Mortality and Longevity

Modeling and Forecasting Cause-of-Death Mortality by Socioeconomic Factors
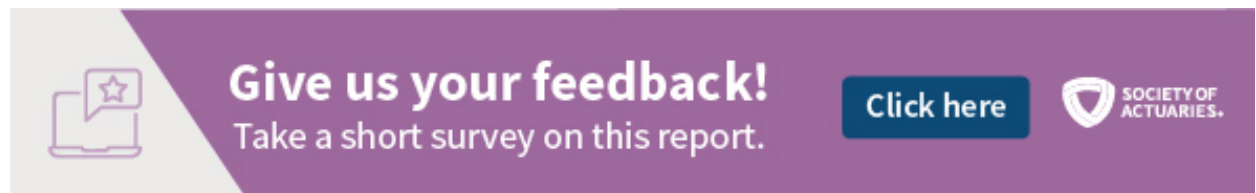
June 2021

Modeling and Forecasting Cause-of-Death Mortality by Socioeconomic Factors

AUTHOR	Alexandre Boumezoued Jean-Baptiste Coulomb Al Klein, FSA, MAAA Damien Louvet Eve Elisabeth Titon Milliman	**SPONSOR**	SOA Research Expanding Boundaries Pool

Give us your feedback!
Take a short survey on this report.

Click here

SOCIETY OF ACTUARIES.

CONTENTS

Modeling and Forecasting Cause-of-Death Mortality by Socioeconomic Factors

Executive Summary

A first output of this report is to analyze mortality by socioeconomic factors and thus to better predict deaths by cause for a population characterized with high socioeconomic characteristics and which therefore differs from the general population.

This report provides a toolkit for actuaries and other interested parties to model and forecast mortality by cause and socioeconomic factors. The benefits of the work output are fourfold:

- First, it relies on a unique redesign of the International Classification of Diseases (ICD) coding, allowing the user of the tool to perform a more precise analysis of the role of drug- and smoking-related deaths in particular.

- Second, the modeling framework using the Lee-Carter formulation as a building block is robust and allows the user to analyze, not only the main parameters for each cause, but also the cross-correlations between the time series underlying the evolution of the cause-specific rates.

- Third, the model developed allows the user to project mortality for a cluster population in reference to the general population. As such, it also allows for comparing the national and cluster population cause-of-death mortality evolutions. The population retained within the richest cluster contains 24,432,088 exposures and 110,576 deaths in 2016, which provides sufficient data for both inference and forecasts.

- Finally, the report provides a detailed comparison to aggregate forecasts, which does not rely on cause-of-death information.

The results of this study confirm that there are differences between a national and a richest cluster population, in terms of both mortality rate at the different ages and life expectancy. Overall, as expected the richest cluster shows lower mortality, however in terms of death rates, these differences depend on the cause-of-death and the ages considered. Thus, for drug deaths at young ages and dementia deaths at old ages, the gap is expected to widen with time; this is driven by a difference in the initial level of the death rates, to which improvement rates of similar order of magnitude are applied. The evolution of the death rates for a same cause between the cluster and the national population are similar both because the alpha coefficients (see Table 8) are close to 1 and because the national population trends are close to 0. These national trends are small because we observe minor evolution of the death rates during the last few years and because the model used to forecast the national death rates captures the most recent historical trends (see previous SOA project "Modeling and Forecasting Cause-of-Death Mortality"). The absolute increases and the final gaps are different because of the difference of the initial levels, noticing overall that the relational model used to link the cluster population with the national population remains vulnerable to the fact that the national population trends are small and that the alpha parameters are close to 1.

Therefore, the historical levels of death rates are different between the cluster population and the national population but the forecasts are rather similar for most causes-of-death.

In terms of life expectancy, the gap between the national and the cluster population remains stable over the years as the evolution of the underlying causes offset, which confirms the utility of modeling specific death rates to compare both populations. Note that the projections in this report are only available for short-term forecasts since the projections are based on the most recent historical trends and thus do not consider possible changes in future trends that may happen (see previous SOA project "Modeling and Forecasting Cause-of-Death Mortality"). The cause-of-death findings in terms of expected forecast based on the modelling framework developed in this study are described below in more details.

In terms of distribution of death rates by age ranges for the richest cluster of counties, the drug cause at ages 30–34 is the top cause-of-death in 2017 for both males and females and the related death rate is expected to increase. Note that this may appear counterintuitive to some actuaries if they have identified that insured portfolios are not impacted by the opioid crisis in particular. However, this is first due to the fact that the national population is segmented based on economic information at county level in this study, whereas more granular information (involving lifestyle) could lead to a different clustering, therefore different conclusions. Moreover, it is worth mentioning that the life insurance underwriting process is able to eliminate some of those more at risk for drug abuse based on such lifestyle factors. As such, a proxy of the insured population based on counties always include a basis risk. In this report, we also provide an additional analysis of the most recent data from the Centers for Disease Control and Prevention (CDC), showing a stabilization of the number of deaths because of drug overdose (see Appendix E).

As for the other age groups from the best socioeconomic cluster, several conclusions have been derived in this report. At ages 50–54, cardiovascular was the top cause-of-death for males in 2017 but is expected to become the second cause-of-death after 2023, being replaced by drug. For the female population, neoplasm was the top cause-of-death in 2017 and it is expected to remain the top cause in the future. At ages 80–84, cardiovascular was the top cause-of-death for both males and females in 2017 but its death rate is expected to decrease with time according to the model developed, although this decrease would be smaller than in the past (see Sections 6 and 8).The death rate for the cause of neoplasm is expected to decrease for both males and females. However, neoplasm would remain the second leading cause-of-death for males through 2026. For females, neoplasm would move from the second to third leading cause-of-death in 2025, with dementia becoming the second leading cause-of-death.

In terms of evolution of life expectancy for the richest cluster population, on one hand, the partial life expectancy between ages 65 and 85 and the total life expectancy at 65 are both expected to increase. On the other hand, the total life expectancy at birth is expected to remain stable because of the increase of the death rates due to the drug deaths at young ages. Recall that these forecasts have been obtained using the modelling framework developed in this study. Note in particular that the scenario of an increase in drug-related deaths may not occur if the death rate for drugs stops its exponential increase during the next or future years, see further analysis provided in appendix related to the recent evolution of drug-related deaths. Besides, the authors have not included in their study the lifestyle variables of the cluster population, which may imply some differences between the mortality rates produced and the mortality rates of a true insured population.

Section 1: Introduction

To date, aggregate mortality tables of general populations have been used for providing both historical mortality analysis and future scenarios based on appropriate forecasting tools. For such data, the Human Mortality Database (HMD) has become one of the primary reference providers of mortality estimates since its launch in 2002.

In February 2017, the Society of Actuaries (SOA) provided support to the HMD to expand the database by including cause-of-death information for a set of countries. Beyond the World Health Organization data by causes-of-death, the release of more homogeneous and user-friendly data on cause-of-death mortality rates, therefore, opens the way for the profession to analyze and measure mortality and longevity risks at a more granular level.

In this context, a first cause-of-death modeling project has previously been completed by the authors and published by the SOA in December 2019, "Modeling and Forecasting Cause-of-Death Mortality". The main outputs of the previous project were the redesign of 11 causes-of-death that differ from the usual retained lists for cause-of-death modeling, along with a modeling framework allowing for the ability to provide forecasts of age and gender specific cause-of-death mortality rates in the future.

Following this project, it was decided to have a second phase which would include socioeconomic factors such as income and geography, among others, in order to better predict cause-of-death mortality and to provide estimated forecasts closer to an insured population. This report details such study involving cause-of-death mortality estimates at a lower level of information than the split by age and gender. As a core building block, the data we propose to use are those of the National Vital Statistics System which is part of the Centers for Disease Control and Prevention, and which will allow us to access cause-of-death information within each county.

As such, one of the key objectives of this project is to provide a reference case study for the forecasting of cause-of-death mortality at a refined granularity, by using county level information. The resulting projections could be used to compare against experts' opinions on advancements and deteriorations in mortality that may be cause-of-death related at a county level, or more specifically insured population forecasts for the several causes-of-death at stake.

The focus of this report is United States mortality, although the methodology described in general could be extended to other countries. This report is organized as follows:

- Section 2: provides a literature review of the studies linking cause-of-death mortality with socioeconomic and other individual characteristics.
- Section 3: contains a detailed definition of cause-of-death rates, as well as a description of supporting data used in this study.
- Section 4: describes the clustering methodology used to identify the "richest" U.S. counties.
- Section 5: analyzes the historical pattern of U.S. mortality by cause-of-death according to the classification chosen in this project, and differentiated between the national evolution and that of the richest counties.
- Section 6: discusses the modeling framework used.
- Section 7: contains the forecasts for cause-of-death rates, as well as a comparison to the all-cause projection.
- Section 8: addresses the comparison between the cause-specific forecast and the aggregate (all causes) forecast focusing on partial and total life expectancy.
- Finally, acknowledgments, appendices and references are detailed in the corresponding sections.

This report is published with an additional tool that aims at providing an easy-to-use cause-of-death forecasting framework. The model implemented in the tool relies on the same assumptions as the by-cause model presented in this report. However, the tool also offers the option for a user to input external opinions about the future pattern of causes-of-deaths, see Appendix G. The authors encourage the readers to familiarize themselves with the cause-of-death framework through this tool.

Section 2: Literature Review

Several studies on U.S. mortality have attempted to link cause-of-death mortality with socioeconomic factors. This section presents the main references in this area which have inspired the present study.

Lourés and Cairns (2019) exploited data from multiple public sources, including highly detailed cause-of-death data from the United States Centers for Disease Control and Prevention, to explore the mortality gap between the better and worse off in the U.S. during the period 1989–2015, using education as a proxy. They used death rates for the U.S. population born between 1914 and 1970 for years 1989–2015, separated by educational attainment and cause-of-death. They found that there was a gap in all-cause mortality between different education groups, at all ages analyzed. The gap was wider at younger ages, meaning that education plays a bigger role in early mortality than it does at higher ages. They also found that for both genders in the period 1989–2015, there had been an increase in the mortality gap between the two education groups. This was mostly driven by a stagnation in the death rates of the lower educated group, which had been almost constant over the period of the analysis, while a mortality improvement was clearly seen in the higher educated population.

Mortality differentials have also been studied in the light of income and geography, as underlined by Chetty, Stepner, Abraham et al (2016). They tried in their paper to measure the level, time trend, and geographic variability in the association between income and life expectancy, and they identified factors related to small area variation. Income data for the U.S. population were obtained from 1.4 billion deidentified tax records between 1999 and 2014. Mortality data were obtained from Social Security Administration death records. These data were used to estimate race and ethnicity-adjusted life expectancy at 40 years of age by household income percentile, sex, and geographic area, and to evaluate factors associated with differences in life expectancy. Pretax household earnings were used as a measure of income. The results of their paper is that higher income was associated with greater longevity throughout the income distribution. The gap in life expectancy between the richest 1% and poorest 1% of individuals was 14.6 years for males and 10.1 years for females. Moreover, inequality in life expectancy increased over time. Between 2001 and 2014, life expectancy increased by 2.34 years for males and 2.91 years for females in the top 5% of the income distribution, but by only 0.32 years for males and 0.04 years for females in the bottom 5%. Furthermore, life expectancy for low-income individuals varied substantially across local areas. In the bottom income quartile, life expectancy differed by approximately 4.5 years between areas with the highest and lowest longevity. Changes in life expectancy between 2001 and 2014 ranged from gains of more than 4 years to losses of more than 2 years across areas. Finally, geographic differences in life expectancy for individuals in the lowest income quartile were significantly correlated with health behaviors such as smoking, but were not significantly correlated with access to medical care, physical environmental factors, income inequality, or labor market conditions. Life expectancy for low-income individuals was positively correlated with the local area fraction of immigrants, fraction of college graduates, and government expenditures.

Mokdad, Ballestros, Echko et al (2018) began their paper with the observation that several studies have measured health outcomes in the United States, but none have provided a comprehensive assessment of patterns of health by state. The aim of their study was to use the results of the Global Burden of Disease

(GBD) Study to report trends in the burden of diseases, injuries, and risk factors at the state level from 1990 to 2016. In order to achieve this aim, a systematic analysis of published studies and available data sources estimated the burden of disease by age, sex, geography, and year. Prevalence, incidence, mortality, life expectancy, healthy life expectancy (HALE), years of life lost (YLLs) due to premature mortality, years lived with disability (YLDs), and disability-adjusted life-years (DALYs) for 333 causes and 84 risk factors with 95% uncertainty intervals (UIs) were computed. The results were that between 1990 and 2016, overall death rates in the United States declined from 745.2 per 100,000 persons to 578.0 per 100,000 persons. The probability of death among adults aged 20 to 55 years declined in 31 states from 1990 to 2016. In 2016, Hawaii had the highest life expectancy at birth (81.3 years) and Mississippi had the lowest (74.7 years), a 6.6-year difference. Minnesota had the highest HALE at birth (70.3 years), and West Virginia had the lowest (63.8 years), a 6.5-year difference. The leading causes of DALYs in the United States for 1990 and 2016 were ischemic heart disease and lung cancer, while the third leading cause in 1990 was low back pain, and the third leading cause in 2016 was chronic obstructive pulmonary disease. One notable fact was that opioid use disorders moved from the 11th leading cause of DALYs in 1990 to the 7th leading cause in 2016, representing a 74.5% change. In 2016, each of the following six risks individually accounted for more than 5% of risk-attributable DALYs: tobacco consumption, high body mass index (BMI), poor diet, alcohol and drug use, high fasting plasma glucose, and high blood pressure. Across all US states, the top risk factors in terms of attributable DALYs were due to 1 of the 3 following causes: tobacco consumption (32 states), high BMI (10 states), or alcohol and drug use (8 states).

As a major cause-of-death, Drug abuse has dramatically increased in the recent years in the United States. According to the Society of Actuaries (2019), the total economic burden of the opioid crisis in the U.S. from 2015 to 2018 was at least 631 billion dollars. 205 billion dollars of this total was attributable to excess health care spending for individuals and 253 billion dollars was attributable to mortality costs (driven by lost lifetime earnings). Lost productivity cost 96 billion dollars, criminal justice activities was 39 billion dollars and the rest (government-funded child, family assistance programs and education programs) was another 39 billion dollars. Many sectors are impacted. 29% of the total economic burden of the opioid crisis would be borne by federal, state and local governments, while the remainder would be borne by the private sector and individuals. The paper of the SOA projects the costs for 2019, and finds an increase in costs, especially for health care, mortality and lost productivity costs. The midpoint of the total cost in 2019 is estimated at 188 billion dollars, with low and high cost estimates ranging from 172 billion dollars to 214 billion dollars.

One of the main interests in the present research is to quantity mortality differentials related to drug abuse in terms of level and trend for the richest counties compared to the general population, as well as to anticipate its evolution through forecasts based on appropriate projection models.

Note that the Society of Actuaries (2019) has already provided U.S. population mortality observations studying the evolution of death rates by cause-of-death and by income quartile. The results of the study were that there was a more marked improvement for the most favored categories, with some exceptions for some causes: "Relative to All Counties' mortality, the Top 15% generally decreased over 1999–2017 while the Bottom 15% increased. Exceptions to this among the ten key causes-of-death, were increases from 1999 to 2017 in the Top 15% for Alzheimer's - dementia and accidents and a decrease for accidents in the Bottom 15%. No cause-of-death was higher in the Top 15% than the Bottom 15% in 2017." Moreover, the study found that there was an increase in inequalities in terms of life expectancy over the last years.

The present report will look at the differences between the national population and the richest counties, which will also allow to further study these inequalities by cause-of-death.

Section 3: Causes-of-Death Definition, Classification and Data Review

The purpose of this section is to introduce the general modeling assumptions used for this project. The authors introduce the standardized notion of a cause-of-death and the specific classifications of causes of death used. Then, they present the geographical level, the age groups and the database used for the study. Finally, they explain the calculation of the death rates.

3.1 CAUSE-OF-DEATH DEFINITION

The cause-of-death definition is the same as defined in the previous SOA project "Modeling and Forecasting Cause-of-Death Mortality". This is a redesign and aggregation of the International Classification of Diseases (ICD) coding, allowing for a precise analysis of the role of drug and smoking related deaths in particular. The details of the construction of this classification in terms of ICD-10 codes is provided in Appendix F.

Table 1
LIST OF THE CAUSES OF DEATH USED IN THIS STUDY

Cause no	Working list
1	Cardiovascular diseases
2	Cerebrovascular diseases
3	Neoplasms directly induced by smoking (Neosmok)
4	Neoplasms (not directly induced by smoking)
5	Dementia
6	Diabetes
7	Influenza
8	Respiratory diseases
9	Drug abuse
10	External causes
11	Other

3.2 GEOGRAPHICAL LEVEL

The objective is to define groups of individuals with similar socioeconomic factors, allowing for identifying richest groups which are expected to be closer to an insured population County level was the geographical level chosen to capture sufficiently the differentials of some socioeconomic factors, and death rates by cause are available at this level. The counties are grouped by similar socioeconomic factors called "clusters" by using classification techniques and the mortality is forecasted by cause for each cluster, using a forecasting model which leverages general population mortality forecasts, as developed in the SOA project on "Modeling and Forecasting Cause-of-Death Mortality".

3.3 DATA SOURCE

The data source that has been chosen contains death rates by county and is extracted from the Centers for Disease Control and Prevention (CDC) database. The death numbers and exposures are at the granularity of gender by age, by year, by cause-of-death, by county which allows for a precise mortality analysis. However, three main challenges have been encountered:

- Exposure was not available for ages > 84 years. Therefore, the oldest age group studied in this research is 80–84.

- As the authors wanted to compare insured population to national population, they identified the insured population to a "cluster population", using methodologies found in the literature review. They assimilated the cluster population to the richest counties in terms of income, net earnings, etc. (see Beck and Webb, 2003 and Frees and Sun, 2010). The richest counties were grouped into a unique cluster. CDC Wonder (the online extraction tool of CDC data) allowed the aggregation of counties upstream the extraction. The authors also aggregated the ICD-10 causes-of-death into the 11 main causes-of-death, which form the working list as described in Table 1.

- In the database, when the number of deaths for a given level of extraction (gender by cause-of-death, by age, by county) is between 0 and 9, the information in the database is not available due to privacy purposes. To work around this issue, two complementary solutions have been found:

    - Aggregating the data avoided losing much of the data with the limited number of deaths (the less granular the data is, the greater the number of deaths). The aggregated numbers of deaths have directly been extracted at the granularity {gender by group of cause-of-death, by age, by cluster of counties} needed to retain this data.

    - In order to complete the still deleted lines due to the number of deaths between 0 and 9, the authors used an estimation based on national death rates, at the granularity {gender by age and by year}, following the methodology described below.

    Let us consider some age class a, gender g and year y and assume that the death rates for the cluster of counties $\mu_{k,a,g,y}^{cluster}$ are unknown for each cause $k = 1, \dots, m$ due to privacy purposes, whereas $\mu_{k,a,g,y}^{cluster}$ are known for each cause $k = m + 1, \dots, n$. Then the death rates for any cause $k = 1, \dots, m$ can be estimated using available cause-of-death mortality rates $\mu_{k,a,g,y}^{nat}$ for the general population by

    $$\hat{\mu}_{k,a,g,y}^{cluster} = \mu_{k,a,g,y}^{nat} \times \frac{\mu_{tot,a,g,y}^{cluster} - \sum_{i=m+1}^{n} \mu_{i,a,g,y}^{cluster}}{\sum_{i=1}^{m} \mu_{i,a,g,y}^{nat}},$$

    where $\hat{\mu}_{k,a,g,y}^{cluster}$ is the estimate of cluster-specific mortality for cause k, $\mu_{tot,a,g,y}^{cluster}$ is the cluster-specific mortality all causes aggregated, and $\mu_{tot,a,g,y}^{cluster} - \sum_{i=m+1}^{n} \mu_{i,a,g,y}^{cluster}$ is by construction the cluster specific mortality for causes 1 to m aggregated.

    Note that this methodology relies on the following decomposition of the aggregate death rate into the sum of the cause-specific death rates, which holds under the so-called independence assumption as discussed at the end of this section:

    $$\mu_{tot,a,y,g}^{cluster} = \sum_{k=1}^{n} \mu_{k,a,g,y}^{cluster}.$$

    Moreover, when the information is suppressed, it means that the number of deaths should be lower than 9.

    Thus, an estimated number of death in the cluster is for the cause k can finally be recovered as:

    $$\min(9, Expo_{a,g,y}^{cluster} \times \hat{\mu}_{k,a,g,y}^{cluster}),$$

where $Expo_{a,g,y}^{cluster}$ denotes the exposure-to-risk for the cluster of counties, for age class a, gender g, and year y. The number of deaths are then aggregated into {gender × year x age}, which is compared with the totals extracted from the same granularity; due to non-perfect match, excess deaths are then attributed to the cause "Other". In this approach, the historical death rates for the general population come from the SOA report "Modeling and Forecasting Cause-of-Death Mortality" (2019).

3.4 AGE GROUP CLASSIFICATION

The following age groups are available in the CDC and have been used:

- 0
- 1–4
- 5–9 and each subsequent five year age group up to 80–84

Note that for some of the graphics presented in this report, the age groups have been named according to their CDC prefixes or the first age of the group. For instance, "5" will refer to 5–9.

Although this aggregation by five-year age groups (the standard format provided) creates more stable historical death rates than one-year age groups, Gelman & Auerbach (2016) have shown that there might be some aggregation bias within mortality tables. The bias is due to the evolution of the age structure of the inner population within an age group. Since the group is aging over time, it may show a higher or lower number of deaths independent of the underlying (one-year age) mortality rates time pattern. The authors have performed a detailed analysis of possible aggregation biases in the previous research "Modeling and Forecasting Cause-of-Death Mortality", and have concluded that the aggregation bias was not significant for the age group lengths of five years.

3.5 SOCIOECONOMIC FACTORS BY COUNTY

In order to select the counties that will be used as a proxy for the insured population, an analysis is performed using socioeconomic factors available from the Bureau of Economic Analysis (BEA). The socioeconomic factors have been analyzed and those retained for the present study are focused on the economic situation of individuals, as it was proved by several references that the income has an impact on mortality (see Section 3). These factors are listed in Table 2 and their description is provided in Appendix D.

Table 2

EXPLANATORY VARIABLES

Explanatory variables
110 – Per capita personal income
120 – Per capita net earnings
130 – Per capita personal current transfer receipts
140 – Per capita income maintenance benefits
150 – Per capita unemployment insurance compensation
160 – Per capita retirement and other
170 – Per capita dividends, interest, and rent
250 – Wage and salary employment (% Total employment – number of jobs)
270 – Farm proprietors employment (% Proprietors employment)
2100 – Retirement and disability insurance benefits – Per capita
2200 – Medical benefits – Per capita
2600 – Education and training assistance – per capita

The observations available for these socioeconomic factors are from 1999 to 2017, by county. In this study, the authors retained the last year of historical data (2017) for each explanatory variable, and the corresponding evolution feature over the period defined as the variation of the socioeconomic factor between 1999 and 2017. Both indicators therefore provide complementary information on the level of the economic factor as well as its trend since 1999.

The correlations calculated among these variables are illustrated in Figure 1.

