nm_x$	n	$_na_x$	$_nq_x$	$_np_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0-5	2.287	2	0,000875	6	2,000	0,005229	0,994771	100.000	523	597.908	8.369.569	83,7
5-10	4.794	0	0,000000	5	2,500	0,000000	1,000000	99.477	0	497.386	7.771.661	78,1
10-15	9.367	2	0,000214	5	2,500	0,001067	0,998933	99.477	106	497.120	7.274.275	73,1
15-20	18.121	2	0,000110	5	2,500	0,000552	0,999448	99.371	55	496.718	6.777.155	68,2
20-25	39.981	11	0,000275	5	2,500	0,001375	0,998625	99.316	137	496.239	6.280.437	63,2
25-30	89.042	23	0,000258	5	2,500	0,001291	0,998709	99.180	128	495.578	5.784.198	58,3
30-35	150.395	64	0,000426	5	2,500	0,002125	0,997875	99.052	211	494.732	5.288.620	53,4
35-40	199.444	102	0,000511	5	2,500	0,002554	0,997446	98.841	252	493.574	4.793.888	48,5
40-45	210.493	166	0,000789	5	2,500	0,003935	0,996065	98.589	388	491.973	4.300.314	43,6
45-50	178.919	242	0,001353	5	2,500	0,006740	0,993260	98.201	662	489.349	3.808.340	38,8
50-55	131.140	274	0,002089	5	2,500	0,010393	0,989607	97.539	1.014	485.160	3.318.991	34,0
55-60	88.080	312	0,003542	5	2,500	0,017556	0,982444	96.525	1.695	478.389	2.833.832	29,4
60-65	67.297	421	0,006256	5	2,500	0,030798	0,969202	94.831	2.921	466.851	2.355.442	24,8
65-70	58.731	525	0,008939	5	2,500	0,043718	0,956282	91.910	4.018	449.505	1.888.591	20,5
70-75	53.732	896	0,016675	5	2,500	0,080040	0,919960	87.892	7.035	421.872	1.439.086	16,4
75-80	30.998	1101	0,035518	5	2,500	0,163109	0,836891	80.857	13.188	371.314	1.017.214	12,6
80-85	9.537	635	0,066583	5	2,500	0,285406	0,714594	67.669	19.313	290.060	645.901	9,5
85-90	2.077	216	0,103996	5	2,500	0,412686	0,587314	48.356	19.956	191.888	355.840	7,4
90-95	379	55	0,145119	5	2,500	0,532430	0,467570	28.400	15.121	104.197	163.952	5,8
95<	72	16	0,222222	∞	4,500	1,000000	0,000000	13.279	13.279	59755	59.755	4,5

2010-2017 Surinamese

Age x	${}_nN_x$	${}_nD_x$	${}_nm_x$	n	${}_na_x$	${}_nq_x$	${}_np_x$	l_x	${}_nd_x$	${}_nL_x$	T_x	e_x
0-5	1.515	1	0,000660	6	2,000	0,003950	0,996050	100.000	395	598.420	8.160.861	81,6
5-10	4.181	1	0,000239	5	2,500	0,001195	0,998805	99.605	119	497.727	7.562.441	75,9
10-15	10.369	0	0,000000	5	2,500	0,000000	1,000000	99.486	0	497.430	7.064.714	71,0
15-20	21.270	8	0,000376	5	2,500	0,001879	0,998121	99.486	187	496.963	6.567.284	66,0
20-25	40.290	7	0,000174	5	2,500	0,000868	0,999132	99.299	86	496.280	6.070.321	61,1
25-30	60.671	26	0,000429	5	2,500	0,002140	0,997860	99.213	212	495.533	5.574.042	56,2
30-35	81.511	47	0,000577	5	2,500	0,002879	0,997121	99.000	285	494.290	5.078.508	51,3
35-40	130.307	93	0,000714	5	2,500	0,003562	0,996438	98.715	352	492.698	4.584.219	46,4
40-45	188.993	233	0,001233	5	2,500	0,006145	0,993855	98.364	604	490.308	4.091.520	41,6
45-50	210.500	392	0,001862	5	2,500	0,009268	0,990732	97.759	906	486.532	3.601.213	36,8
50-55	206.233	692	0,003355	5	2,500	0,016638	0,983362	96.853	1.611	480.238	3.114.681	32,2
55-60	174.591	927	0,005310	5	2,500	0,026200	0,973800	95.242	2.495	469.971	2.634.443	27,7
60-65	125.570	1116	0,008887	5	2,500	0,043471	0,956529	92.747	4.032	453.653	2.164.472	23,3
65-70	81.395	1134	0,013932	5	2,500	0,067316	0,932684	88.715	5.972	428.644	1.710.819	19,3
70-75	51.774	1168	0,022560	5	2,500	0,106776	0,893224	82.743	8.835	391.627	1.282.175	15,5
75-80	31.953	1179	0,036898	5	2,500	0,168909	0,831091	73.908	12.484	338.330	890.548	12,0
80-85	17.316	1135	0,065546	5	2,500	0,281589	0,718411	61.424	17.296	263.880	552.218	9,0
85-90	8.212	891	0,108500	5	2,500	0,426745	0,573255	44.128	18.831	173.561	288.338	6,5
90-95	2.806	539	0,192088	5	2,500	0,648850	0,351150	25.297	16.414	85.448	114.777	4,5
95<	799	242	0,302879	∞	3,302	1,000000	0,000000	8.883	8.883	29328	29.328	3,3

III. Life Expectancy – Confidence Intervals

95%-Confidence 2010-2017	Surinamese				Moroccan				Turkish				Dutch			
Ages	CI-	e_x	CI+	CI-Width	CI-	e_x	CI+	CI-Width	CI-	e_x	CI+	CI-Width	CI-	e_x	CI+	CI-Width
0-5	81,05	81,66	82,11	1,05	83,12	83,70	84,34	1,22	81,68	81,99	82,31	0,63	82,46	82,41	82,50	0,05
5-10	75,64	75,92	76,18	0,54	77,80	78,13	78,48	0,68	75,85	76,11	76,40	0,55	76,76	76,78	76,80	0,04
10-15	70,79	71,01	71,23	0,43	72,82	73,13	73,49	0,67	70,84	71,11	71,39	0,55	71,78	71,81	71,83	0,04
15-20	65,80	66,01	66,22	0,42	67,87	68,20	68,54	0,67	65,94	66,20	66,44	0,51	66,81	66,83	66,86	0,04
20-25	60,93	61,13	61,33	0,40	62,90	63,24	63,57	0,66	60,98	61,23	61,48	0,50	61,87	61,89	61,91	0,04
25-30	55,98	56,18	56,39	0,41	57,99	58,32	58,66	0,67	56,01	56,27	56,52	0,51	56,96	56,98	57,00	0,04
30-35	51,10	51,30	51,49	0,39	53,09	53,39	53,73	0,65	51,11	51,35	51,60	0,49	52,05	52,07	52,09	0,04
35-40	46,26	46,44	46,62	0,36	48,18	48,50	48,84	0,66	46,21	46,45	46,69	0,48	47,15	47,17	47,19	0,04
40-45	41,40	41,60	41,77	0,37	43,29	43,62	43,97	0,69	41,31	41,54	41,80	0,49	42,28	42,30	42,32	0,04
45-50	36,65	36,84	37,02	0,36	38,46	38,78	39,13	0,67	36,45	36,68	36,93	0,48	37,47	37,49	37,51	0,04
50-55	31,98	32,16	32,35	0,37	33,70	34,03	34,38	0,68	31,68	31,93	32,17	0,49	32,76	32,77	32,79	0,04
55-60	27,49	27,66	27,85	0,36	29,06	29,36	29,75	0,69	27,08	27,31	27,57	0,49	28,18	28,20	28,22	0,03
60-65	23,16	23,34	23,53	0,37	24,53	24,84	25,17	0,64	22,60	22,84	23,09	0,48	23,80	23,82	23,83	0,03
65-70	19,10	19,28	19,46	0,36	20,22	20,55	20,88	0,66	18,31	18,56	18,79	0,48	19,61	19,62	19,64	0,03
70-75	15,31	15,50	15,69	0,39	16,04	16,37	16,72	0,68	14,31	14,57	14,83	0,52	15,66	15,67	15,69	0,03
75-80	11,87	12,05	12,23	0,35	12,22	12,58	13,00	0,78	10,78	11,06	11,33	0,55	12,01	12,03	12,04	0,03
80-85	8,80	8,99	9,17	0,36	9,13	9,55	9,98	0,85	8,03	8,35	8,65	0,62	8,72	8,73	8,74	0,02
85-90	6,36	6,53	6,73	0,37	6,85	7,36	7,92	1,07	5,92	6,29	6,73	0,81	5,98	5,99	6,00	0,02
90-95	4,34	4,54	4,74	0,40	5,16	5,77	6,45	1,29	4,47	4,95	5,49	1,02	3,86	3,87	3,88	0,02
95<	-	3,30	-	-	-	4,50	-	-	-	3,07	-	-	-	2,37	-	-