Figure 1
CORRELATION BETWEEN SOCIOECONOMIC VARIABLES
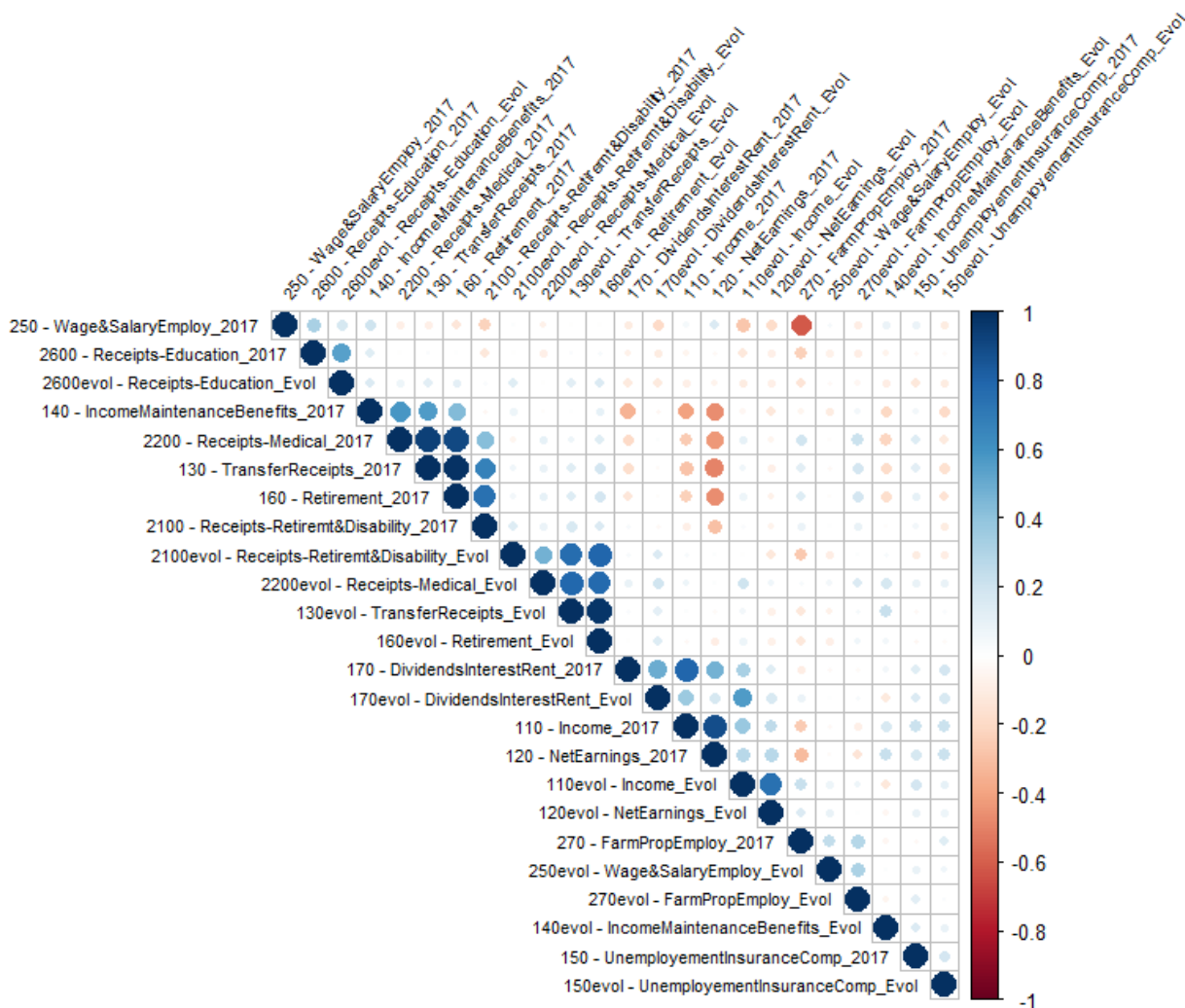
On this figure, large blue dots represent high positive correlation close to 1 whereas large red dots are associated with highly negative correlation close to −1. In between, small dots are linked with relatively low correlation close to zero.

Positive correlations between socioeconomic variables linked to the revenue (income, dividends, etc.) are observed: for instance, there is a correlation of 79.9% between "110-Income_2017" and "170-DividendsInterestRent_2017". There are also high positive correlations between socioeconomic variables linked to transfer of revenue (retirement, medical, etc.): as an example, the features "130-TransferReceipts_2017" and "160-Retirement_2017" are correlated at 98.8%. Correlations between transfer variables and revenue variables are mainly negative: for instance, "130-TransferReceipts_2017" and "120- NetEarnings_2017" are negatively correlated at −49.7%.

3.6 DEATH RATE DEFINITION AND INDEPENDENCE ASSUMPTION

For this report, the authors focused on the death rate as the force of the mortality. Thus, the death rate for an age group a, a gender g, a cause k and a year y is estimated by:

$$\mu_{k,a,g,y} = \frac{D_{k,a,g,y}}{E_{a,g,y}},$$

where $D_{k,a,g,y}$ refers to the number of deaths by cause k during year y of individuals of gender g in age group a last birthday. $E_{a,g,y}$ is an estimate of the so-called exposure-to-risk, that is the total life duration in the year y of individuals with gender g with age in goup a last birthday. The exposure does not relate to any cause and the sum of the by-cause estimates gives the total death rate estimate:

$$\mu_{a,g,y} = \frac{D_{a,g,y}}{E_{a,g,y}} = \frac{\sum_k D_{k,a,g,y}}{E_{a,g,y}} = \sum_k \mu_{k,a,g,y}.$$

Beyond this empirical description, below the authors show a theoretical clarification about the definition and main assumption on causes-of-death independence in the competing risks. This description may be omitted on first reading.

The competing risks framework is based on two causes, *A* and *B*. A cause-specific lifetime is associated with each cause, as

- τ^A : lifetime for cause *A* (such as cancer)

- τ^B : lifetime for cause *B* (such as all other causes)

The random duration τ^A can be interpreted as the lifetime in a world where only cause *A* would exist. The authors denote by τ the total lifetime which can be expressed as the minimum between cause-specific lifetimes as

$$\tau = \min(\tau^A, \tau^B)$$

so that in the competing risks framework life ends when one of the two clocks rings. The aggregate death rate (or force of mortality), denoted $\mu(a)$, is defined as the (instantaneous) probability of death before age $a + \delta$ for an individual aged a, for small increment δ. In comparison, the cause-specific death rate $\mu_i(a)$ corresponds to the (instantaneous) probability of death if only cause i exists, given the survival at age a.

The survival function at age a is defined as the probability that all lifetimes by cause will be higher than a.

$$S(a) = P(\tau > a) = P(\tau^A > a, \tau^B > a) = \exp\left(-\int_0^a \mu(y)dy\right)$$

The key issue is that the cause-specific death rate $P(\tau_i < a + \delta | \tau_i \geq a)$ (called "net" probability) cannot be estimated in practice in the general case since one only observes the "duration" of a given cause if death occurs from this cause, while the other durations remain right-censored (it is only known that they are longer than current lifetime). That is, the so-called "crude" probability can be estimated in practice:

$$P(\tau_i < a + \delta | \tau = \tau_i, \tau \geq a)$$

In this work, as it is most often the case in cause-of-death analysis, it is assumed that cause-specific lifetimes τ^A and τ^B are independent, which implies two key consequences:

- The net and the crude death rates are equal. In other words, the (net) cause-specific death rate can directly be estimated from the data using the formula $\mu_{k,a,g,y} = \frac{D_{k,a,g,y}}{E_{a,g,y}}$.

- The survival function can be rewritten as:

$$S(a) = P(\tau^A > a) \times P(\tau^B > a) = \exp\left(-\int_0^a \mu_A(y)dy\right) \times \exp\left(-\int_0^a \mu_B(y)dy\right)$$

Thus, an aggregate mortality rate can be expressed as the sum of the underlying cause-specific rates.

For a more detailed discussion on the dependency structure between cause-specific lifetimes, refer to Dimitrova et al. (2013), Arnold et al. (2018), and references therein.

Section 4: Clustering Counties by Socioeconomic Factors

The purpose of this section is to develop and detail the clustering method by socioeconomic factors that has been retained. In the first subsection, an overview of the clustering methodology is presented. Then, the results are detailed step by step in the second subsection.

4.1 OVERVIEW OF THE CLUSTERING PROCESS

The idea is to use explanatory variables (economic factors) to cluster the counties into homogeneous groups of characteristics. The methodology chosen was to perform an Ascending Hierarchical Classification (AHC) on all explanatory variables in order to group counties. The cluster dendrogram was used to determine the number of clusters. Then, the authors performed a Principal Components Analysis (PCA) to better understand the composition of each cluster and to represent AHC's clusters on the axis defined by the PCA; this allows us to identify the main variables which define each cluster.

The three main steps of the clustering process:

Step 1: The hierarchical clustering dendrogram allows us to group the initial data into clusters which are more and more precise, thanks to a calculation of the distance between the observations. This dissimilarity is represented by the arrows in the space of the characteristics, see Figure 4. Then, each county is assigned to a given cluster.

Step 2: The second step of the clustering process consists in visualizing the observations on the axis defined by the PCA. On the axis defined by the PCA, the observations (here, U.S. county codes) are plotted as well as their associated cluster (represented by a color). Thus, we can visualize the position of each cluster on the axis.

Step 3: The next step is to analyze the position of the socioeconomic variables on the PCA's axis. To each cluster are associated specific variables that characterize the most the cluster.


A first clustering by counties is made on the U.S. population. The cluster dendrogram suggests four clusters. The "richest" cluster contains 116 counties. A second clustering is made on these 116 counties to refine the clustering and keep only the top of the richest counties.

4.2 DETAILED RESULTS OF THE CLUSTERING PROCESS

4.2.1 FIRST CLUSTERING

An AHC is performed on all U.S. counties. As depicted in Figure 2, the first dendrogram separates counties via different characteristics (branches). The branches are numerous but we can distinguish four main branches. Thus, the first dendrogram suggests four clusters.

Figure 2
CLUSTERING PROCESS: AHC
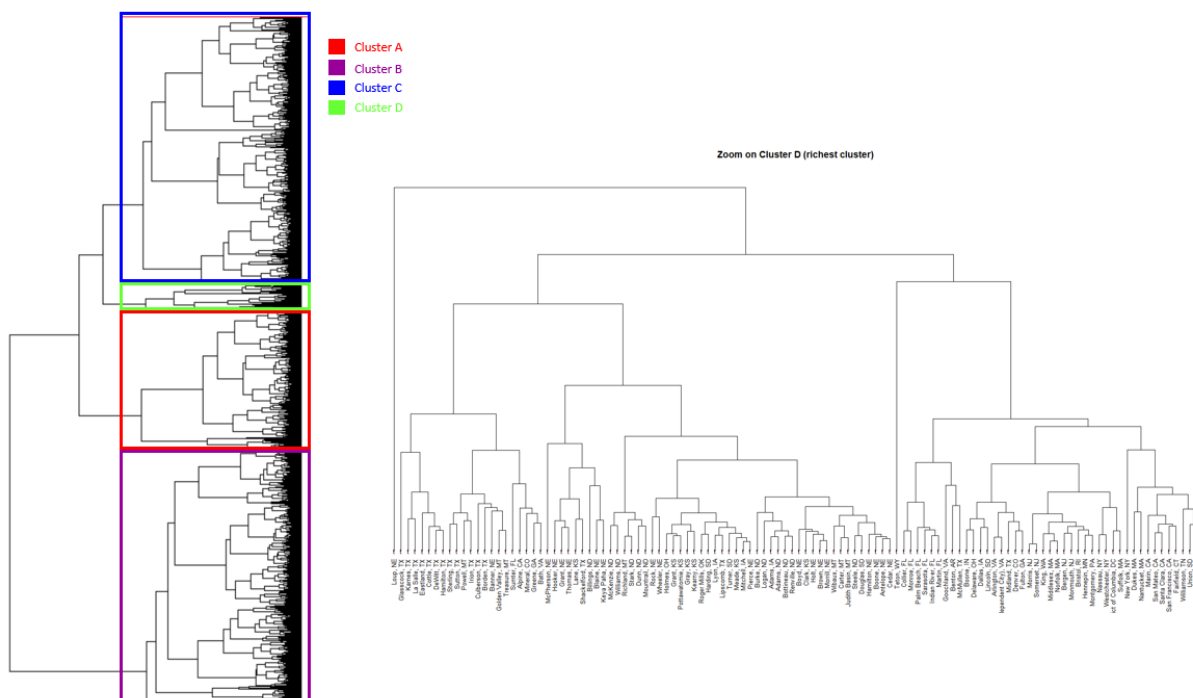
Then, a PCA is performed. The U.S. county codes are plotted as well as their associated cluster (represented by a color), on the axis defined by the PCA (see Figure 3). This allows the visualization of the position of each cluster on the axis. At the same time, the position of the socioeconomic variables on the PCA's axis are analyzed. Each cluster is associated with specific variables that characterize it the most, see Figure 4.

Figure 3
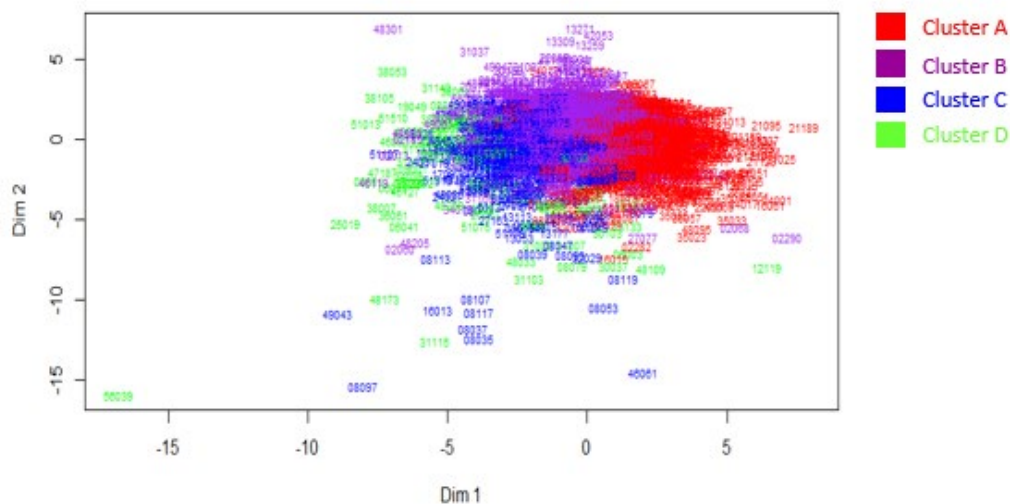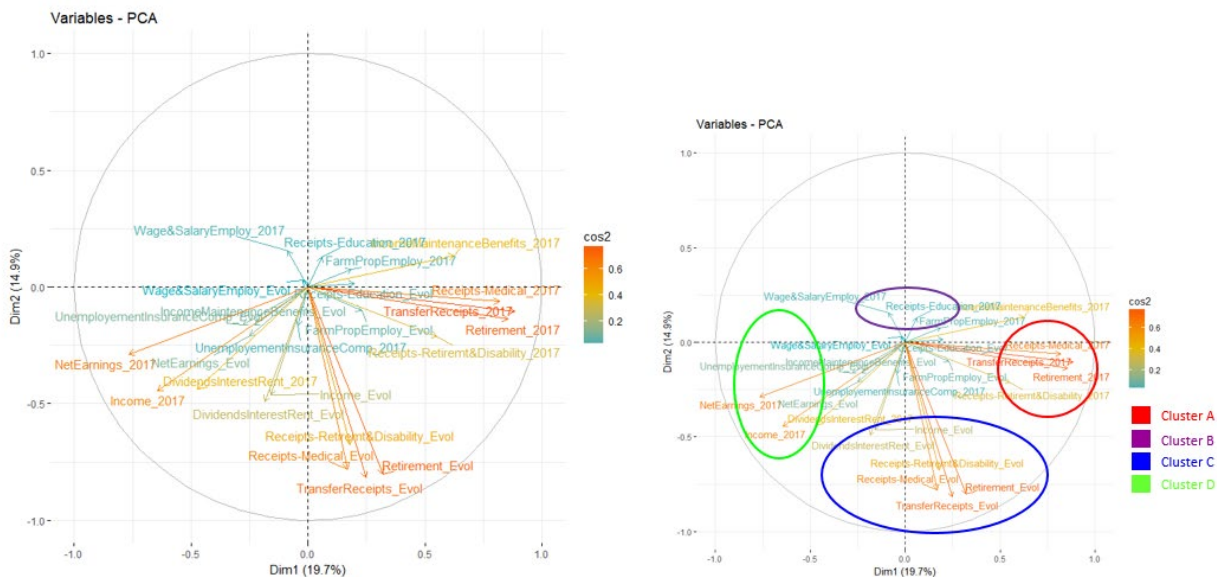CLUSTERING PROCESS: PCA – observations

Figure 4
CLUSTERING PROCESS: PCA – variables

Once the clusters have been identified and analyzed (in terms of socioeconomic variables), they can be represented on the U.S. map, Figure 5. The richest counties are represented in green, whereas the most deprived counties are filled in red.

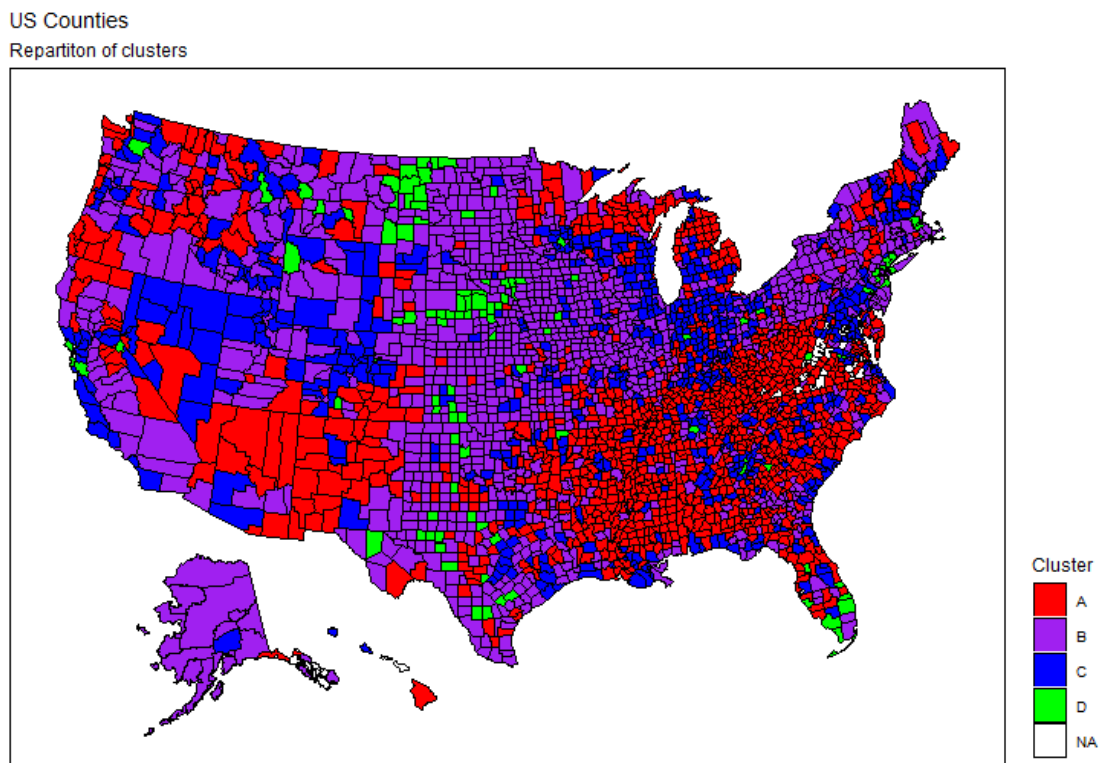**Figure 5**

GRAPHICAL REPRESENTATION OF THE RICHEST COUNTIES

US Counties
Repartiton of clusters

Cluster
A
B
C
D
NA

Table 3

FIRST CLUSTERING DESCRIPTION

First clustering description	Cluster A	Cluster B	Cluster C	Cluster D
110 – Per capita income 2017	$36,912	$41,700	$46,673	**$68,783**
110evol – Per capita income evolution (1999–2017)	+0.83%	+0.87%	+0.78%	**+1.52%**
250 – Wage and salary employment (% of total number of jobs)	70%	69%	76%	65%
120 – Per capita net earnings 2017	$19,045	$24,237	$29,741	**$40,627**
120evol – Per capita net earnings evolution (1999–2017)	+0.56%	+0.75%	+0.63%	**+1.84%**
170 – Per capita dividends, interest, rent 2017	$6,578	$8,044	$8,981	**$19,600**

From Figure 4, the PCA highlights that income, net earnings, dividends, interests and rents are major explanatory variables for cluster D. Other variables such as wage and salary employment displayed in Table 3 do not characterize cluster D. The first clustering underlines that cluster D is the richest cluster compared to A, B and C. It concentrates the highest values of income, net earnings, dividends, interest and rent and thus, we believe best represents an insured population.

The income per capita and earnings distribution by county in each cluster are displayed below in Figures 6 and 7.

Figure 6
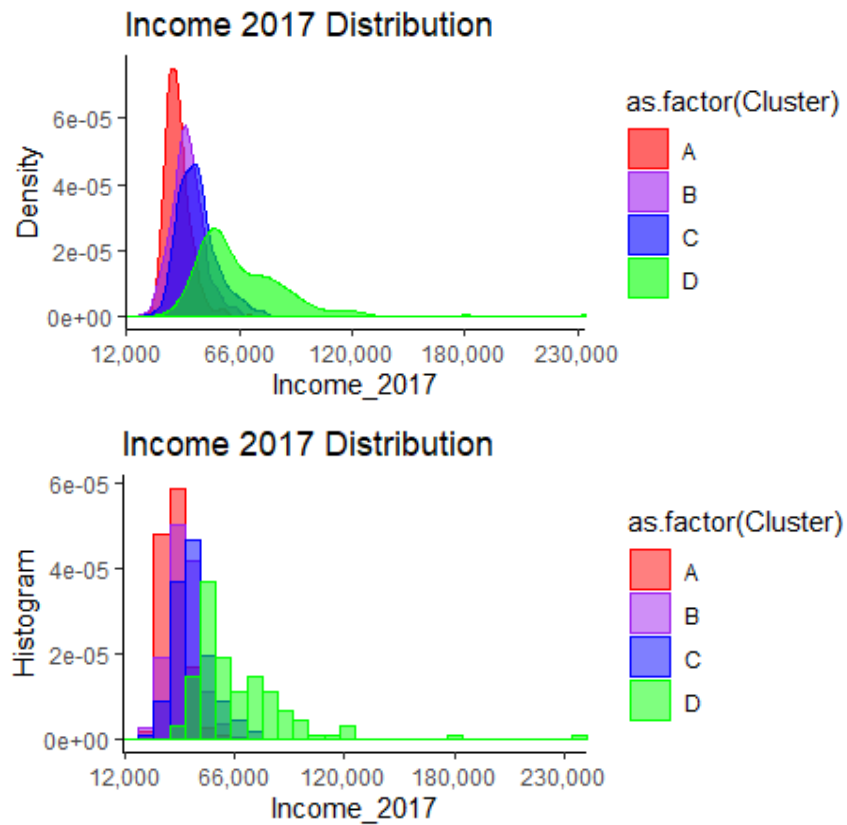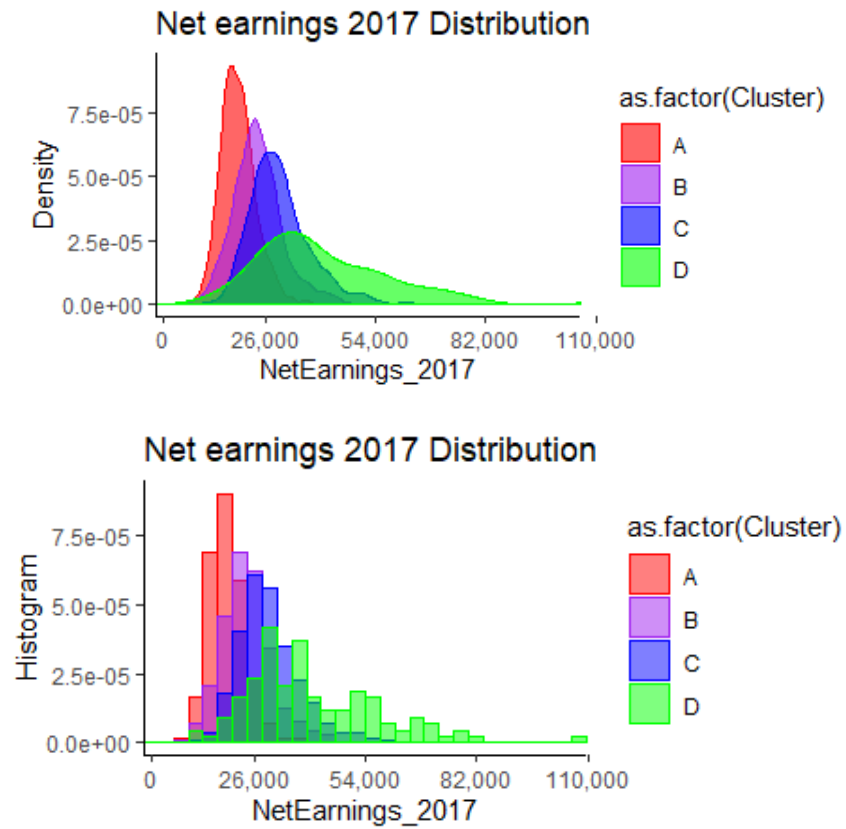INCOME PER CAPITA BY COUNTY 2017, DISTRIBUTION OF THE FIRST CLUSTERING