IV. Age-Specific Death Rates – Confidence Intervals

95%-Confidence

2010-2017	**Dutch**				**Surinamese**				**Turkish**				**Moroccan**			
Ages	CI+	$_nm_x$	CI-	CI-Width	CI+	$_nm_x$	CI-	CI-Width	CI+	$_nm_x$	CI-	CI-Width	CI+	$_nm_x$	CI-	CI-Width
0-5	0,001	0,001	0,001	0,000	0,002	0,001	0,000	0,002	0,001	0,000	0,000	0,001	0,002	0,001	0,000	0,002
5-10	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
10-15	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,001	0,001	0,000	0,000	0,001
15-20	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
20-25	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
25-30	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
30-35	0,000	0,000	0,000	0,000	0,001	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,001	0,000	0,000	0,000
35-40	0,001	0,001	0,001	0,000	0,001	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,001	0,001	0,000	0,000
40-45	0,001	0,001	0,001	0,000	0,001	0,001	0,001	0,000	0,001	0,001	0,001	0,000	0,001	0,001	0,001	0,000
45-50	0,002	0,002	0,002	0,000	0,002	0,002	0,002	0,000	0,002	0,001	0,001	0,000	0,002	0,001	0,001	0,000
50-55	0,003	0,003	0,003	0,000	0,004	0,003	0,003	0,000	0,003	0,003	0,002	0,000	0,002	0,002	0,002	0,000
55-60	0,005	0,005	0,005	0,000	0,006	0,005	0,005	0,001	0,005	0,004	0,004	0,001	0,004	0,004	0,003	0,001
60-65	0,007	0,007	0,007	0,000	0,009	0,009	0,008	0,001	0,008	0,007	0,006	0,001	0,007	0,006	0,006	0,001
65-70	0,012	0,012	0,012	0,000	0,015	0,014	0,013	0,002	0,013	0,012	0,011	0,002	0,010	0,009	0,008	0,002
70-75	0,020	0,020	0,019	0,000	0,024	0,023	0,021	0,003	0,025	0,023	0,022	0,003	0,018	0,017	0,016	0,002
75-80	0,033	0,033	0,033	0,000	0,039	0,037	0,035	0,004	0,050	0,047	0,045	0,005	0,038	0,036	0,033	0,004
80-85	0,062	0,062	0,061	0,001	0,069	0,066	0,062	0,008	0,086	0,080	0,075	0,011	0,072	0,067	0,061	0,010
85-90	0,117	0,117	0,116	0,001	0,116	0,108	0,101	0,014	0,146	0,130	0,115	0,030	0,118	0,104	0,091	0,027
90-95	0,225	0,224	0,223	0,002	0,208	0,192	0,176	0,032	0,194	0,155	0,119	0,075	0,185	0,145	0,108	0,077
95<	0,426	0,423	0,419	0,007	0,342	0,303	0,265	0,076	0,449	0,326	0,213	0,236	0,333	0,222	0,125	0,208