Figure 7
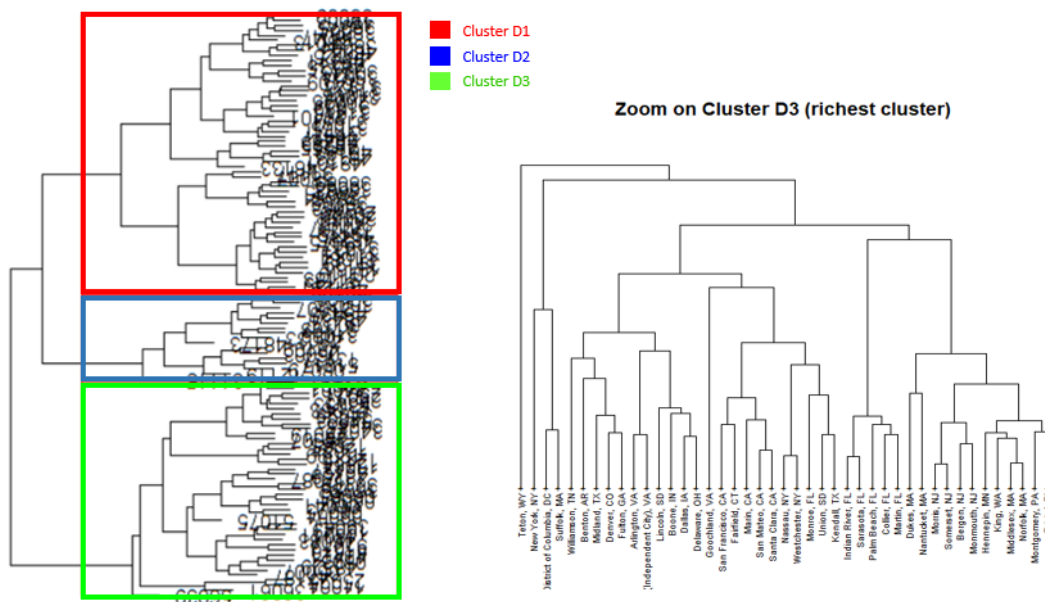NET EARNINGS PER CAPITA BY COUNTY 2017, DISTRIBUTION OF THE FIRST CLUSTERING

4.2.2 SECOND CLUSTERING

A second clustering was made on the cluster D, which contains 116 counties in order to see if we could improve the results by clustering more.
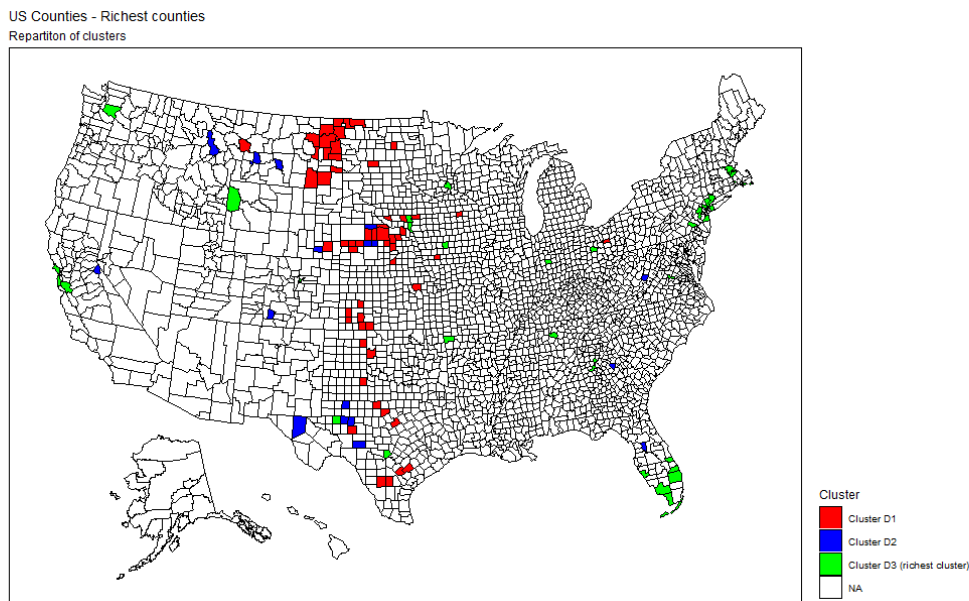
Figure 8

CLUSTER DENDROGRAM – RICHEST COUNTIES

Again, the dendrogram separates counties via different characteristics (branches). We can distinguish three main branches and wealth is the main driver of this split. Therefore, the cluster dendrogram of the second clustering suggests three clusters. The PCA analysis allows us to identify the richest cluster (in green). It contains 43 counties. In terms of population, this cluster represents 98% of the total population of the 116 counties (sum of the population of the three clusters), see Table 5.

Figure 9

GRAPHICAL REPRESENTATION OF THE RICHEST COUNTIES

US Counties - Richest counties
Repartiton of clusters

Cluster
Cluster D1
Cluster D2
Cluster D3 (richest cluster)
NA

These 43 counties within cluster D3 in green as depicted in Figure 9 will be retained as approximating an insured population. These counties are listed in the Appendix A.

Table 4

SECOND CLUSTERING DESCRIPTION

Second clustering description	Cluster D1	Cluster D2	Cluster D3
110 – Per capita income 2017	$55,538	$55,757	**$91,182**
110evol – Per capita income evolution (1999–2017)	+1.64%	+2.17%	+1.11%
250 – Wage and salary employment (% of total number of jobs)	60%	59%	**74%**
120 – Per capita net earnings 2017	$33,712	$28,658	**$54,364**
120evol – Per capita net earnings evolution (1999–2017)	+2.06%	+3.33%	+0.96%
170 – Per capita dividends, interest, rent 2017	$13,400	$16,698	**$28,823**

Income, wages and salary employment, net earnings, dividends, interest and rent are major explanatory variables for cluster D3. Other variables such as per capita income evolution and per capita net earnings evolution displayed in Table 4 do not characterize cluster D3 but allow to compare the new clusters with Table 3. This second clustering highlights that cluster D3 is the richest. It contains 43 counties that are the richest among the richest and thus, best represents an insured population.

The income per capita and earnings distribution by county in each cluster for the second clustering is displayed below in Figures 10 and 11.

Figure 10

INCOME PER CAPITA BY COUNTY 2017, DISTRIBUTION OF THE SECOND CLUSTERING
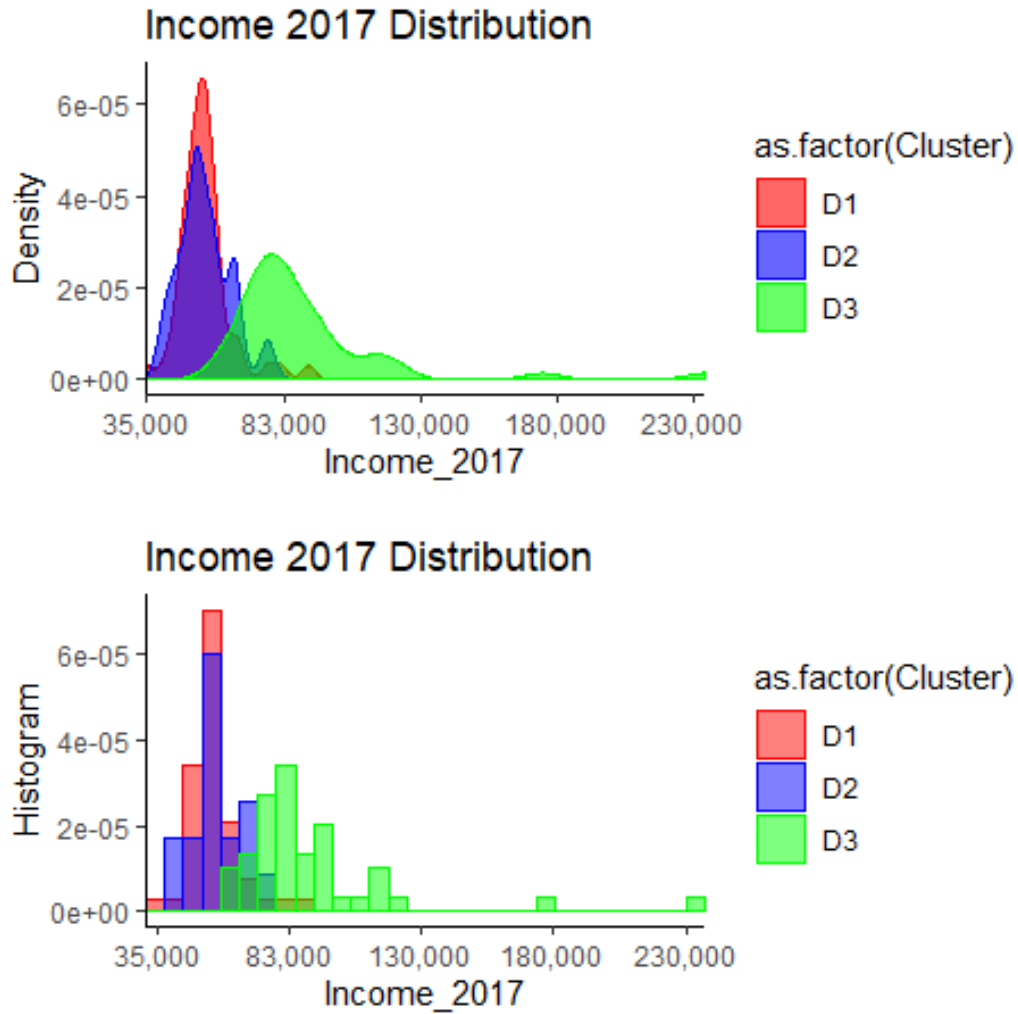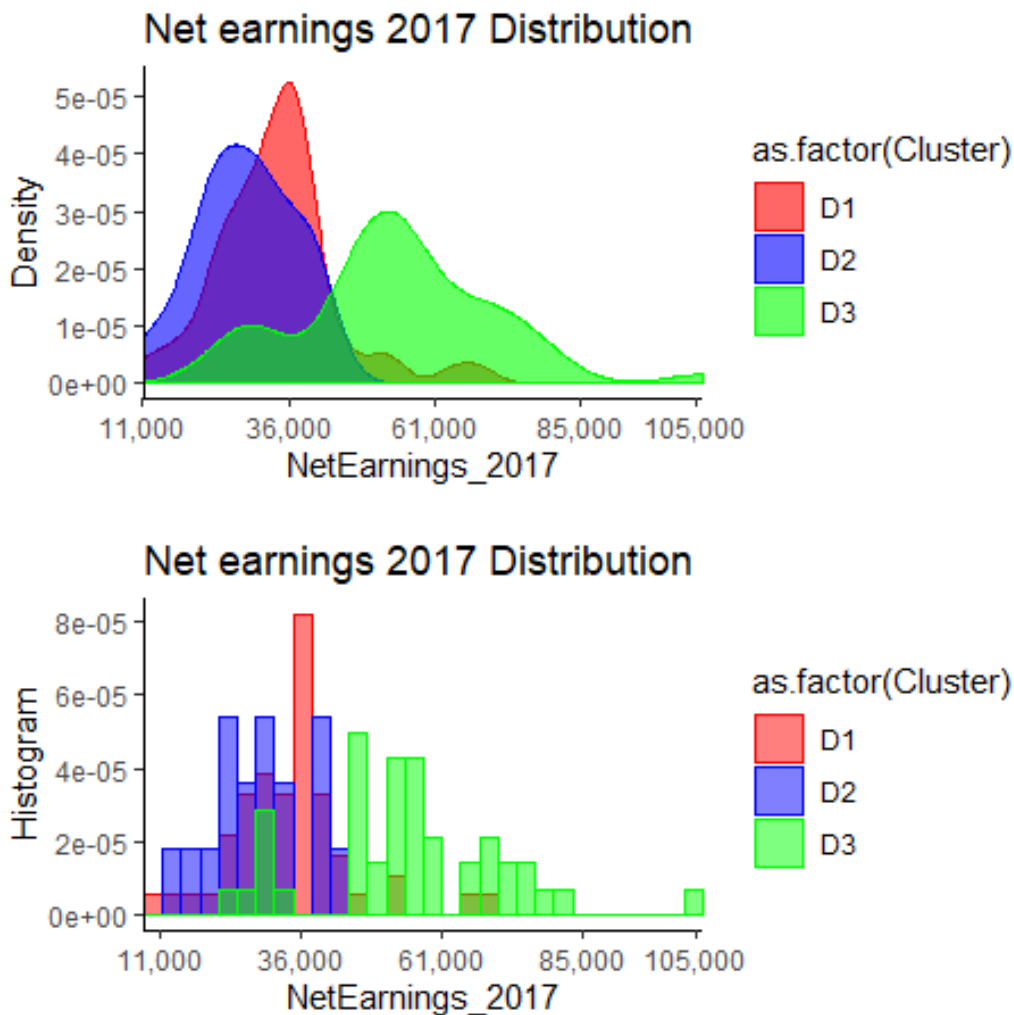
Figure 11
NET EARNINGS PER CAPITA BY COUNTY 2017, DISTRIBUTION OF THE SECOND CLUSTERING

As the cluster D3 is the richest, this cluster will be retained as best representing the insured population. As depicted in Table 5, the population of this cluster contains more than 24 million exposures and 110 thousand deaths (all causes) in 2016, which appears to be sufficient to infer mortality rates. A detail of the comparison of exposures and deaths between the cluster population and the general population is provided in Tables 6 and 7. We note that cluster D3 contains the majority of the exposure of cluster D (97.8%), see Table 5. Thus, there is no need to cluster more than twice as the second clustering significantly reduces the number of clusters but does not significantly change the exposure compared to the first clustering.

Table 5
POPULATION OF THE SECOND CLUSTERING

Second clustering description	Cluster D1	Cluster D2	Cluster D3
Exposure 2016	394,219	161,493	24,432,088
Distribution	1.6%	0.06%	97.8%
Deaths 2016			110,576

Table 6

COMPARISON OF THE FEMALE POPULATION OF CLUSTER D3 WITH THE NATIONAL POPULATION

FEMALE, 2016	Cluster D3		National		Comparison	
Age	Exposure	Deaths number	Exposure	Deaths number	ExposureD3/ ExposureNat	DeathsD3/ DeathsNat
0–1	140,400	502	1,935,841	10,089	7.3%	5.0%
1–4	546,468	92	7,800,077	1,773	7.0%	5.2%
5–9	692,650	67	9,994,620	1,091	6.9%	6.1%
10–14	699,476	50	10,106,986	1,271	6.9%	3.9%
15–19	722,979	156	10,334,167	3,006	7.0%	5.2%
20–24	793,608	291	10,892,162	5,337	7.3%	5.5%
25–29	965,822	457	11,259,234	7,319	8.6%	6.2%
30–34	950,749	549	10,838,813	9,537	8.8%	5.8%
35–39	863,921	624	10,417,907	12,044	8.3%	5.2%
40–44	809,641	750	9,955,829	16,036	8.1%	4.7%
45–49	864,637	1,344	10,581,069	25,916	8.2%	5.2%
50–54	882,207	2,305	11,113,759	42,434	7.9%	5.4%
55–59	866,317	3,340	11,287,717	63,024	7.7%	5.3%
60–64	769,990	4,365	10,168,608	80,371	7.6%	5.4%
65–69	675,497	5,764	8,854,098	102,809	7.6%	5.6%
70–74	493,931	6,747	6,379,430	119,187	7.7%	5.7%
75–79	367,628	8,286	4,656,383	141,282	7.9%	5.9%
80–84	280,905	11,437	3,420,320	178,767	8.2%	6.4%
Total	12,386,826	47,126	159,997,020	821,294	7.7%	5.7%

Table 7

COMPARISON OF THE MALE POPULATION OF CLUSTER D3 WITH THE NATIONAL POPULATION

Male, 2016	Cluster D3		National		Comparison	
Age	Exposure	Deaths number	Exposure	Deaths number	ExposureD3/ ExposureNat	DeathsD3/ DeathsNat
0–1	146,797	629	2,025,815	12,682	7.2%	5.0%
1–4	570,468	94	8,162,659	2,358	7.0%	4.0%
5–9	720,328	78	10,427,099	1,394	6.9%	5.6%
10–14	725,716	85	10,530,762	1,740	6.9%	4.9%
15–19	740,990	380	10,810,282	7,453	6.9%	5.1%
20–24	797,702	967	11,493,228	15,932	6.9%	6.1%
25–29	981,200	1,223	11,659,004	19,101	8.4%	6.4%
30–34	965,670	1,172	10,997,251	20,299	8.8%	5.8%
35–39	864,325	1,282	10,402,812	22,000	8.3%	5.8%
40–44	803,328	1,492	9,812,627	26,280	8.2%	5.7%
45–49	853,621	2,313	10,384,925	40,205	8.2%	5.8%
50–54	862,819	3,745	10,731,937	65,547	8.0%	5.7%
55–59	827,185	5,554	10,673,935	98,971	7.7%	5.6%
60–64	701,903	6,944	9,311,773	124,440	7.5%	5.6%
65–69	585,789	8,136	7,907,485	144,563	7.4%	5.6%
70–74	408,153	8,551	5,474,111	151,081	7.5%	5.7%
75–79	286,350	9,390	3,735,627	160,466	7.7%	5.9%
80–84	202,918	11,415	2,460,452	173,088	8.2%	6.6%
Total	12,045,262	63,450	157,001,784	1,087,599	7.7%	5.8%

Section 5: Death Rates Historical Comparison

The purpose of this section is to analyze the historical data obtained with the death rates classification, before any projection.

5.1 AGGREGATE DEATH RATES

The historical aggregate death rates are plotted in Figures 12 and 13 below.

Figure 12
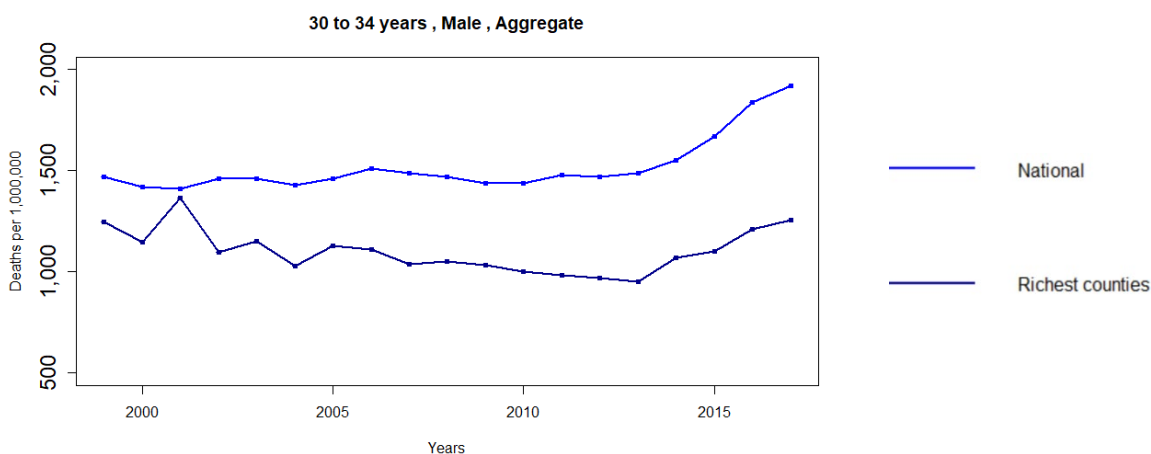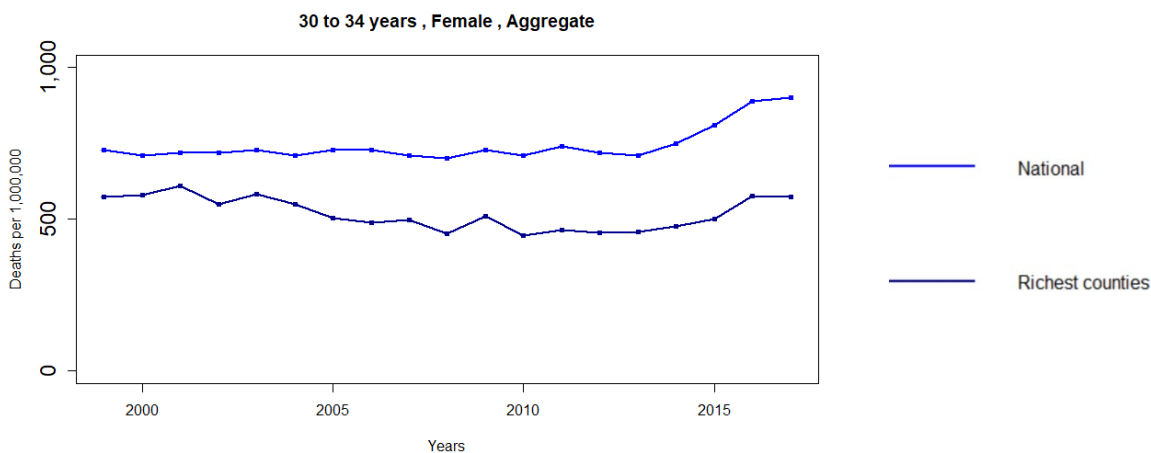HISTORICAL AGGREGATE DEATH RATES FOR MALE AGES 30–34

Figure 13
HISTORICAL AGGREGATE DEATH RATES FOR FEMALE AGES 30–34

The national death rates at ages 30 to 34 have stagnated between 1999 and 2013 and then have increased since 2013. For the richest counties, the death rates have decreased a little until 2013 and then have increased between 2013 and 2017. However, the increase was lower for the richest counties than for the national population, showing an increasing gap between the two populations. Overall, the magnitude of the difference between the level of death rates underlines the inequalities between the two populations.

Another observation is that the gap for males is larger than females and has steadily increased. For females, the gap is smaller and appears to be consistent for a fair amount of time.

Figure 14

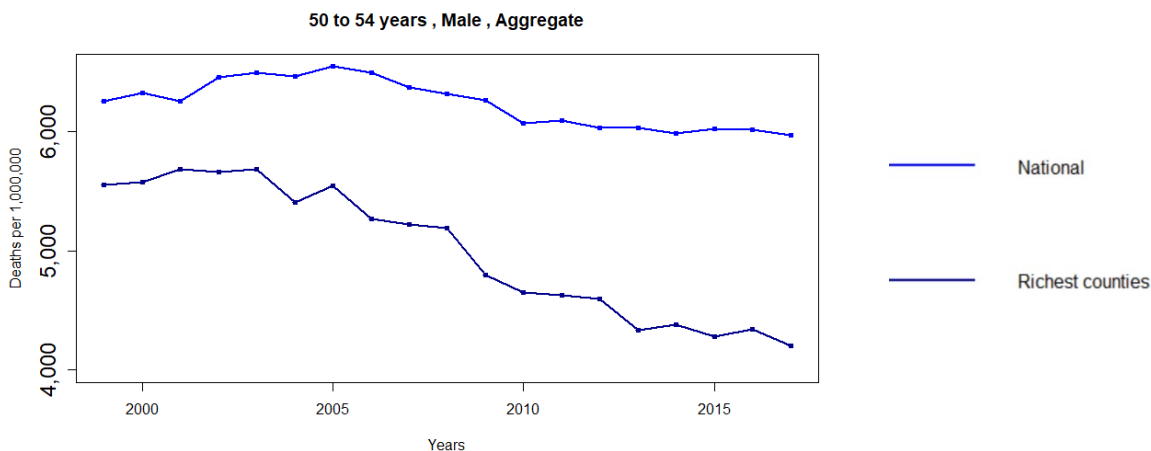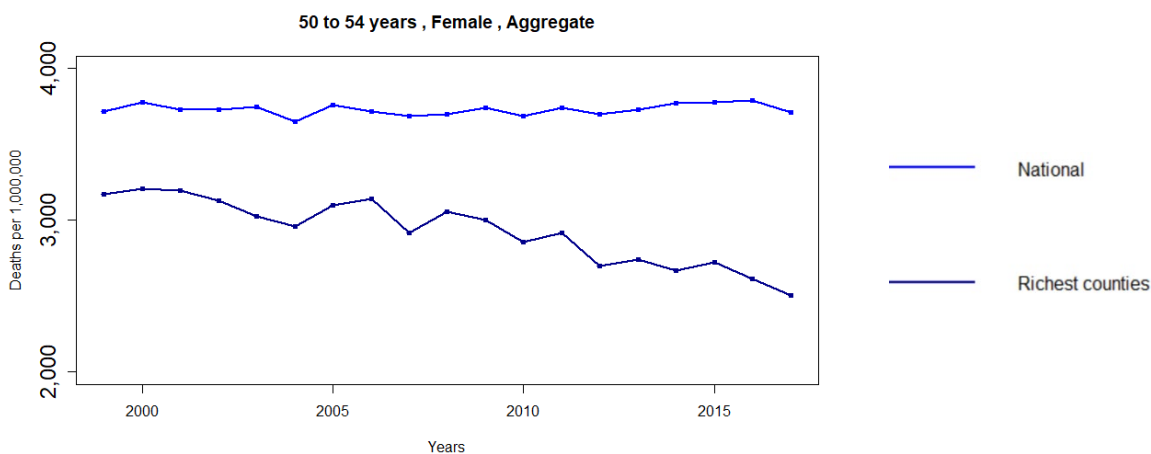HISTORICAL AGGREGATE DEATH RATES FOR MALE AGES 50–54

50 to 54 years , Male , Aggregate

Figure 15

HISTORICAL AGGREGATE DEATH RATES FOR FEMALE AGES 50–54

50 to 54 years , Female , Aggregate

For females, the national death rates at ages 50 to 54 have stagnated between 1999 and 2017 whereas the death rates for the richest counties have decreased, especially since 2006. For males, the national death rates have increased between 1999 and 2005 and then have decreased between 2005 and 2017. For males in the richest counties, this decrease is greater since 2003. The level of death rates is also lower for the richest counties for both males and females, and the gap between the national and the richest counties populations has widened since 1999, which reinforces the inequalities between the two populations in terms of mortality rates at intermediate ages 50–54. Overall, the gap between national and richest was

about the same for males and females until around 2003, when it began widening for males. The gap from then forward was greater for males, although the gap for females began to widen in 2009.

Figure 16
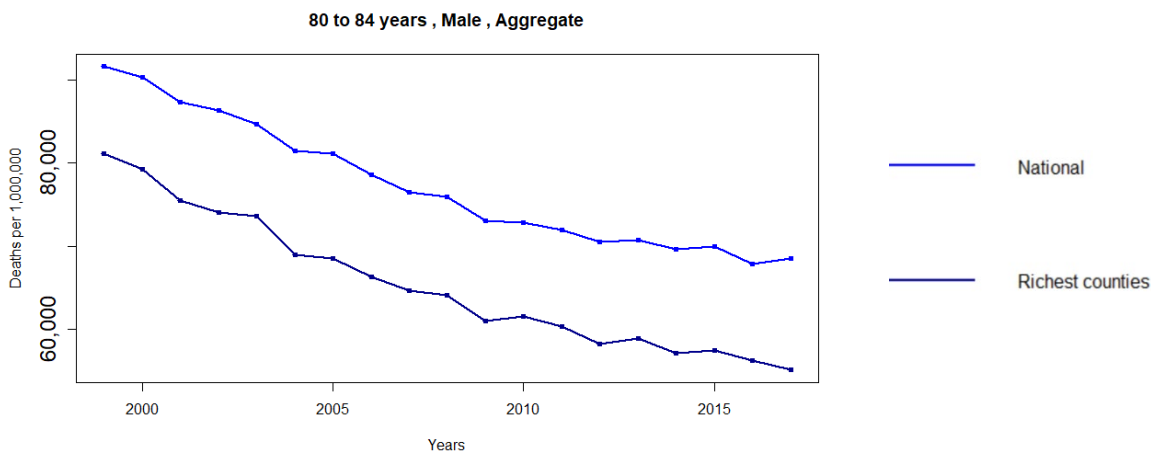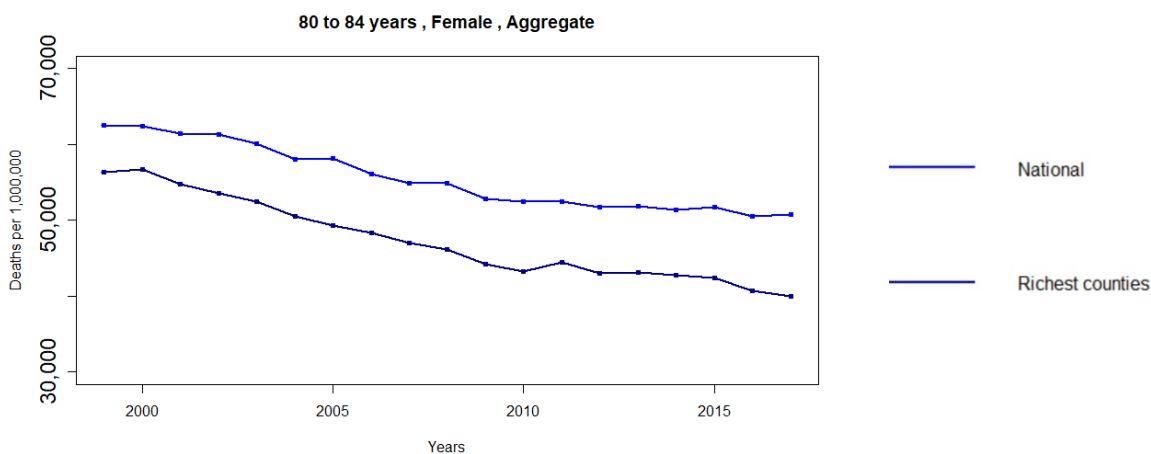HISTORICAL AGGREGATE DEATH RATES FOR MALE AGES 80–84

Figure 17
HISTORICAL AGGREGATE DEATH RATES FOR FEMALE AGES 80–84

The national death rates at ages 80 to 84 have decreased since 1999 for both males and females. While the level of death rates within the richest counties remains lower, the decrease is rather similar for the richest counties to that seen at the national level. There was a slight increase in the gap for both males and females beginning in about 2013.

5.2 BY-CAUSE DEATH RATES

This section provides cause-of-death historical analysis for sample age groups and focuses on the top four causes-of-death in each age group, split by the richest counties and national population, and by gender.

Figure 18 looks at external, drug, other, and cardiovascular causes for males ages 30–34 for the richest counties and nationally.

Figure 18

HISTORICAL DEATH RATES FOR MALE AGES 30–34

30 to 34 years , Male , Richest counties

30 to 34 years , Male , National
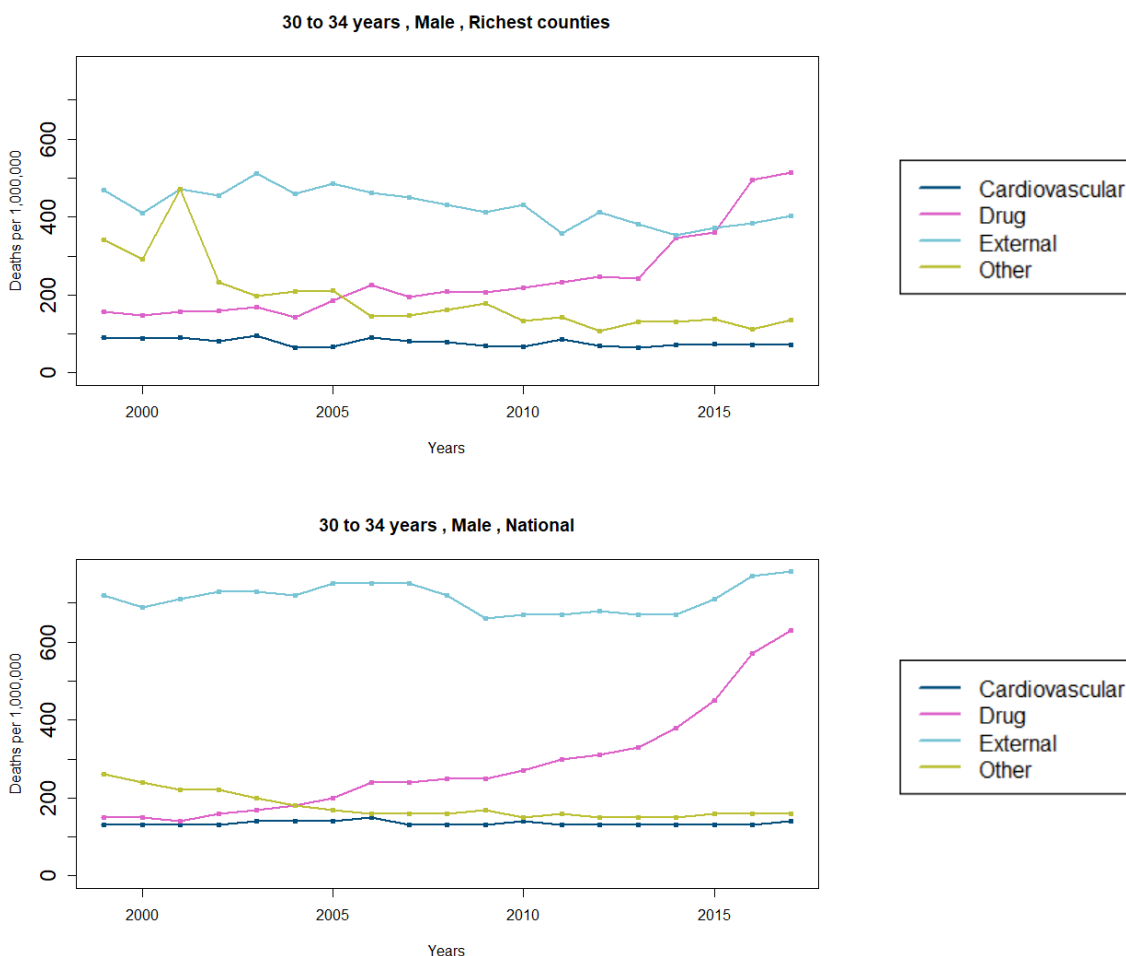
For males aged 30 to 34, external mortality is clearly higher in the national population as compared to the richest counties population. The external cause-of-death and the drug cause-of-death are the two main causes-of-death for people aged 30 to 34. The death rates for drugs began increasing for both the richest counties and nationally in the early 2000s and exceeds the death rates for cause external for the richest counties population in 2016 and 2017. Drug deaths did not exceed external causes for the national population because the death rates for cause external are much higher. Regarding the two next most common causes, other and cardiovascular, one observes a roughly stable evolution since 2005.

Figure 19

HISTORICAL DEATH RATES FOR FEMALE AGES 30–34

30 to 34 years , Female , Richest counties

**30 to 34 years , Female , National**
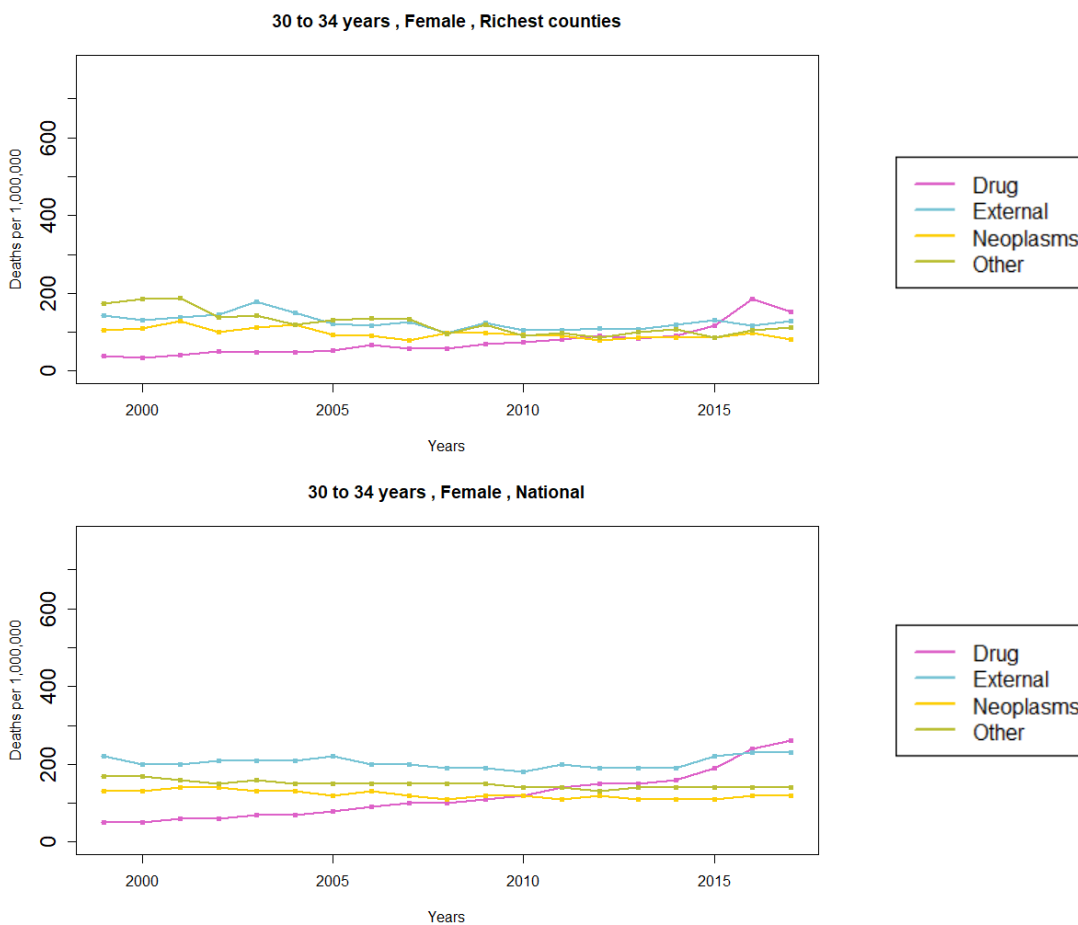
In 2016 and 2017, drugs exceeded the cause external for females, for both the richest counties and the national populations. The increase in death rates for drugs is more important for the national population than the overall death rates by cause because the national population is naturally larger. Death rates by other causes and external for the richest counties alternate being the highest until 2016 when drug death rates are highest. Death rates are more erratic for the cluster population at low ages due to smaller exposure.

Figure 20 shows death rates for cardiovascular, neoplasms, other, and drugs for male ages 50–54 for the richest counties and nationally.

Figure 20

HISTORICAL DEATH RATES FOR MALE AGES 50–54

The main cause-of-death at age 50 to 54 is neoplasms for females, and cardiovascular for males. The death rates are globally decreasing with time except for the cause drug.

50 to 54 years , Male , Richest counties

50 to 54 years , Male , National
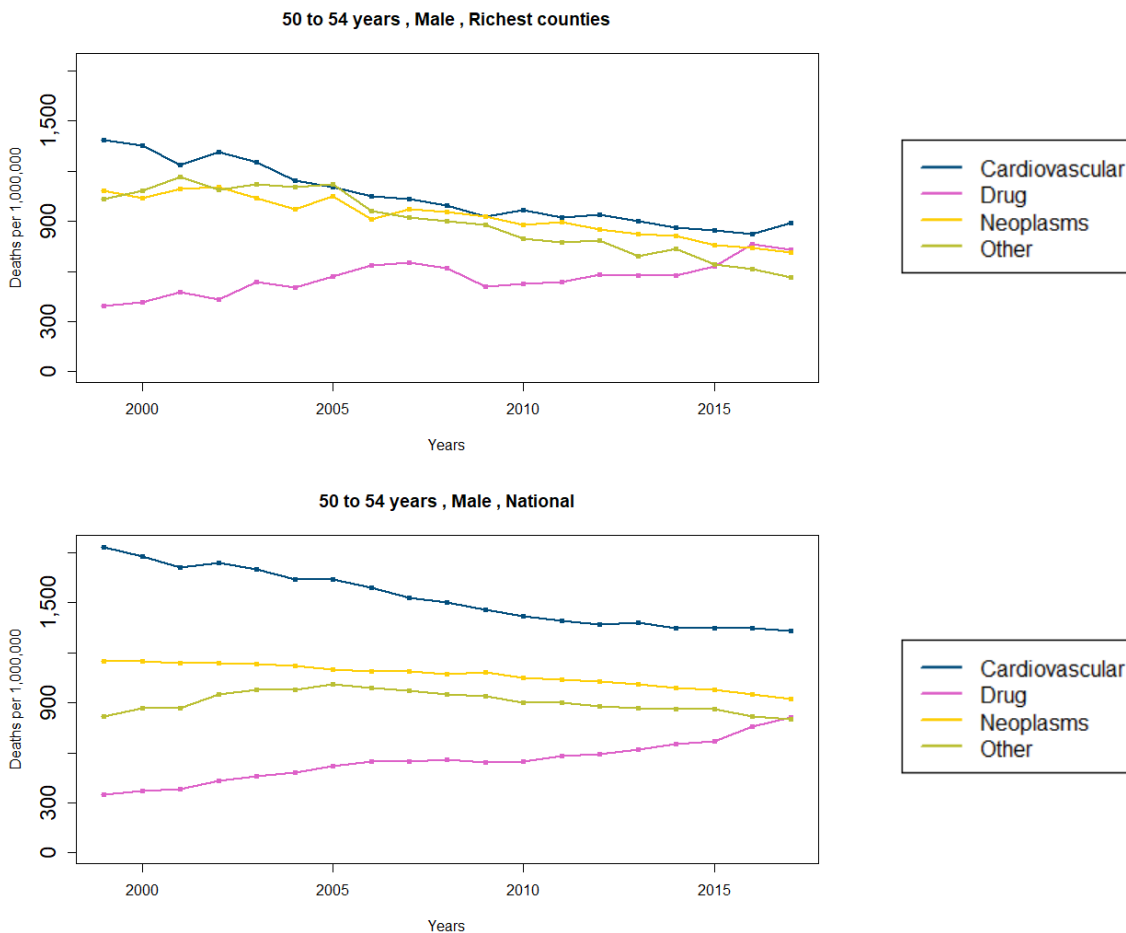
The order from highest to lowest of the four causes viewed are generally cardiovascular, neoplasms, other causes, and drugs, and the first three are declining for both the richest counties and nationally. Drugs generally increased to 2016, where they were in second place for the richest counties and tied for third nationally. The first three causes for the richest counties were generally relatively close while they were considerably further apart nationally. The death rates for neoplasm were rather similar for both the richest counties and nationally, although slightly lower in the richest cluster. In addition, for males the cardiovascular death rates are significantly higher for the national population compared to the richest counties. Finally, trends for the other cause differ, since the trend is clearly steadily decreasing in the cluster population.

Figure 21 shows the death rates for female ages 50–54 for neoplasms, other causes, cardiovascular, and drugs for the richest counties and nationally.

Figure 21

HISTORICAL DEATH RATES FOR FEMALE AGES 50–54

50 to 54 years , Female , Richest counties

50 to 54 years , Female , National
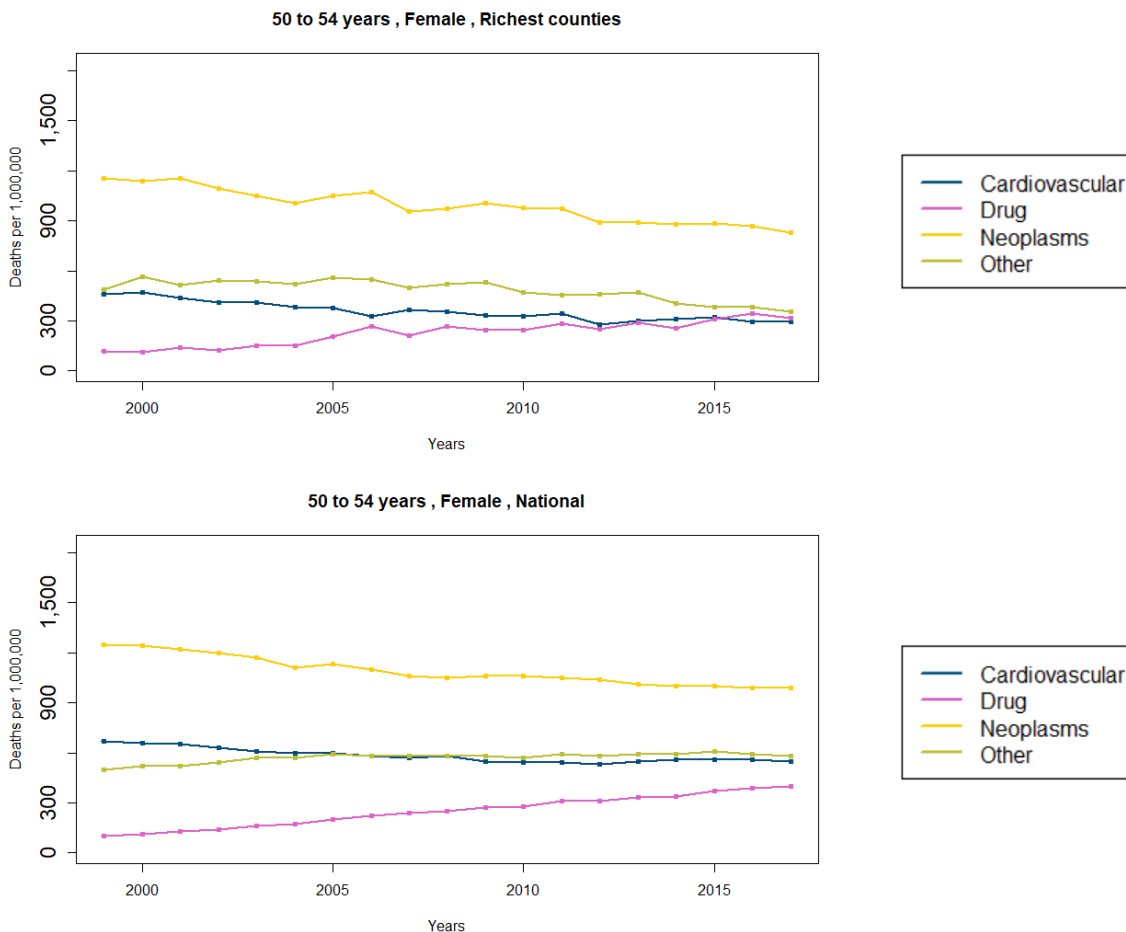
For females, the mortality trends by cause over the period are similar for the cluster and the national population. Neoplasms have the highest death rates for both populations and are decreasing while drugs have the lowest death rate, except beginning in 2016 for the richest counties, and are increasing in both populations. Slightly different patterns can be observed for the cause "other", with an increasing trend in the general population and a decreasing trend for the richest counties.

Figure 22 shows the death rates for male ages 80–84 for cardiovascular, neoplasms, dementia, and other causes for both the richest counties and national population.