V. Age and Cause-specific Decomposition of Differences in Life Expectancy

	Turkish-Dutch						
	Age-Decomposition				**Cause-Decomposition**		
Age	Direct Effect	Indirect+Interaction	Total Effect	Total Effect	Total Effect CVD	Total Effect Cancer	Total Effect Abroad
0-5	0,01232689	0,234539736	0,24686663		0,005808842	-0,10220007	0,002323537
5-10	0,00090566	0,025759499	0,02666516		0,001339637	0,0090585	0,00057413
10-15	-0,00235659	-0,062401469	-0,06475806		-0,029393563	0,008644066	0,000557682
15-20	0,00112623	0,027583521	0,02870975		0,003450671	0,009424967	-0,013739763
20-25	0,00161559	0,036365914	0,03798151		-0,001376448	0,010666128	-0,008466893
25-30	0,00040831	0,008387148	0,00879546		-0,004580029	0,005717926	-0,006832707
30-35	9,5095E-05	0,001766968	0,00186206		-0,003793732	0,001594556	-0,02038241
35-40	0,00207119	0,034417897	0,03648909		0,001855569	0,020174932	-0,01171621
40-45	0,00320141	0,046967306	0,05016872		-0,001365106	0,03066144	-0,013719454
45-50	0,00198263	0,025321177	0,02730381		-0,007541579	0,033704521	-0,030212292
50-55	0,00267359	0,029208166	0,03188175		0,006051538	0,070285803	-0,07261764
55-60	0,00621993	0,056822542	0,06304247		-0,00359183	0,118216833	-0,106330445
60-65	0,00541428	0,040192631	0,04560691		0,02015015	0,144942619	-0,149657534
65-70	-0,00260411	-0,015171536	-0,01777564		-0,013637146	0,192126445	-0,244007293
70-75	-0,03624243	-0,160272904	-0,19651533		-0,019569623	0,157058977	-0,316608751
75-80	-0,11580027	-0,386651087	-0,50245136		-0,01443428	0,103715619	-0,426781072
80-85	-0,11195046	-0,28156991	-0,39352037		-0,035263107	0,106958865	-0,401363417
85-90	-0,04820957	-0,095521673	-0,14373124		-0,086148151	0,079056411	-0,224362626
90-95	0,10481002	0,128663341	0,23347337		-0,051862477	0,047396808	-0,098590508
95<	0,05251851	-	0,05251851	e20(2)-e20(1)=	0,003682132	-0,00357776	-0,024273976
			-0,664870	**-0,664870**	**-0,211424119**	**1,118700122**	**-2,155923227**

	Moroccan-Dutch						
	Age-Decomposition				**Cause-Decomposition**		
Age	Direct Effect	Indirect+Interaction	Total Effect		Total Effect CVD	Total Effect Cancer	Total Effect Abroad
0-5	-0,00285711	-0,055803043	-0,05866015		0,005940213	0,015170389	-0,213019732
5-10	0,00090566	0,0264907	0,02739636		0,001376372	0,009306899	0,000589874
10-15	-0,00166583	-0,045444235	-0,04711007		0,001435023	0,008897146	0,000574009
15-20	0,00094951	0,024017565	0,02496707		0,003559226	-0,008420275	0,001646806
20-25	3,6121E-05	0,000842628	0,00087875		0,005024481	0,003438497	0,004137808
25-30	0,00088057	0,018806355	0,01968692		0,006251148	0,010680944	-0,014864633
30-35	-0,0001287	-0,002496811	-0,00262551		0,005757776	-0,000802422	-0,006737751
35-40	0,00086024	0,01500912	0,01586936		0,006511641	-0,006367051	-0,013515203
40-45	0,0018857	0,029251951	0,03113766		0,016615602	0,01826579	-0,014751278
45-50	0,00338872	0,046123728	0,04951245		0,023435193	0,03482417	-0,026697298
50-55	0,00868848	0,102032276	0,11072076		0,038152819	0,045741298	-0,031000442
55-60	0,01405912	0,139682653	0,15374178		0,027975405	0,106713095	-0,064086297
60-65	0,01338506	0,110015892	0,12340095		0,022083568	0,153330267	-0,101530872
65-70	0,03205163	0,209917358	0,24196899		0,061766536	0,256169547	-0,13921192
70-75	0,02780245	0,13990654	0,16770899		0,061337326	0,223509192	-0,183106238
75-80	-0,02229856	-0,085136526	-0,10743509		0,045751519	0,171024074	-0,296042193
80-85	-0,03092109	-0,091017368	-0,12193846		0,058375358	0,176624969	-0,362364995
85-90	0,04735012	0,109340728	0,15669085		0,14839935	0,075047178	-0,340909935
90-95	0,12269177	0,220845192	0,34353697		0,153515836	0,044581229	-0,206333862
95<	0,15947443	-	0,15947443	e20(2)-e20(1)=	0,06166042	0,021580558	-0,055091703
			1,342330	**1,342330**	**0,742613978**	**1,334361335**	**-1,852106812**