Figure 22

HISTORICAL DEATH RATES FOR MALE AGES 80–84

80 to 84 years , Male , Richest counties


80 to 84 years , Male , National
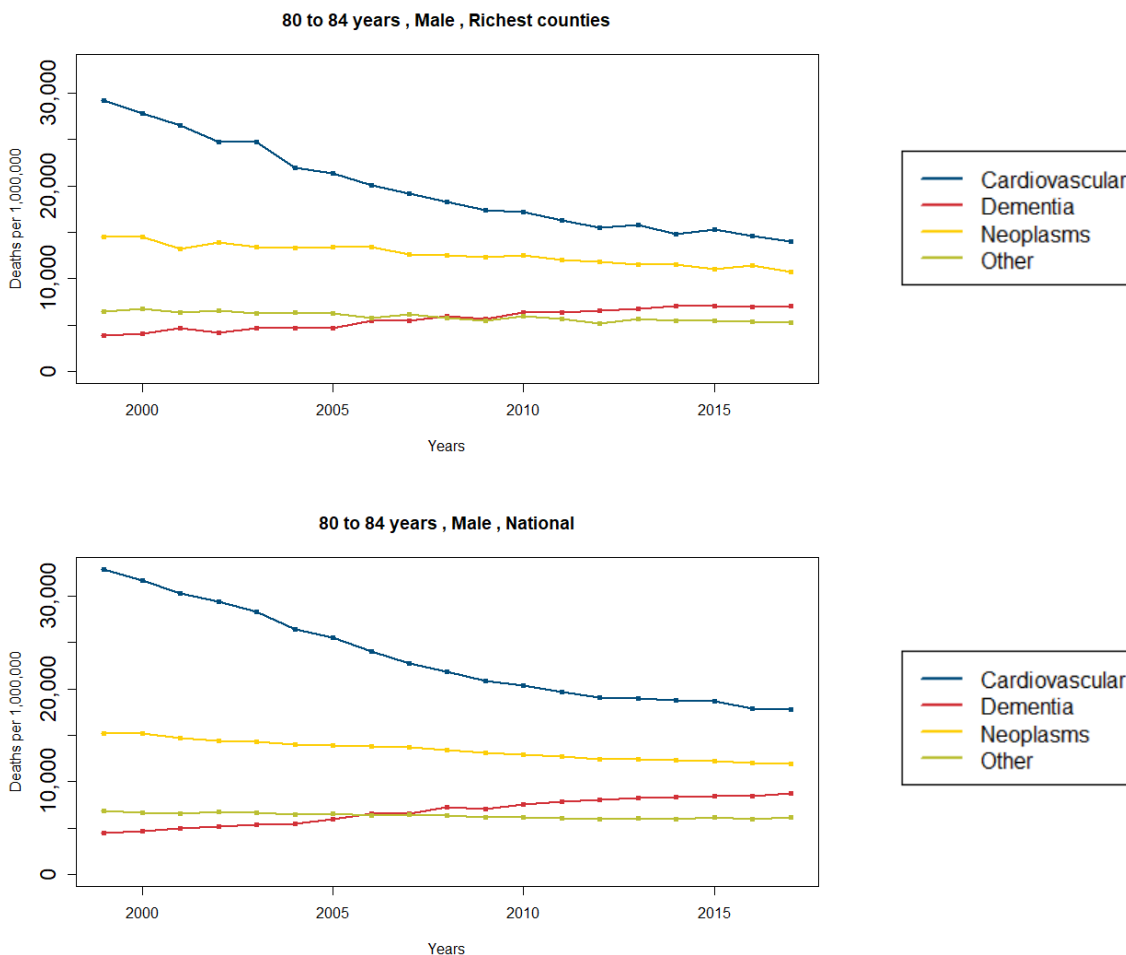
No strong mortality differential between national and cluster population seems to appear at high ages. The mortality trends by cause over the period are similar for the cluster and the national population. Slight differences can be still observed for cardiovascular diseases related death rates, where the rate is higher in the national population compared to the cluster population. The deaths due to cardiovascular diseases are decreasing whereas the deaths due to dementia are increasing for both males and females. In fact,

dementia overtakes other causes as the third leading cause of death about 2006–2007 for both populations.

Figure 23 shows the death rates for female ages 80–84 for cardiovascular, neoplasms, dementia, and other causes for both the richest counties and national populations.

Figure 23

HISTORICAL DEATH RATES FOR FEMALE AGES 80–84

80 to 84 years , Female , Richest counties

80 to 84 years , Female , National
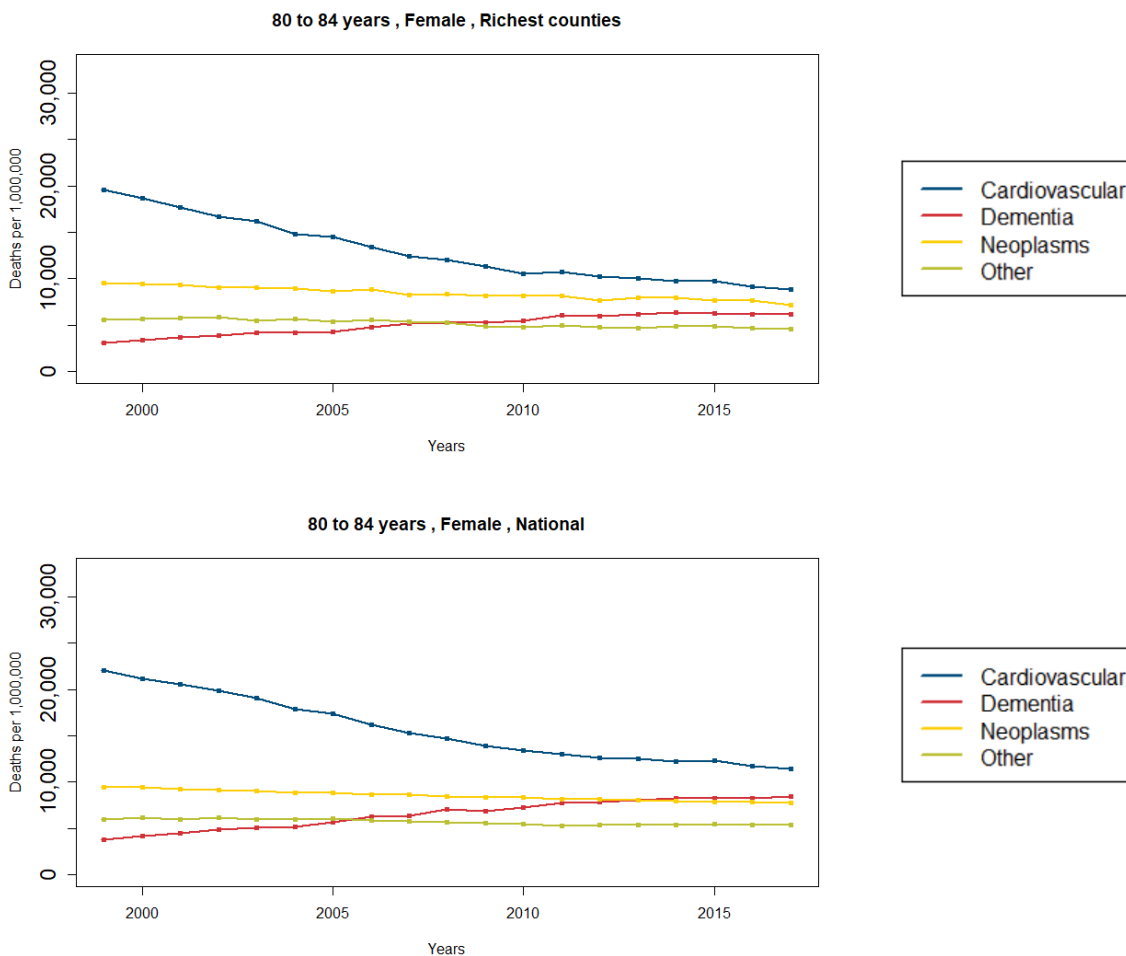
The deaths due to cardiovascular diseases were higher in the national population, but decreased in both populations. Neoplasms and other causes were relatively flat and at about the same level for both populations. Dementia-related deaths increased in both populations, the cause dementia becoming the third leading cause of death in about 2008 for the richest counties and the third leading cause in about 2006 and second leading cause in about 2013 in the general population.

Section 6: By-Cause Death Rates Modeling

The purpose of this section is to develop and analyze the death rates projections using a relational model.

6.1 NATIONAL DEATH RATES

The death rates of the baseline population (national population) $\mu_{k,a,g,y}^{nat}$ have been defined in the previous SOA project "Modeling and Forecasting Cause-of-Death Mortality" and have been taken from this previous project. The authors adapted the Lee-Carter model (1992) to cause-of-death modeling in a multivariate framework. In this model, the age a time mortality surface of each cause k and each gender g is decomposed into a static age function $\alpha_{k,a,g}$ (cause-specific age structure), a time series $\kappa_{k,g,y}$ driving the cause-specific mortality stochastic evolution, and an age sensitivity parameter $\beta_{k,a,g}$ that captures the sensitivity of the age class a to the cause k specific increase or decrease.

In formula, the logarithm of the mortality rate is given as follows:

$$\ln(\mu_{k,a,g,y}) = \alpha_{k,a,g} + \beta_{k,a,g}\kappa_{k,g,y}$$

The model is calibrated using the likelihood method, assuming that the number of deaths follows a Poisson distribution, see Brouhns et al. (2002).

To project mortality and capture the interaction (correlation) between cause-specific mortality rates, the time series $(\kappa_{k,g,y})_{k,g}$ have been jointly modeled and forecasted using ARIMA processes.

The methodology of calibration is:

- The authors first calibrated the model separately for each gender g and cause-of-death k, allowing in particular to get the time series parameter $\kappa_{k,g,y}$.
- For each cause-of-death, the authors calibrated the $\kappa_{k,g,y}$ dynamics (ARIMA model) by maximum likelihood estimation.
- To take into account the correlation between the different causes-of-death, for each gender, the authors calibrated a correlation matrix based on the residuals obtained during the calibration of the $\kappa_{k,g,y}$ dynamics of each cause-of-death and gender.

Historically, and in many adaptations of the Lee-Carter model according to the literature, the time component $\kappa_{k,g,y}$ $1 \leq k \leq m$ is assumed to follow a random walk with drift. This is an ARIMA(0,1,0) process given by:

$$\kappa_{k,g,y} = \kappa_{k,g,y-1} + \Delta_{k,g} + \varepsilon_{k,g,y}$$

with $\Delta_{k,g}$, the trend parameter (also called drift), modeling the linear trend of the mortality rates and $\varepsilon_{k,g,y} \sim N(0, \sigma_{k,g})$, a white noise, modeling the deviation of mortality rates from the trend. By $\Sigma \in \mathbb{R}^{m*m}$, the authors denoted the correlation matrix $\Sigma = cor(\varepsilon_{k,g,.})$ modeling the dependency between causes for each gender.

With this model, $\kappa_{k,g,y}$ is expected to reflect the mortality trend. For each gender, the calibration of $\Delta_{k,g}$ and $\sigma_{k,g}$ has been done with a maximum likelihood estimation from each cause. The correlation coefficients have then been determined from the residuals $\varepsilon_{k,g,y}$.

The random walk with drift assumption may not be satisfied over all the historical periods, this is why the trend parameters $\Delta_{k,g}$ have been calibrated on a more recent historical period for some causes, after validation by a breakpoint detection method (see "Modeling and Forecasting Cause-of-Death Mortality" for more details).

6.2 RELATIONAL MODEL LEVERAGING FORECASTS

A relational model allows capturing the age and time mortality pattern of a sub-population ("insured population" obtained by clustering) in relation to a reference larger population (national in this paper; obtained from the project "Modeling and Forecasting Cause-of-Death Mortality"). This model has been adapted to a by-cause modeling framework.

For each cause k, age group a, gender g and year y, the relational model assumes that the mortality rate of the cluster population $\mu_{k,a,g,y}^{cluster}$ writes as a function of the mortality rate of the national population $\mu_{k,a,g,y}^{nat}$ as follows:

$$log\left(\mu_{k,a,g,y}^{cluster}\right) \approx \alpha_{k,g,y} \times log\left(\mu_{k,a,g,y}^{nat}\right) + \beta_{k,g,y}$$

The model works well in many situations since the log-linearity with age of the death rate (known as the Gompertz law) is preserved by this linear transformation at the log-scale.

The age range selected would be for every 5 years until 84 years old and the model could allow for a natural extrapolation at higher ages.

In order to determine if the dependence in the year y of the parameters $\alpha_{k,g,y}$ and $\beta_{k,g,y}$ is significant, confidence intervals are calculated for each estimated parameter according to the following formula:

$$CI_{1-\alpha}\left(\alpha_{k,g,y}\right) = \left[\hat{\alpha}_{k,g,y} - s\left(\alpha_{k,g,y}\right)q_{1-\alpha/2}^t; \hat{\alpha}_{k,g,y} + s\left(\alpha_{k,g,y}\right)q_{1-\alpha/2}^t\right]$$

$$CI_{1-\alpha}\left(\beta_{k,g,y}\right) = \left[\hat{\beta}_{k,g,y} - s\left(\beta_{k,g,y}\right)q_{1-\alpha/2}^t; \hat{\beta}_{k,g,y} + s\left(\beta_{k,g,y}\right)q_{1-\alpha/2}^t\right]$$

Where $s\left(\alpha_{k,g,y}\right)$ and $s\left(\beta_{k,g,y}\right)$ are the estimator of the standard deviation of $\alpha_{k,g,y}$ and $\beta_{k,g,y}$ respectively, and $q_{1-\alpha/2}^t$ is the percentile at level $1-\alpha/2$ of the Student's t-distribution with the number of degrees of freedom given as the size of the sample minus one.

With $\alpha = 5\%$, the confidence intervals of the parameters for a given cause are not disjoint for different years. That is, the following inequalities are satisfied, as depicted in Figure 24:

$$max_y\left(\hat{\alpha}_{k,g,y} - s\left(\alpha_{k,g,y}\right)q_{1-\alpha/2}^t\right) \leq min_y\left(\hat{\alpha}_{k,g,y} + s\left(\alpha_{k,g,y}\right)q_{1-\alpha/2}^t\right)$$

$$max_y\left(\hat{\beta}_{k,g,y} - s\left(\beta_{k,g,y}\right)q_{1-\alpha/2}^t\right) \leq min_y\left(\hat{\beta}_{k,g,y} + s\left(\beta_{k,g,y}\right)q_{1-\alpha/2}^t\right)$$

Figure 24

ALPHA COEFFICIENT AND CONFIDENCE INTERVALS FOR FEMALES AND THE CAUSE CARDIOVASCULAR
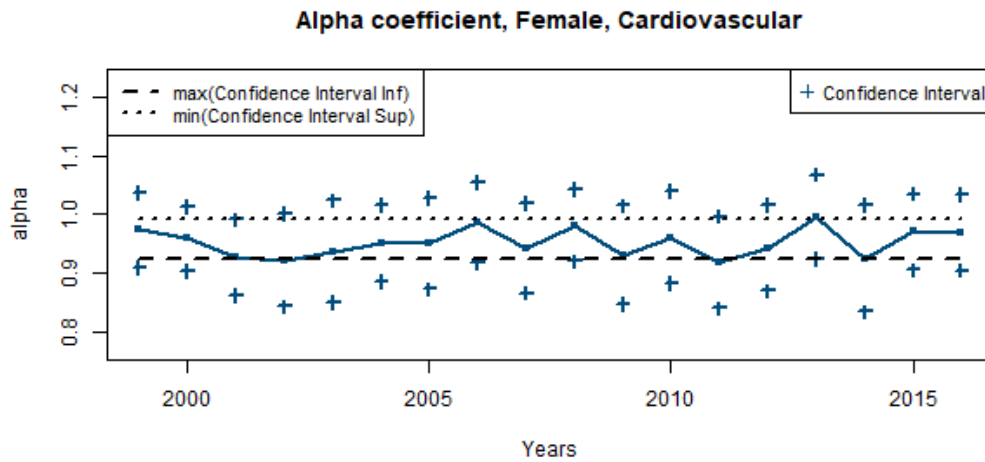
Therefore, for projection purposes, it has been assumed that the coefficients are not time dependent: $\alpha_{k,g,y} = \alpha_{k,g}$ and $\beta_{k,g,y} = \beta_{k,g}$. Thus, there are only two parameters to be estimated for each cause-of-death k and each gender g.

The relational model parameters estimated for the cluster (which represents the cluster population) are shown in Table 8.

Table 8

PARAMETERS FOUND FOR THE CLUSTER POPULATION

Parameter	$\alpha_{k,g}$		$\beta_{k,g}$	
Cause-of-Death\Gender	Male	Female	Male	Female
Cardiovascular	0.965	0.950	−0.558	−0.735
Cerebrovascular	0.927	0.964	−0.829	−0.616
NeoSmok	0.963	0.983	−0.559	−0.322
Neoplasm	0.984	1.006	−0.215	0
Dementia	0.940	0.928	−0.625	−0.784
Diabetes	0.983	0.978	−0.399	−0.544
Influenza	0.935	0.940	−0.753	−0.802
Respiratory	0.906	0.914	−1.093	−1.075
Drug	0.924	0.929	−0.661	−0.765
External	1.116	1.121	0.440	0.591
Other	0.981	0.980	−0.207	−0.334

The fact that $\alpha_{k,g} < 1$ means that death rates increase (or decrease) less for cluster population than for the national population for a specific cause k and gender g. It is the case for all causes, except neoplasm for females and external for both males and females. The fact that $\beta_{k,g} \neq 0$ (observed for all causes, except neoplasm for females) means that apart from death rate improvement, an additional (positive or negative) shift is observed between the death rates of the cluster population and the adjusted death rates of the national population.

In order to perform forecasts, we will rely on the starting value of the death rate for the cluster population at year 2017, then compute the improvement rates given by the relational model as follows:

$$log\left(\frac{\mu_{k,a,g,y}^{cluster}}{\mu_{k,a,g,y-1}^{cluster}}\right) \approx \alpha_{k,g} \times log\left(\frac{\mu_{k,a,g,y}^{nat}}{\mu_{k,a,g,y-1}^{nat}}\right).$$

Thus, the cluster population death rates can be recursively computed according to the following equation, which makes use of the general population mortality forecasts:

$$\mu_{k,a,g,y}^{cluster} \approx \mu_{k,a,g,y-1}^{cluster}\left(\frac{\mu_{k,a,g,y}^{nat}}{\mu_{k,a,g,y-1}^{nat}}\right)^{\alpha_{k,g}}.$$

As the alpha parameters are close to 1, the evolutions of the death rates for the cluster are expected to be similar to those of the national population.

Section 7: By-Cause Death Rates Forecasting

In this section, the results of the death rates projections are displayed and analyzed. Given the data availability at the time of the study, the last year of historical data is 2016 for the U.S. national population and 2017 for the cluster population (richest counties). For the models used in this report, the trajectories are forecasted 10 years ahead as the authors believe that the models achieve better short-term results. Therefore, a reasonable limit for realistic projections may be 2026 and the forecasts presented relate to the period 2017–2026.

The following figures focus especially on three age ranges: 30–34, 50–54 and 80–84, these age ranges reflecting different times in life. If we consider all age ranges between 50 and 84 years, the death rates have decreased historically, except for causes dementia, drug and external. Globally, the forecasts between 50 and 84 years are decreases of death rates, but slower than the historical decreases. The death rate for cause drug is expected to continue to increase. The death rate for cause dementia is also expected to continue to increase, but slower than it did in the past.

Table 9
ANNUAL EVOLUTION OF DEATH RATES (AVERAGE IN AGE RANGES 50–84 YEARS)

Cause-of-Death	Males				Females			
	Cluster		National		Cluster		National	
	Historical	Forecast	Historical	Forecast	Historical	Forecast	Historical	Forecast
Cardiovascular	−3.4%	−0.9%	−2.8%	−1.0%	−3.7%	−1.3%	−2.9%	−1.4%
Cerebrovascular	−2.9%	−0.9%	−2.4%	−1.0%	−3.2%	−1.2%	−2.6%	−1.2%
NeoSmok	−4.3%	−2.1%	−2.2%	−2.2%	−2.9%	−1.0%	−1.1%	−1.1%
Neoplasm	−2.0%	−0.6%	−1.1%	−0.6%	−1.9%	−0.6%	−1.2%	−0.6%
Dementia	2.3%	1.3%	3.4%	1.4%	3.5%	1.4%	4.4%	1.5%
Diabetes	−0.3%	−0.1%	0.2%	−0.1%	−1.7%	−0.4%	−0.8%	−0.4%
Influenza	−2.5%	−0.2%	−0.9%	−0.3%	−2.3%	0.0%	−0.5%	0.0%
Respiratory	−2.4%	−0.7%	−0.8%	−0.8%	−1.5%	0.1%	−0.1%	0.1%
Drug	2.6%	2.1%	1.9%	2.2%	2.5%	3.3%	3.3%	3.6%
External	0.6%	−0.1%	0.3%	−0.1%	0.4%	0.0%	0.2%	0.0%
Other	−1.2%	0.1%	0.2%	0.1%	−1.1%	0.0%	0.2%	0.0%
Aggregate	*−2.0%*	*−0.2%*	*−1.3%*	*−0.4%*	*−1.9%*	*−0.1%*	*−1.0%*	*−0.2%*

Three main points should be noticed to understand the discrepancies between the historical and forecast trends for both the cluster and the general population:

- Because of the intrinsic volatility of the random realization of deaths (sampling risk), the historical trend is only an estimate of the underlying "true" trend.

- The relational model is not time-dependent and thus historical evolutions of death rates for the cluster population are limited to that of the general population based on the α parameter of the relational model; as such, other trend discrepancies between the cluster and the general population are not

captured. Recall that the time-independent relational model has been statistically justified in Section 7, while extending this framework to a time dependent setting remains challenging given the limited number of deaths for each cause in the cluster, compared to a national population.

- The general population trend, which is also used and adjusted to forecast the cluster-specific mortality, is captured on the historical data after the last trend change, using breakpoint detection methods, see the previous SOA project "Modeling and Forecasting Cause-of-Death Mortality" for more details. This is a core source of explanation of the severe forecast scenarios, such as the high increase of death rates for the cause drug, since they reflect the most recent observations.

We may notice that the national population trend in the forecast period is close to 0. Therefore, the death rates forecasted for the cluster population will be similar to those of the national population in terms of evolution over time. Overall, it is noticed that the relational model is vulnerable to the fact that the national population trends are small and that the alpha parameters are close to 1, although statistically the recourse to time-varying coefficients in the relational model cannot be justified.

7.1 AT AGES 30–34

The results of the forecasts for the cluster population with an initial date in 2017 are plot below in Figure 25.

Figure 25
DEATH RATES FORECASTS BY CAUSE AT AGES 30–34
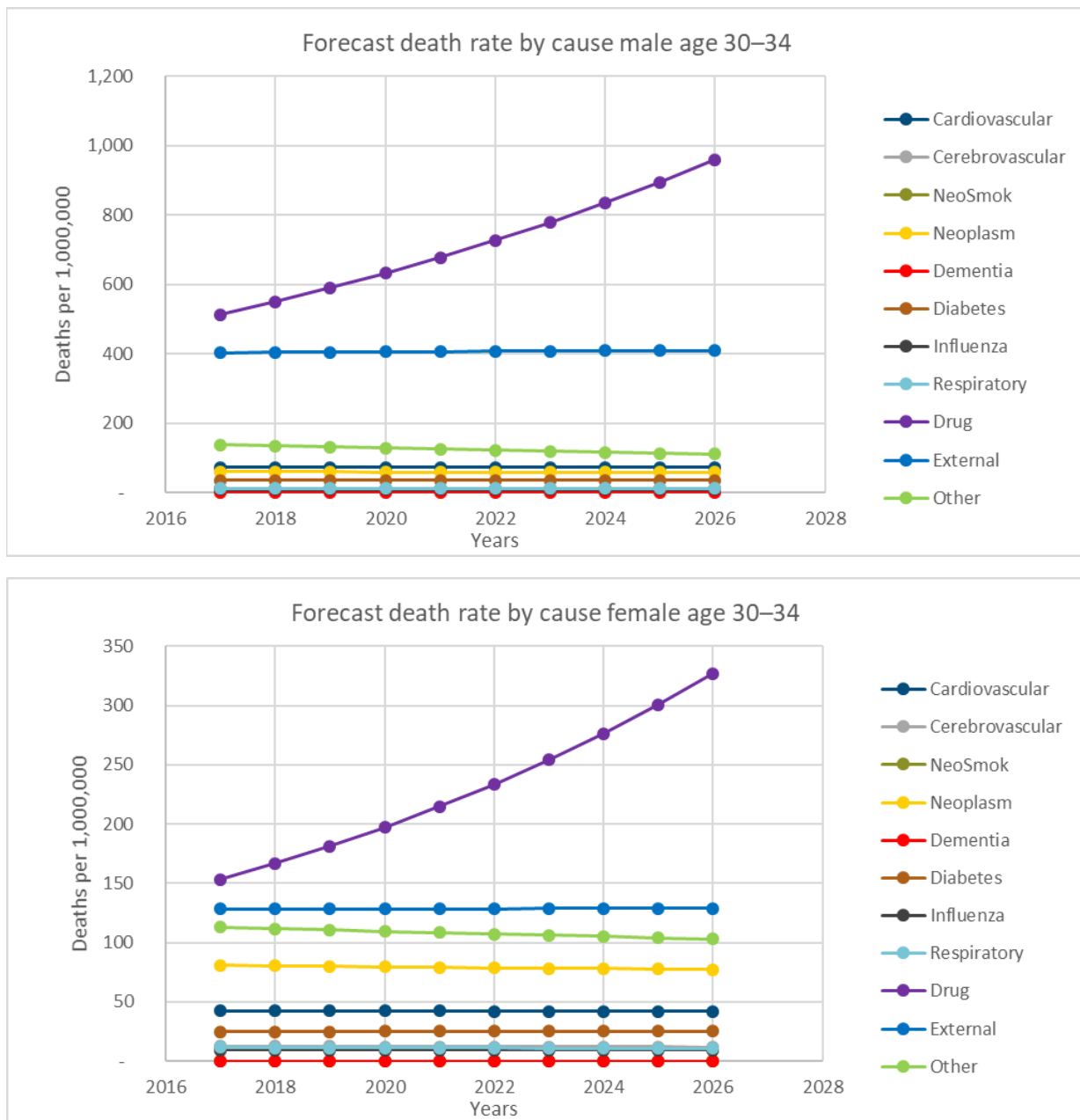
The cause drug at ages 30–34 is the first cause-of-death in 2017 for both males and females, and its death rates are expected to highly increase during the next ten years, widening the gap with the other causes-of-death. This expectation is due to the high increase of the historical rates for this cause during the last five years (momentum effect). The death rates of the other causes-of-death are expected to remain stable, or to decrease a little. For some causes-of-death, the death rates are very low at ages 30–34 (e.g. dementia).

The forecasts of the cause-specific death rates lead to resulting forecasts for the aggregate (all causes) death rate for age band 30–34 depicted in Figure 26.

Figure 26
AGGREGATE DEATH RATES FORECASTS AT AGES 30–34
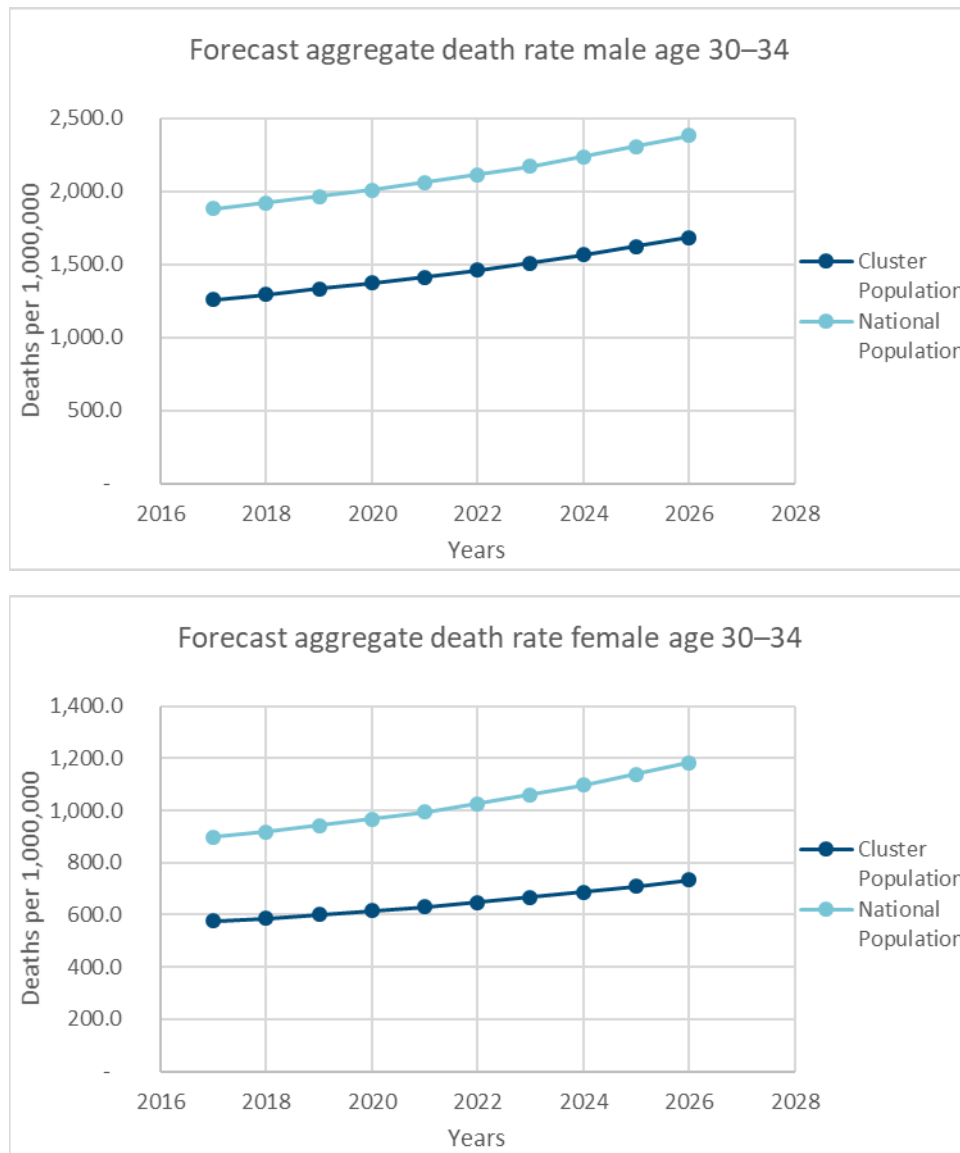
The aggregate death rates at ages 30–34 show that the death rates are 40% higher for the national population than they are for the cluster population. Interestingly, these differences are expected to be stable for the male population, whereas the gap is expected to widen for the female population, which shows a higher level for the general population. Overall, the aggregate death rates are expected to increase at ages 30–34.

Figure 27

COMPARISON DEATH RATES FORECASTS FOR DRUG ABUSE / CLUSTER VS NATIONAL AT AGES 30–34
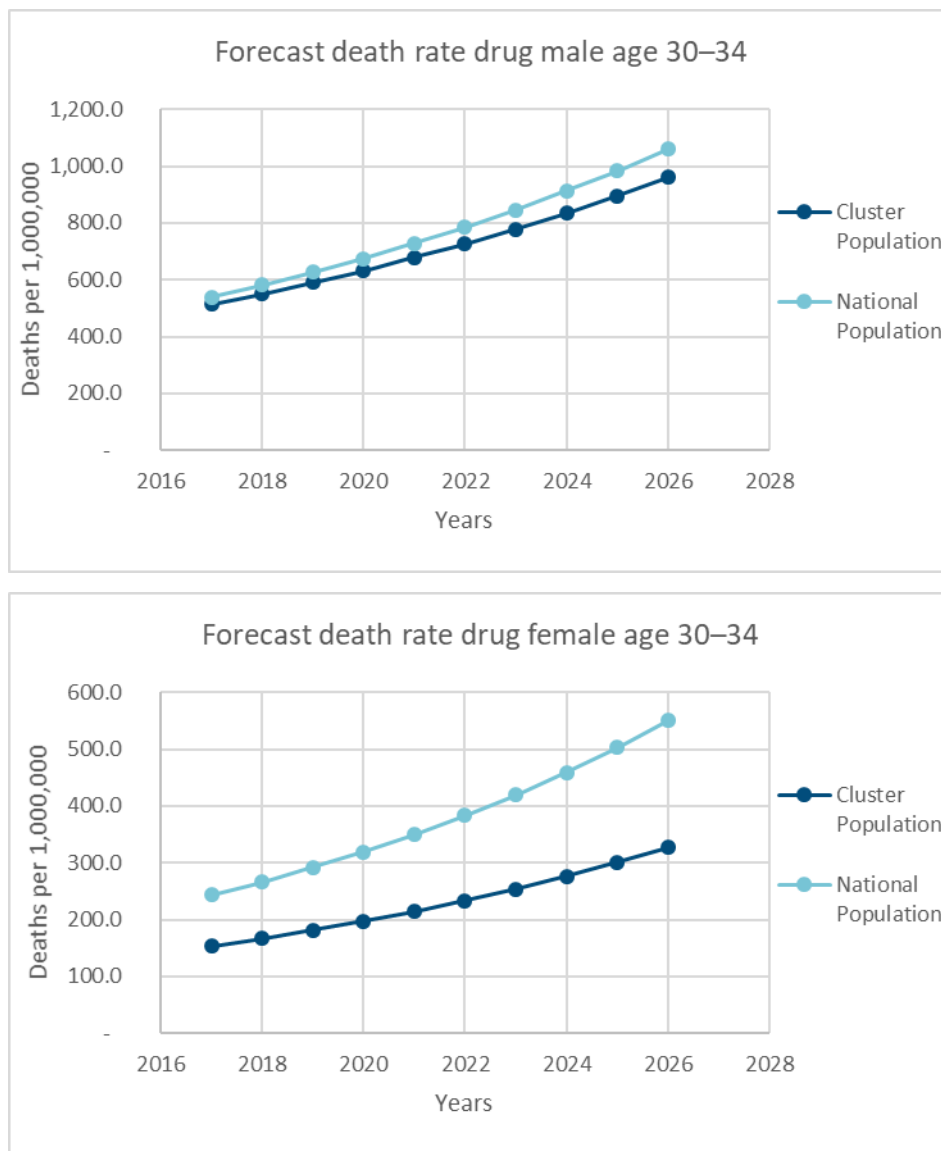
The death rates for cause drug at ages 30–34 are expected to increase, but to increase slower for the cluster population than for the national population. The gap between the cluster population and the national population is much higher for females than it is for males. Note that this is driven by a difference in the initial level of the death rates, to which improvement rates of similar order of magnitude are applied. Indeed, the alpha coefficients of the relational model are close to 1, see Table 8, but the increases and the final gaps are different because of the difference of the initial levels to which the relative improvement rates are applied.

Figure 28
COMPARISON DEATH RATES FORECASTS FOR EXTERNAL / CLUSTER VS NATIONAL AT AGES 30–34
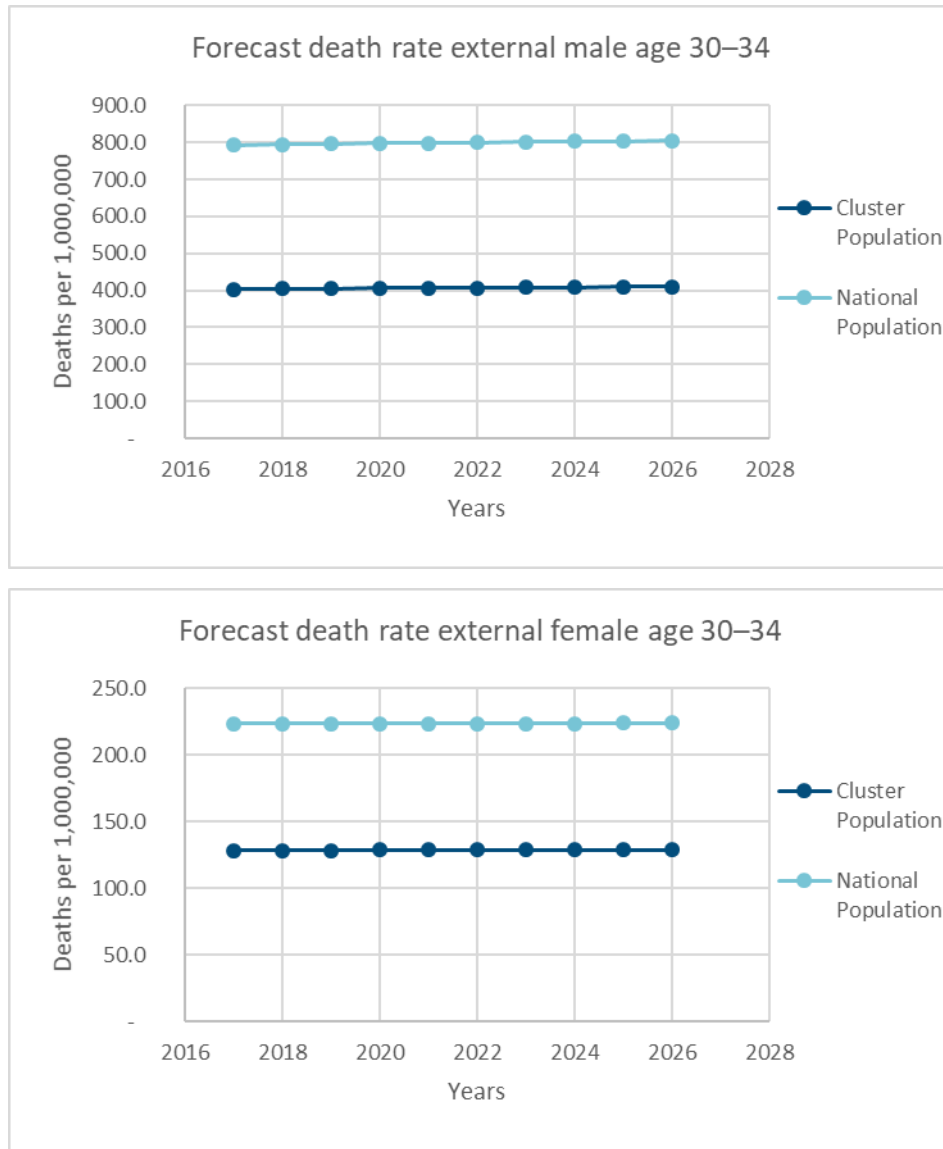
The death rates for cause external at ages 30–34 are expected to remain stable for both males and females. These death rates are lower for the cluster population than for the national population and the gaps are expected to remain stable.
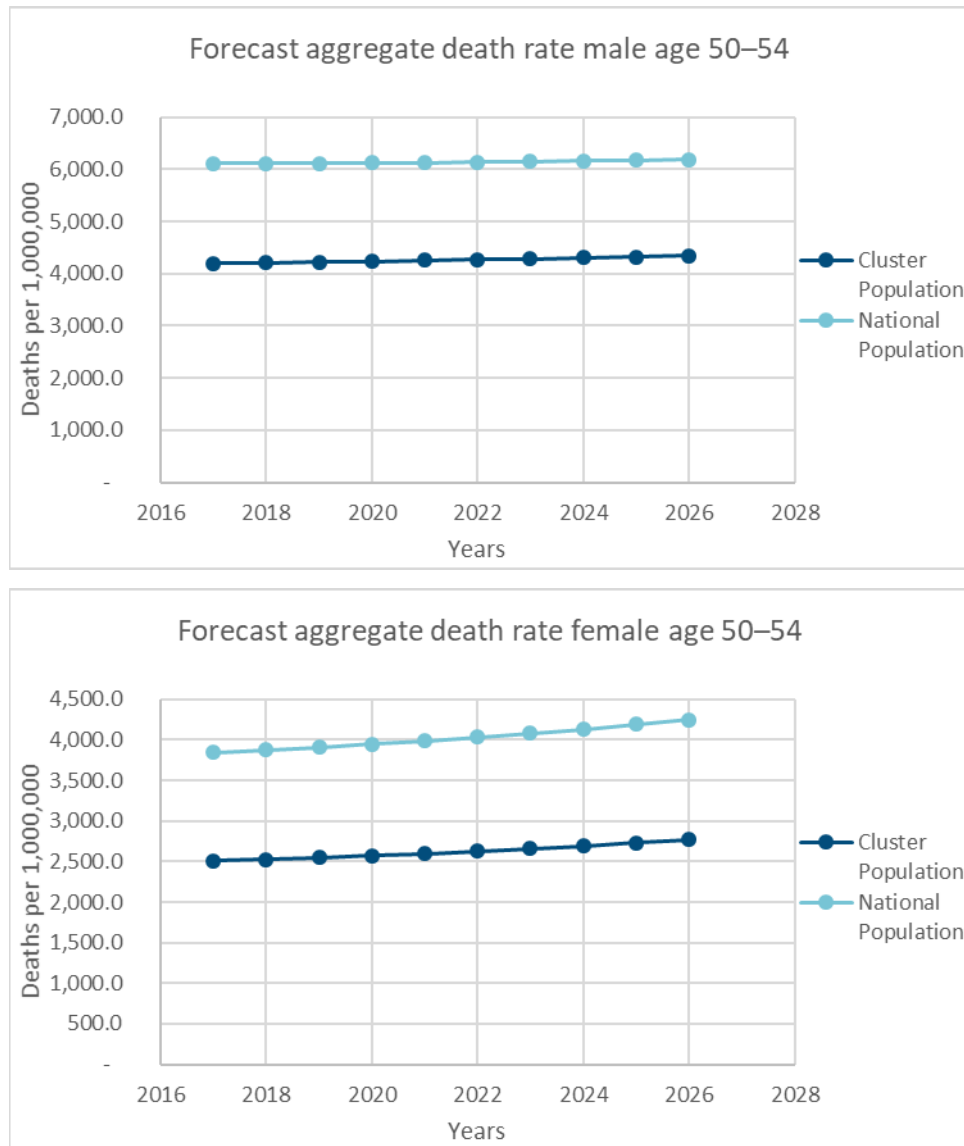
7.2 AT AGES 50–54

Figure 29
DEATH RATES FORECASTS BY CAUSE AT AGES 50–54

Forecast death rate by cause male age 50–54

Forecast death rate by cause female age 50–54
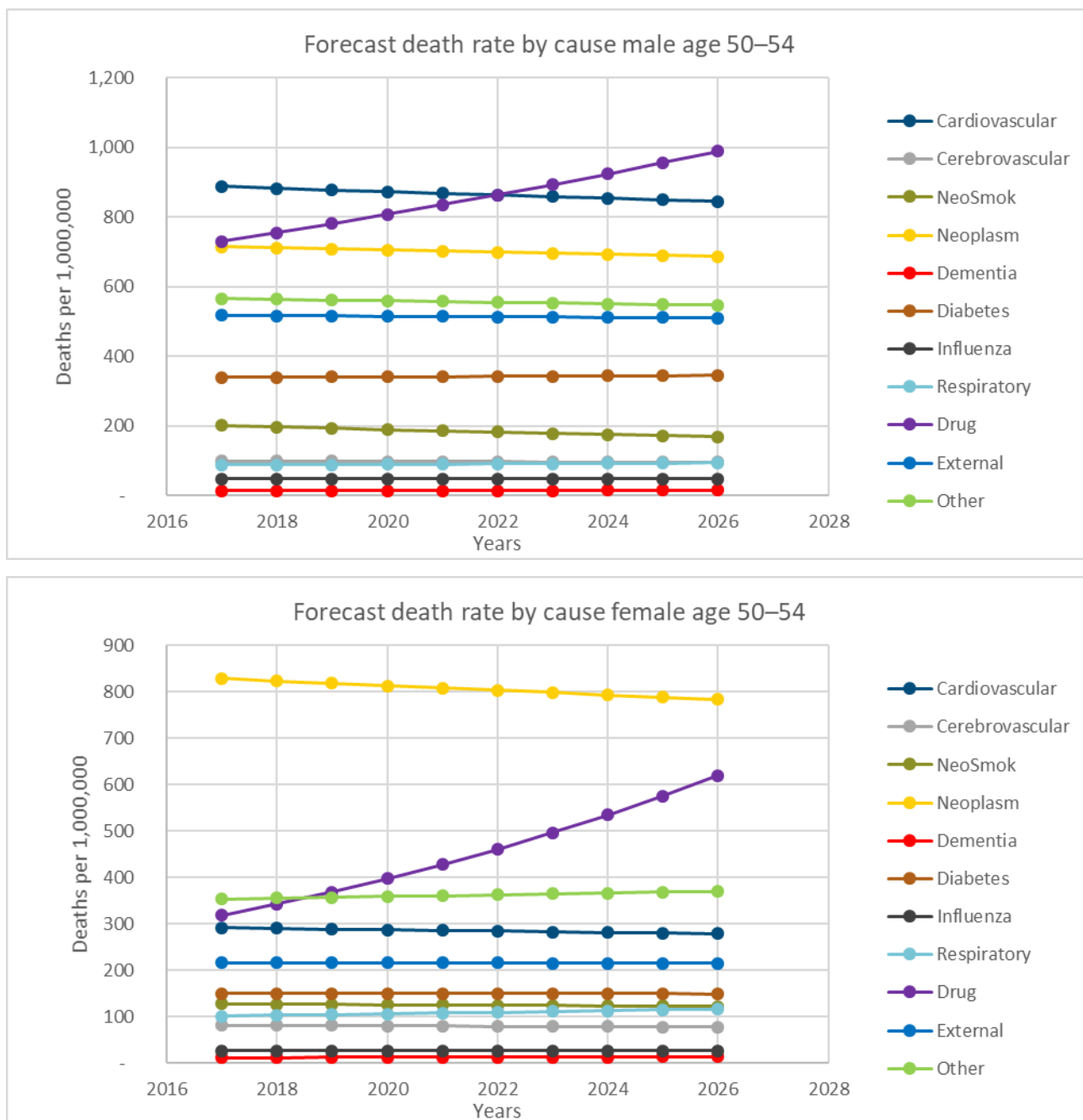
The cause cardiovascular at ages 50–54 is the first cause-of-death for males in 2017. However, it is expected to become the second cause-of-death after 2023. The cause drug would become the first cause-of-death for males. For the female population, the cause neoplasm is the first cause-of-death in 2017 and it is expected to remain. However, the neoplasm death rate is expected to decrease whereas the drug death rate is expected to increase which would make the drug the second cause-of-death in 2019. The death rates of the other causes-of-death are expected to remain stable or to decrease a little.

Figure 30

AGGREGATE DEATH RATES FORECASTS AT AGES 50–54

Forecast aggregate death rate male age 50–54


Forecast aggregate death rate female age 50–54

The aggregate death rates at ages 50–54 show that the death rates are 50% higher for the national population than they are for the cluster population. These differences are expected to remain for the next ten years for both males and females. The aggregate death rates are expected to increase for females and to remain stable for males.