Age	Surinamese-Dutch						
	Age-Decomposition				**Cause-Decomposition**		
Age	Direct Effect	Indirect+Interaction	Total Effect		Total Effect CVD	Total Effect Cancer	Total Effect Abroad
0-5	0,00228999	0,043466458	0,04575645		0,005785972	0,014776482	-0,314398047
5-10	-0,00208733	-0,059290439	-0,06137777		0,001337117	-0,078951316	0,00057305
10-15	0,00100525	0,026543439	0,02754869		0,001391348	0,008626356	0,000556539
15-20	-0,00237139	-0,05798686	-0,06035825		0,003442962	-0,020498096	-0,013357991
20-25	0,00130208	0,0292617	0,03056378		-0,002425212	0,010649018	-0,003280936
25-30	-0,00124078	-0,02545966	-0,02670044		-0,007242807	0,011879281	-0,013911473
30-35	-0,00200656	-0,037272821	-0,03927938		-0,004550273	0,010442621	-0,017652763
35-40	-0,00164765	-0,027414063	-0,02906171		-0,020825188	0,004006774	-0,004861302
40-45	-0,00359505	-0,052973125	-0,05656818		-0,017588798	0,010359517	-0,008951056
45-50	-0,00285131	-0,036677887	-0,0395292		-0,018933836	0,027647903	-0,019800006
50-55	-0,00660162	-0,073041786	-0,07964341		-0,046981218	0,042693811	-0,028623984
55-60	-0,00681023	-0,063573425	-0,07038365		-0,049592591	0,105410083	-0,019804082
60-65	-0,01649739	-0,127257552	-0,14375495		-0,094124894	0,120254539	-0,038609276
65-70	-0,0215568	-0,133616825	-0,15517363		-0,111090823	0,18230822	-0,063785423
70-75	-0,02942366	-0,141815305	-0,17123897		-0,091087549	0,187459957	-0,081308599
75-80	-0,03355824	-0,120678491	-0,15423673		-0,119055527	0,192504461	-0,053155532
80-85	-0,02463419	-0,064385313	-0,08901951		-0,020299621	0,14990252	-0,077783395
85-90	0,03037012	0,055118717	0,08548884		0,032475933	0,096584095	-0,052571114
90-95	0,04561863	0,060246748	0,10586537		0,031357121	0,042818702	-0,018400144
95<	0,06990963	-	0,06990963	e20(2)-e20(1)=	0,030837176	0,005602282	-0,007909773
			-0,762762	**-0,762762**	**-0,509128107**	**1,200523785**	**-0,510408857**

VI. Transformed Gompertz Regression Output

Gompertz regression – Transformed
No. of subjects = 1354073
No. of observations = 11354073
No. of failures = 1131600

_t	Model 1	Model 2*	Model 3**	Model 4***
	Life Expectancy	**Life Expectancy**	**Life Expectancy**	**Life Expectancy**
Dutch (Ref.)	68,3	66,3	67,2	64,0
Turkish	-3,0	-2,7	-2,6	-2,2
Moroccan	-1,2	-0,6	-0,4	-0,0
Surinamese	-1,8	-1,9	-2,1	-0,6
Sex (Ref.: Male)				
Female	-	+3,3	+2,8	+2,4
Cohabitation (Ref.: No)				
Yes	-	-	-1,5	+0,8
Marital Status (Ref.: Single)				
married	-	-	-	+0,8
seperated	-	-	-	-0,5
widowed	-	-	-	+4,7
Intercept	5,76E-07	5,98E-07	4,39E-07	5,13E-07
Gamma	*0,1342436*	*0,1370879*	*0,1395441*	*0,1433329*