Figure 31

COMPARISON DEATH RATES FORECASTS FOR DRUG ABUSE / CLUSTER VS NATIONAL AT AGES 50–54
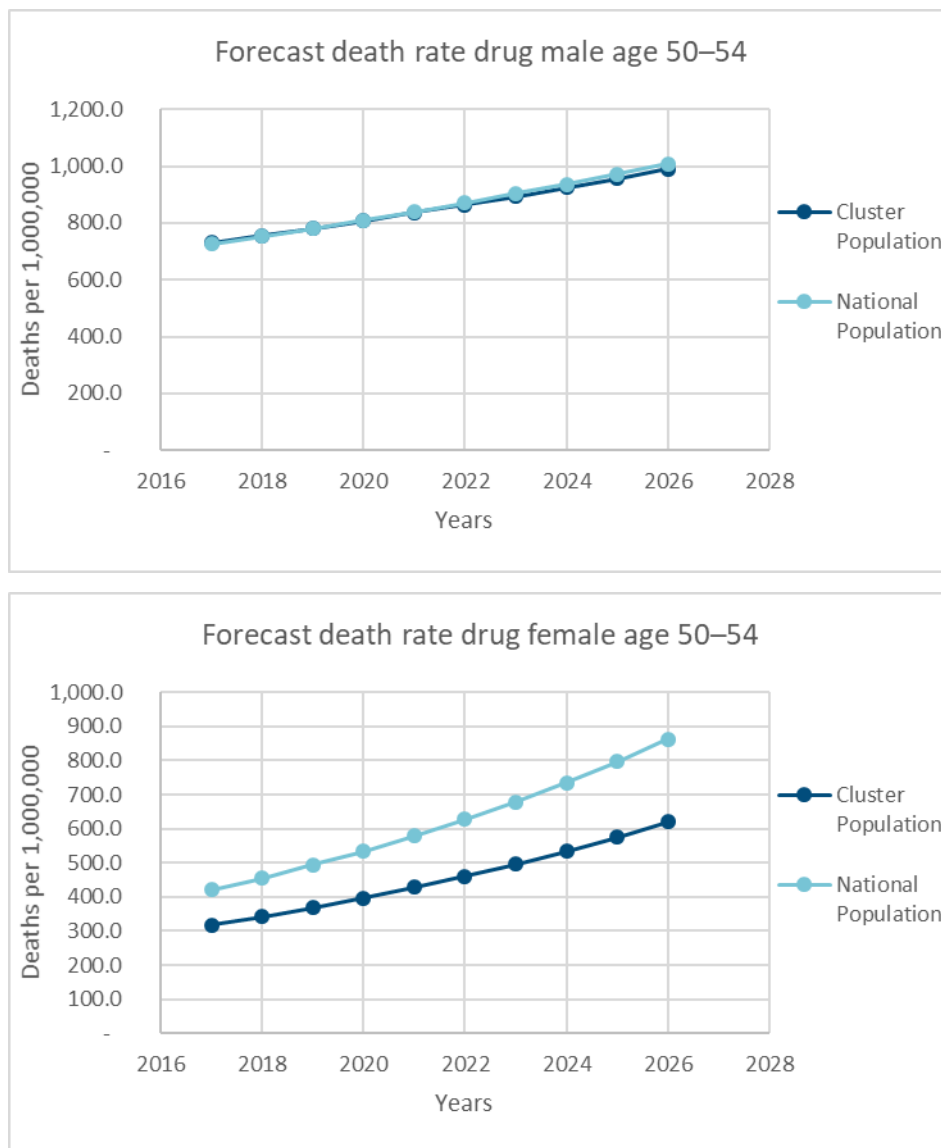
The death rates for cause drug at 50–54 are expected to increase, but to increase slower for the cluster population than for the national population. The gap between the cluster and the national population is significant for females but not for males, while differentials are expected to widen. The fact that the death rates forecasts are similar for the male cluster population and the male national population is counter intuitive. However, we observe that in general, there are less differences in death rates for the cause drug between the two male populations than between the two female populations. Let us also note that the evolutions of the death rates between the cluster and the national population are similar because the alpha coefficients are close to 1 (see Table 8). The absolute increases and the final gaps for the female populations are different because of the difference of the initial levels whereas the levels are rather similar for the male populations. Anyway, an insurer might have a different view of his portfolio's exposure to the drug risks. In such case, it is possible to use the tool provided with the report to implement a different evolution for this specific cause (see Appendix G).

Figure 32
COMPARISON DEATH RATES FORECASTS FOR NEOPLASM / CLUSTER VS NATIONAL AT AGES 50–54

Forecast death rate neoplasm male age 50–54

Forecast death rate neoplasm female age 50–54
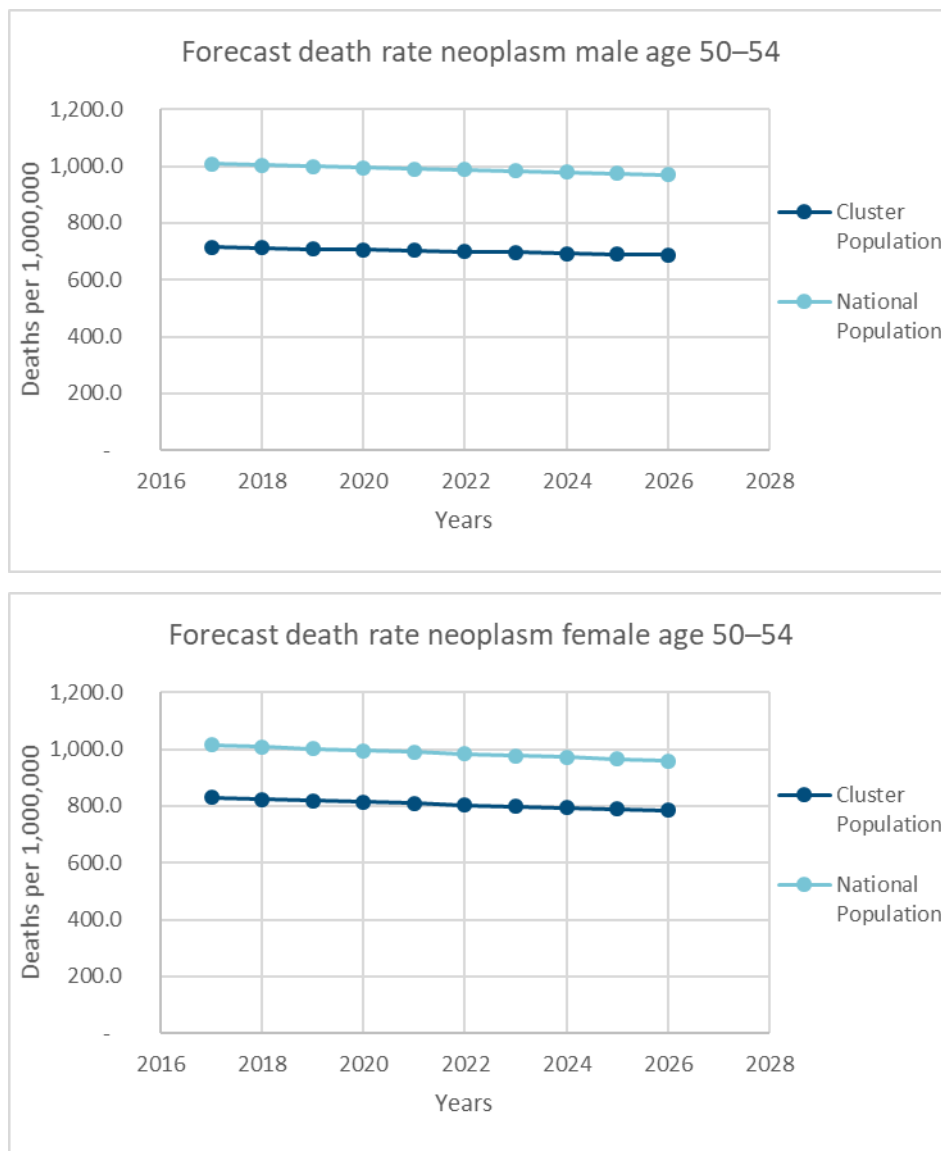
The death rates for cause neoplasm at 50–54 are expected to decrease and these decreases are similar for males and females. The gap between the cluster and the national population is higher for males than it is for females. Differentials are expected to be stable as the decreasing trends of the death rates are similar for the cluster and the national population.

Figure 33
COMPARISON DEATH RATES FORECASTS FOR CARDIOVASCULAR / CLUSTER VS NATIONAL AT AGES 50–54
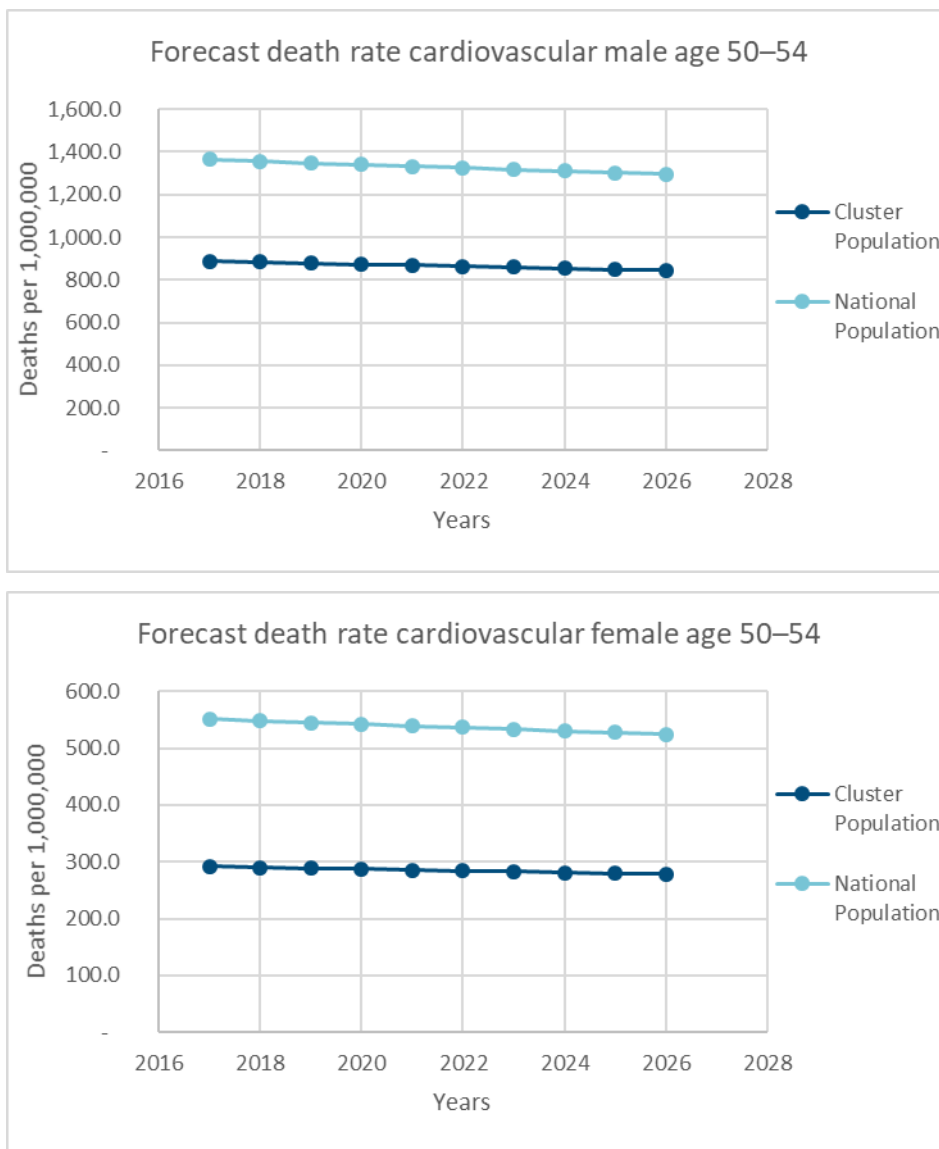
The death rates for cause cardiovascular at 50–54 are expected to decrease and these decreases are similar for both males and females. The gap between the cluster and the national population is more than 50% for males and more than 70% for females. Differentials are expected to be stable.
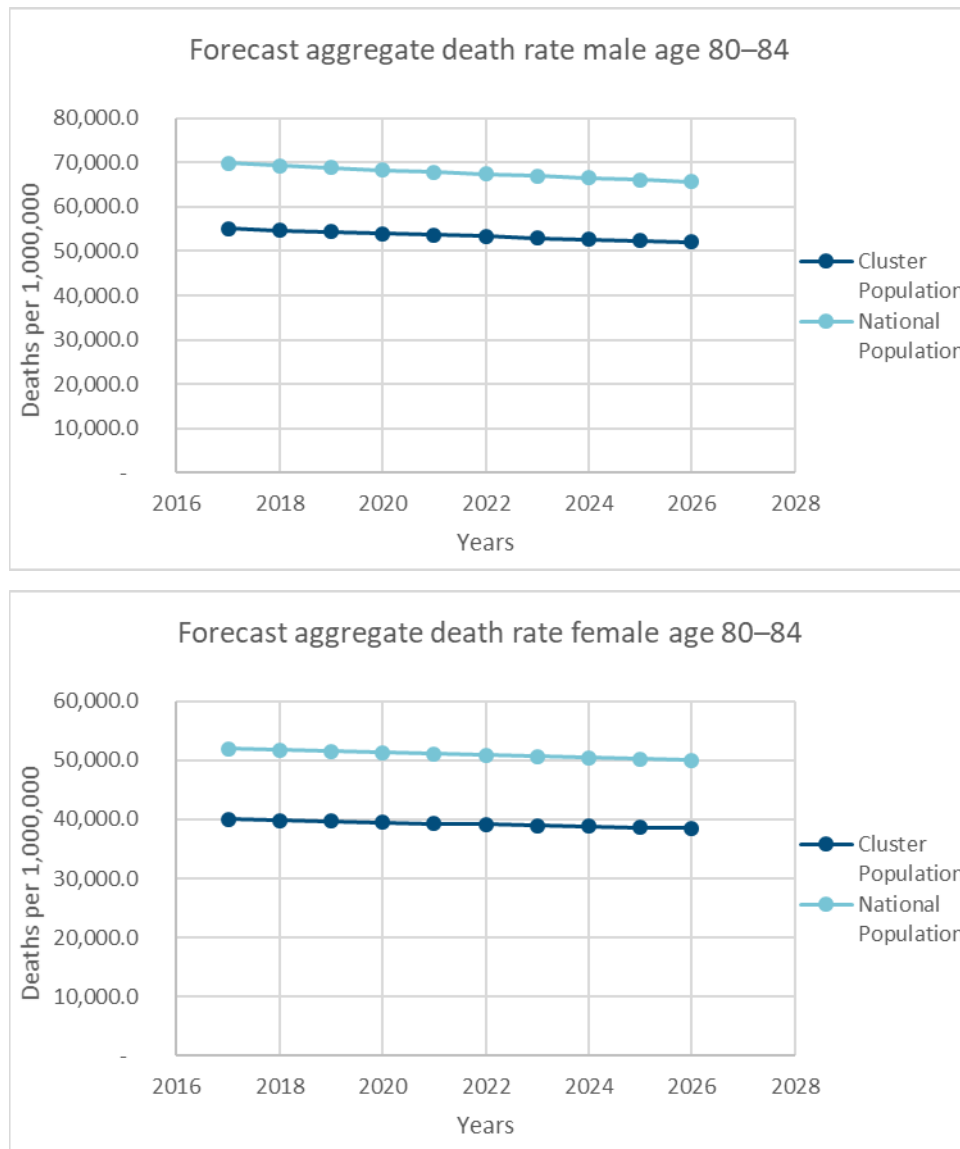
7.3 AT AGES 80–84

Figure 34
DEATH RATES FORECASTS BY CAUSE AT AGES 80–84

Forecast death rate by cause male age 80–84

Forecast death rate by cause female age 80–84
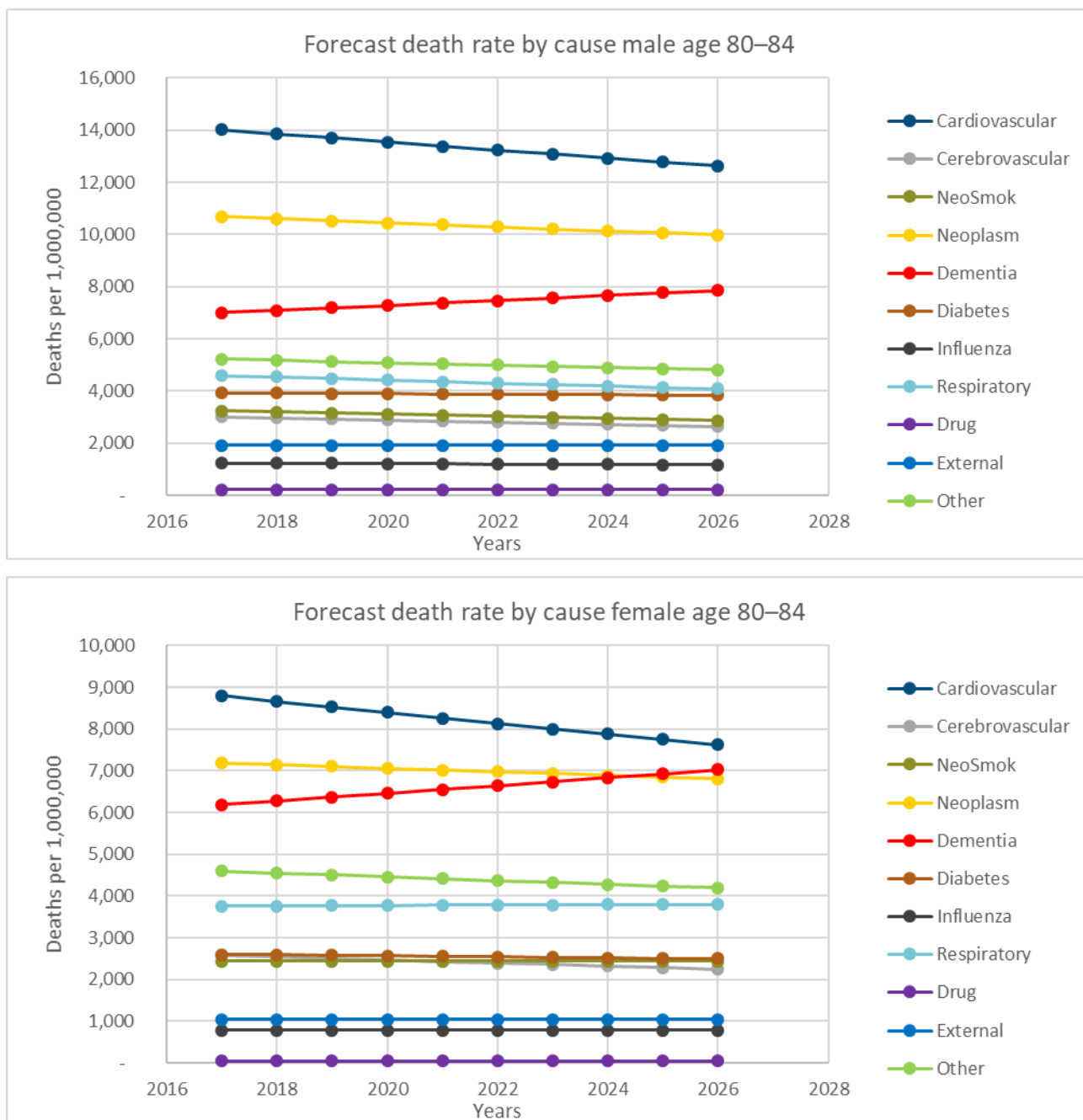
The cause cardiovascular at ages 80–84 is the first cause-of-death for both males and females in 2017. However, its death rate expected to decrease with time. The death rate for cause neoplasm would decrease but neoplasm would remain the second cause-of-death for males, whereas the cause dementia would become the second cause-of-death for females in 2025. The death rates of the other causes-of-death are expected to remain stable or to decrease a little.

Figure 35

AGGREGATE DEATH RATES FORECASTS AT AGES 80–84

The aggregate death rates at ages 80–84 show that the death rates are 25% higher for the national population than they are for the cluster population. These differences are expected to remain for the next ten years, even if the aggregate death rates are expected to decrease a little.

Figure 36

COMPARISON DEATH RATES FORECASTS FOR CARDIOVASCULAR / CLUSTER VS NATIONAL AT AGES 80–84
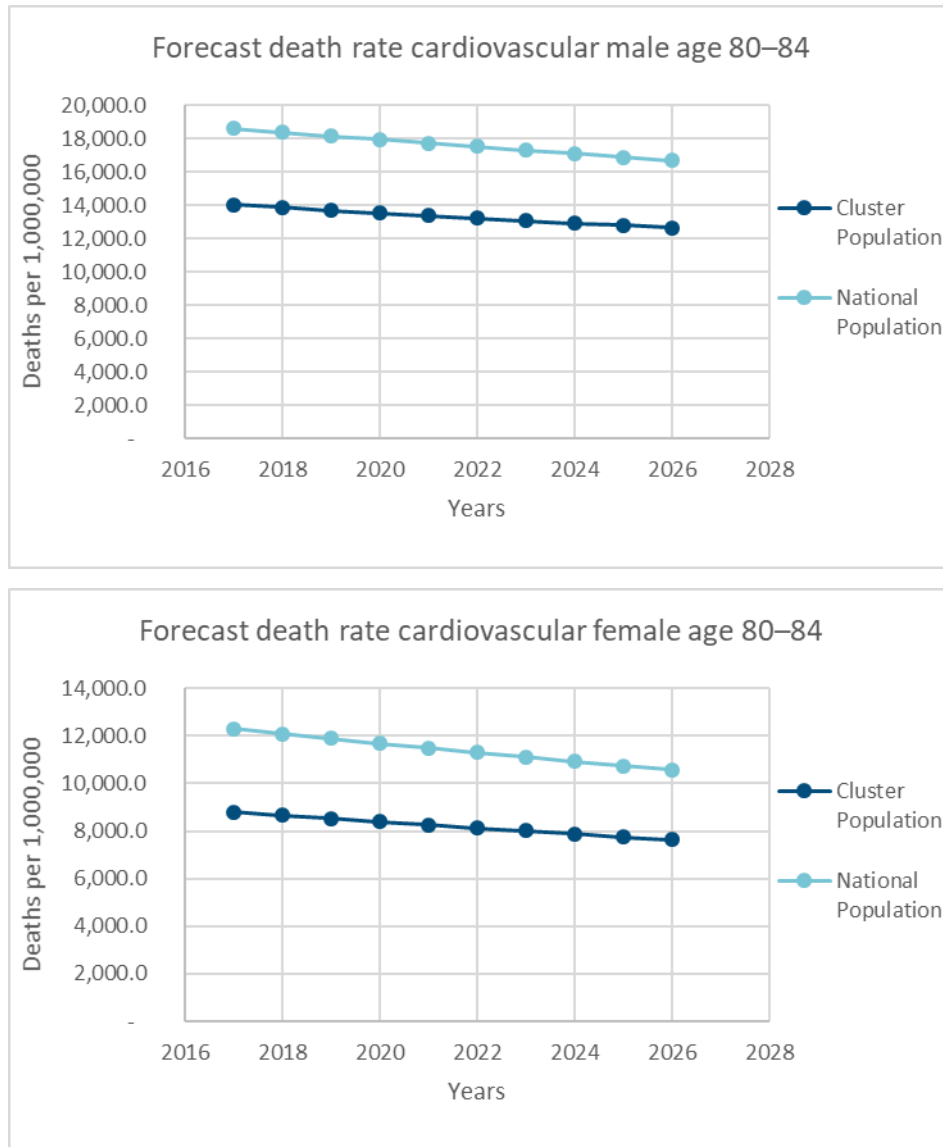

The death rates for cause cardiovascular at 80–84 are expected to decrease and these decreases are similar for males and females. The gap between the cluster and the national population is approximately 25% for both males and females.

Figure 37
COMPARISON DEATH RATES FORECASTS FOR NEOPLASM / CLUSTER VS NATIONAL AT AGES 80–84
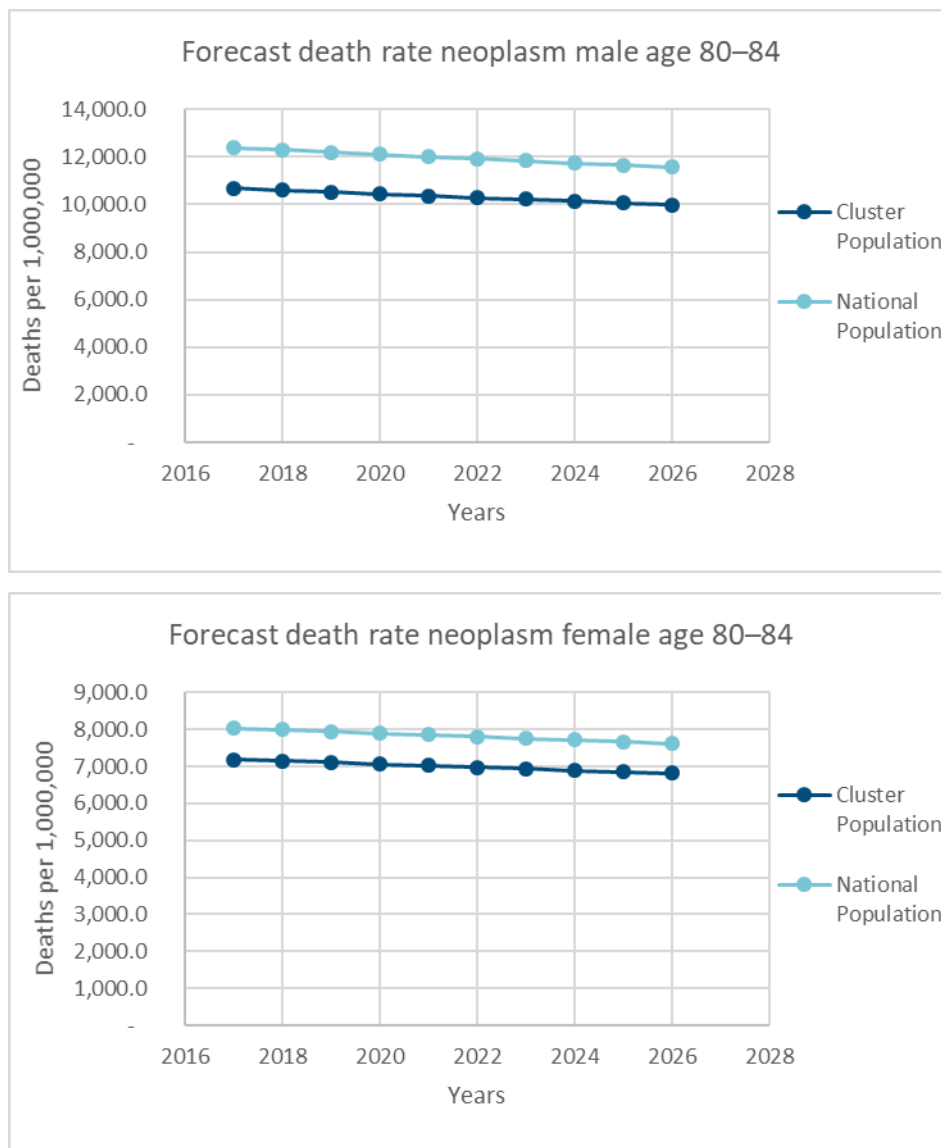
The death rates for cause neoplasm at 80–84 are expected to decrease and these decreases are similar for males and females. The gap between the cluster and the national population is approximately 15% for both males and females.

Figure 38

COMPARISON DEATH RATES FORECASTS FOR DEMENTIA / CLUSTER VS NATIONAL AT AGES 80–84
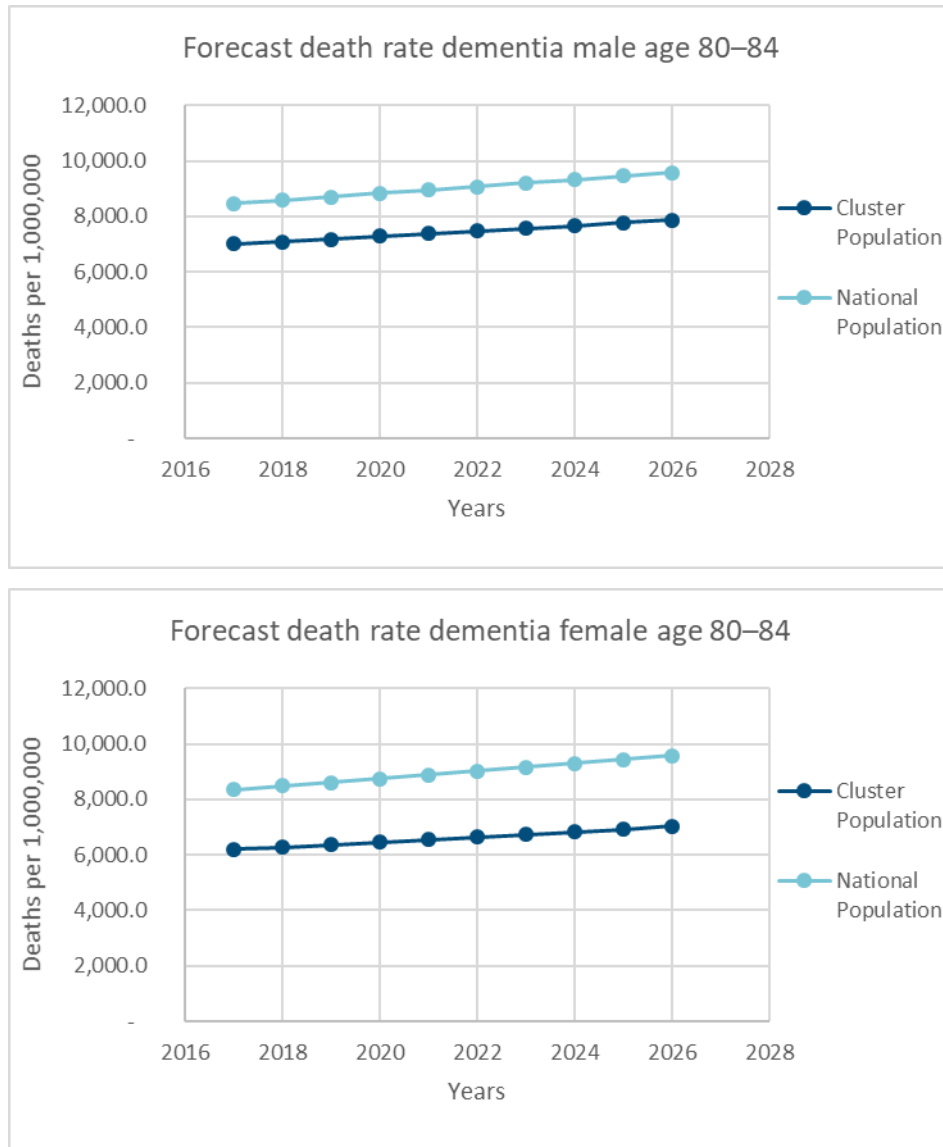
The death rates for cause dementia at 80–84 are expected to increase. The gap between the cluster and the national population is expecting to increase a little. This gap is approximately 15% for males and 25% for females.

Section 8: Life Expectancies Forecasts

The life expectancies forecasts are displayed and analyzed now in this section.

8.1 LIFE EXPECTANCIES CALCULATION

In this study a partial life expectancy calculation is used in order to compare the life expectancies for the national population and for the cluster population. We denote by Y the lifetime of an individual, bounded by some limit age a_{max}. We denote by Y_a the remaining lifetime of an individual with age a, and by ${}_yp_a = P(Y_a > y)$ the survival function starting at age a, that is ${}_yp_a$ is the probability that an individual with age a lives at least y years more.

Then, the following identity holds:

$$ {}_yp_a = P(Y > a + y | Y \geq a) = \frac{P(Y > a + y)}{P(Y \geq a)} = \frac{{}_{a+y}p_0}{{}_ap_0}. $$

We finally denote by ${}_yq_a = 1 - {}_yp_a = P(Y_a < y)$ the probability to die within the next y years for an individual aged a.

Let $e_{a/b}^{partial}$ be the partial life expectancy between ages a and b, written as:

$$ e_{a/b}^{partial} = \int_{y=0}^{b-a} {}_yp_a dy = \sum_{i=0}^{b-a-1} \int_{y=i}^{i+1} {}_yp_a dy = \sum_{i=0}^{b-a-1} \int_{y=i}^{i+1} {}_ip_a \; {}_{y-i}p_{a+i} dy = \sum_{i=0}^{b-a-1} {}_ip_a \int_{y=0}^{1} {}_yp_{a+i} dy $$

Denoting $q_a = {}_1q_a$ the probability of death during the year and using a uniform repartition of deaths among the year, there is for $y \in [0;1]$ ${}_yq_a = yq_a$ which implies:

$$ {}_yp_{a+i} = 1 - {}_yq_{a+i} = 1 - yq_{a+i} $$

Then,

$$ e_{a/b}^{partial} = \sum_{i=0}^{b-a-1} {}_ip_a \int_{y=0}^{1} (1 - yq_{a+i}) dy = \sum_{i=0}^{b-a-1} {}_ip_a \left(1 - \frac{q_{a+i}}{2}\right) = \sum_{i=0}^{b-a-1} {}_ip_a - \frac{1}{2} \sum_{i=0}^{b-a-1} {}_ip_a q_{a+i} $$

Finally,

$$ e_{a/b}^{partial} = \sum_{i=1}^{b-a-1} {}_ip_a + 1 - \frac{1}{2} \sum_{i=0}^{b-a-1} {}_ip_a q_{a+i} $$

With $q_{a+i} = P(Y < a + i + 1 \mid Y \geq a + i) = \frac{{}_ip_a - {}_{i+1}p_a}{{}_ip_a}$ and $q_{a_{max}} = 1$ by definition of the limit age.

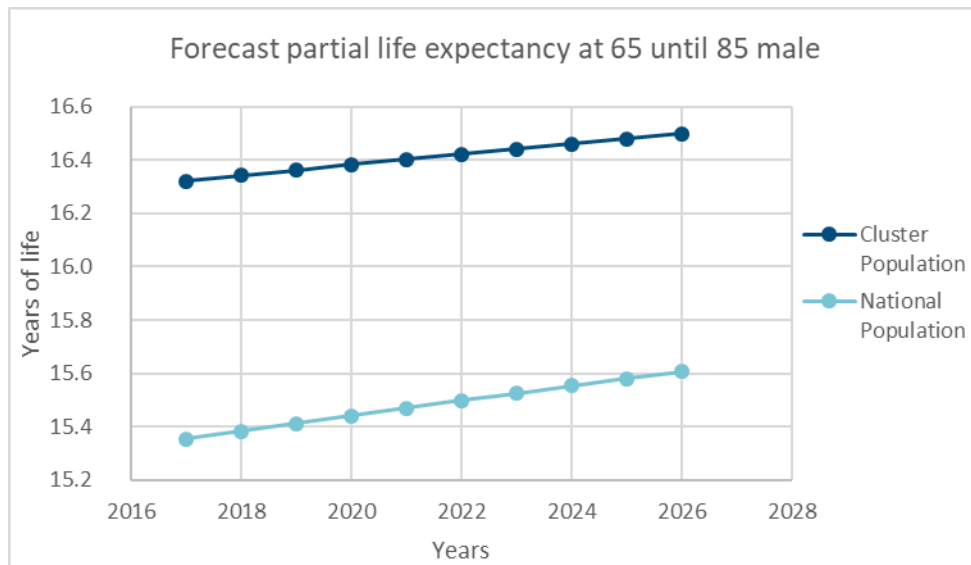Thus, life expectancy at birth, with $a = 0$,

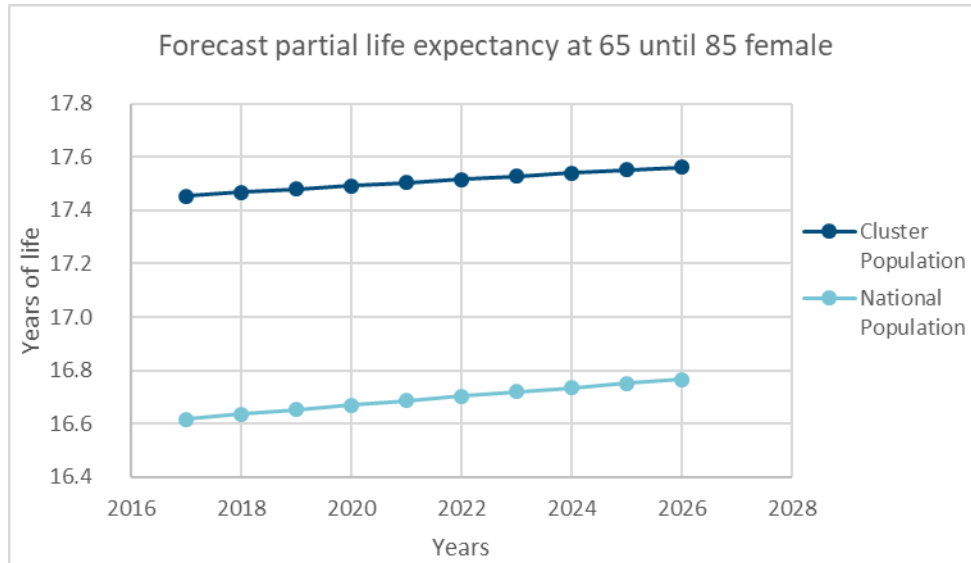$$e_{0/b}^{partial} = \sum_{i=1}^{b-1} {}_i p_0 + 1 - \frac{1}{2}\sum_{i=0}^{b-1} {}_i p_0 q_i$$

The authors will use these equations to calculate the partial life expectancy at birth until 85 years old and the partial life expectancy at 65 until 85 years old.

8.2. PARTIAL LIFE EXPECTANCIES FORECASTS

The partial life expectancies forecasts at 65 until 85 have been computed for males and females and are expected to increase. These partial life expectancy are 1 year higher for the cluster population than they are for the national population for males and 0.8 year higher for females. This confirms that the cluster (and richer) population has a higher life expectancy than the national population, and this difference is expected to remain for the future years. The partial life expectancies at birth until 85 are detailed in Appendix C, but the results are less clear due to the increase of the drug cause at young ages.

Figure 39
PARTIAL LIFE EXPECTANCY FORECASTS AT 65 UNTIL 85

Forecast partial life expectancy at 65 until 85 female

(chart)

- Cluster Population
- National Population

Y-axis: Years of life (16.4 to 17.8)
X-axis: Years (2016 to 2028)
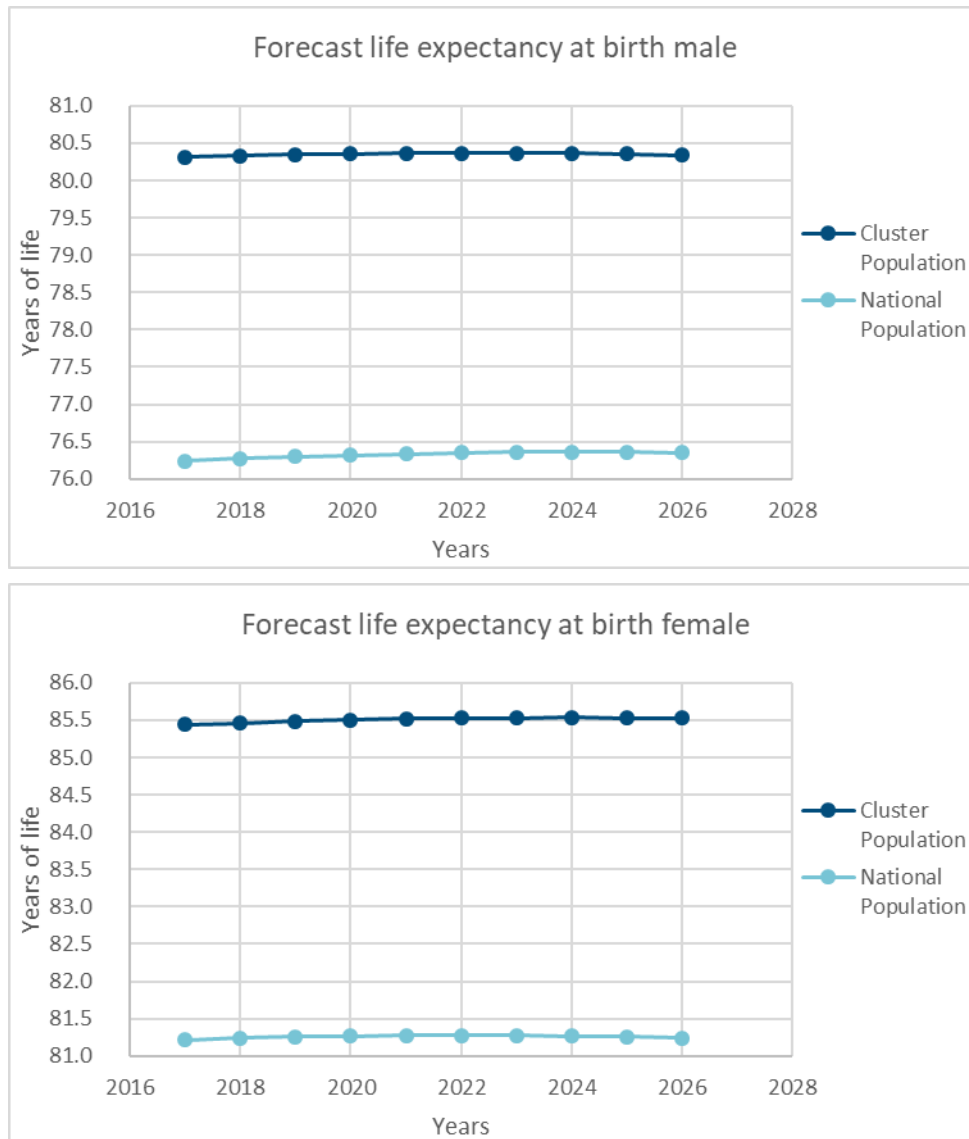
8.3 TOTAL LIFE EXPECTANCIES FORECASTS

The mortality rates after 84 in 2017 are extrapolated from the relational model:

$$log\left(\mu_{k,a,g,y}^{cluster}\right) \approx \alpha_{k,g,y} \times log\left(\mu_{k,a,g,y}^{nat}\right) + \beta_{k,g,y}$$

Then, it is possible to deduce total life expectancies at birth and at 65 years old.

$$e_a^{total} = \sum_{i=1}^{a_{max}-a-1} {}_ip_a + 1 - \frac{1}{2}\sum_{i=0}^{a_{max}-a-1} {}_ip_a q_{a+i} = \sum_{i=1}^{a_{max}-a-1} {}_ip_a + \frac{1}{2}$$

Figure 40
TOTAL LIFE EXPECTANCY FORECASTS AT BIRTH

The total life expectancy at birth is expected to stagnate or to increase very little for both males and females because of the high historical increase of deaths for the cause drug. This hypothesis may not happen in practice if deaths because of drug do not continue to increase as fast as it used to be in the past (see Appendix E related to CDC death forecasts up to June 2019 for drug overdose). The life expectancy at birth of the cluster population is 4 years higher for males than the life expectancy at birth of the national population, and this difference reaches 4.5 years for females.

Figure 41
TOTAL LIFE EXPECTANCY FORECASTS AT 65
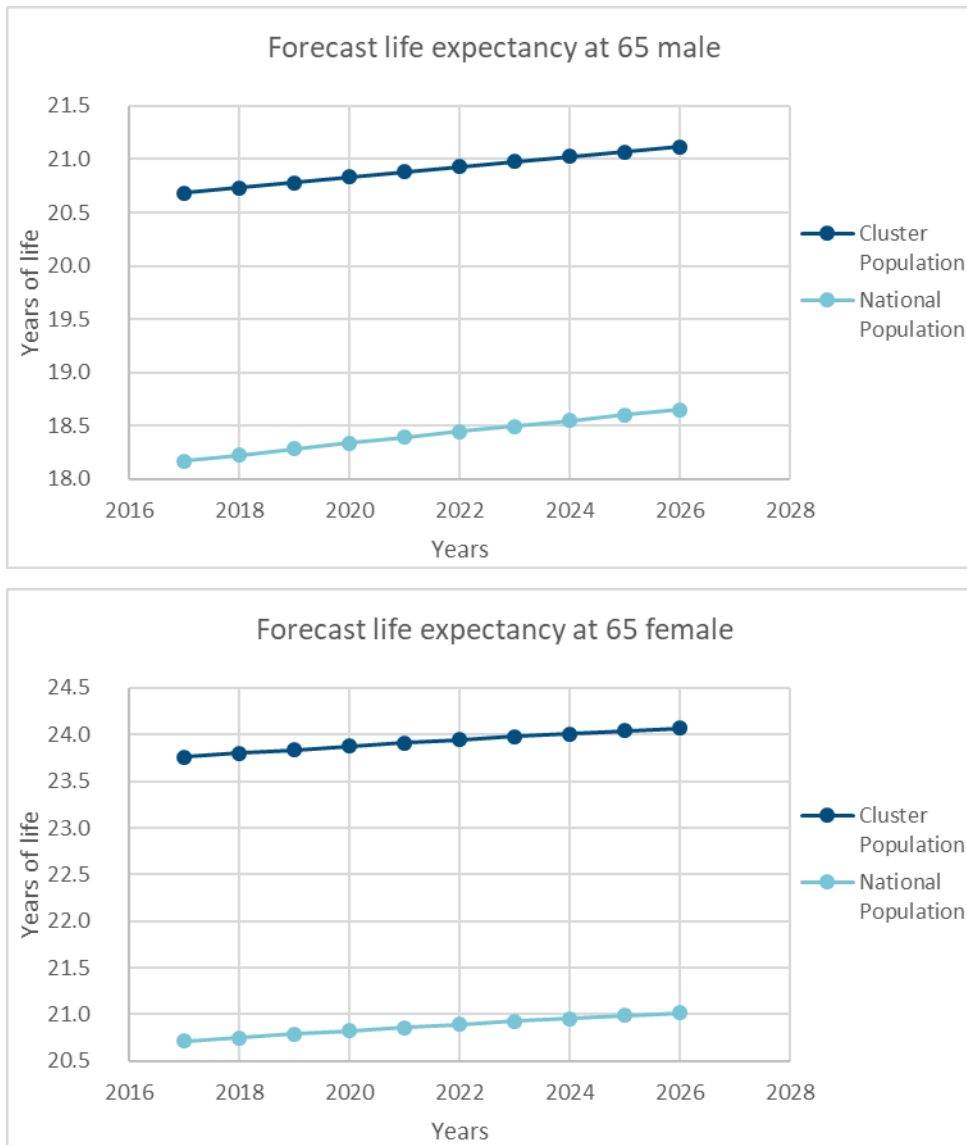
The total life expectancy at 65 is expected to continue to increase for both males and females, but a bit slower than it did in the past. For males, the life expectancy at 65 of the cluster population is 2.5 years higher than the life expectancy at birth of the national population, and for females, the difference reaches 3 years.

Section 9: Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group and others for their insights, advice, guidance, and diligent work reviewing and editing this report for accuracy and relevance.

Project Oversight Group Members:

Jim Filmore, Chairperson, FSA, MAAA

Mary Pat Campbell, FSA, MAAA

Henry Egesi, FSA

Sophie (Lemeng) Feng, ASA, ACIA, CERA

Jean-Marc Fix, FSA, MAAA

Blake Hill, FSA, FCIA

Jason McKinley, FSA

Larry Stern, FSA, MAAA

Mark Walker, FSA, MAAA

Daniel Zamarripa, MD

At the Society of Actuaries:

R. Dale Hall, FSA, CERA, MAAA, Managing Director of Research

Jan Schuh, Senior Research Administrator

Ronora Stryker, ASA, MAAA, Senior Practice Research Actuary

Appendix A: Counties of the Cluster Approximating An Insured Population

The counties of the cluster approximating an insured population are shown in Table 10.

Table 10
COUNTIES OF THE CLUSTER APPROXIMATING AN INSURED POPULATION

State	County
Arkansas	Benton County
California	Santa Clara County
California	Marin County
California	San Francisco County
California	San Mateo County
Colorado	Denver County
Connecticut	Fairfield County
District of Columbia	District of Columbia
Florida	Collier County
Florida	Indian River County
Florida	Palm Beach County
Florida	Sarasota County
Florida	Martin County
Florida	Monroe County
Georgia	Fulton County
Massachussetts	Dukes County
Massachussetts	Middlesex County
Massachussetts	Nantucket County
Massachussetts	Suffolk County
Minnesota	Hennepin County
New Jersey	Somerset County
New Jersey	Monmouth County
New Jersey	Morris County
New Jersey	Bergen County
New York	Westchester County
New York	Nassau County
New York	New York County
Ohio	Delaware County
Pennsylvania	Montgomery County
Rhode Island	Bristol County
South Dakota	Lincoln County
Iowa	Dallas County
Indiana	Boone County
Massachussetts	Norfolk County
South Dakota	Union County
Tennessee	Williamson County
Texas	Kendall County
Texas	Midland County
Virginia	Goochland County
Virginia	Alexandria County
Virginia	Arlington County
Washington	King County
Wyoming	Teton County

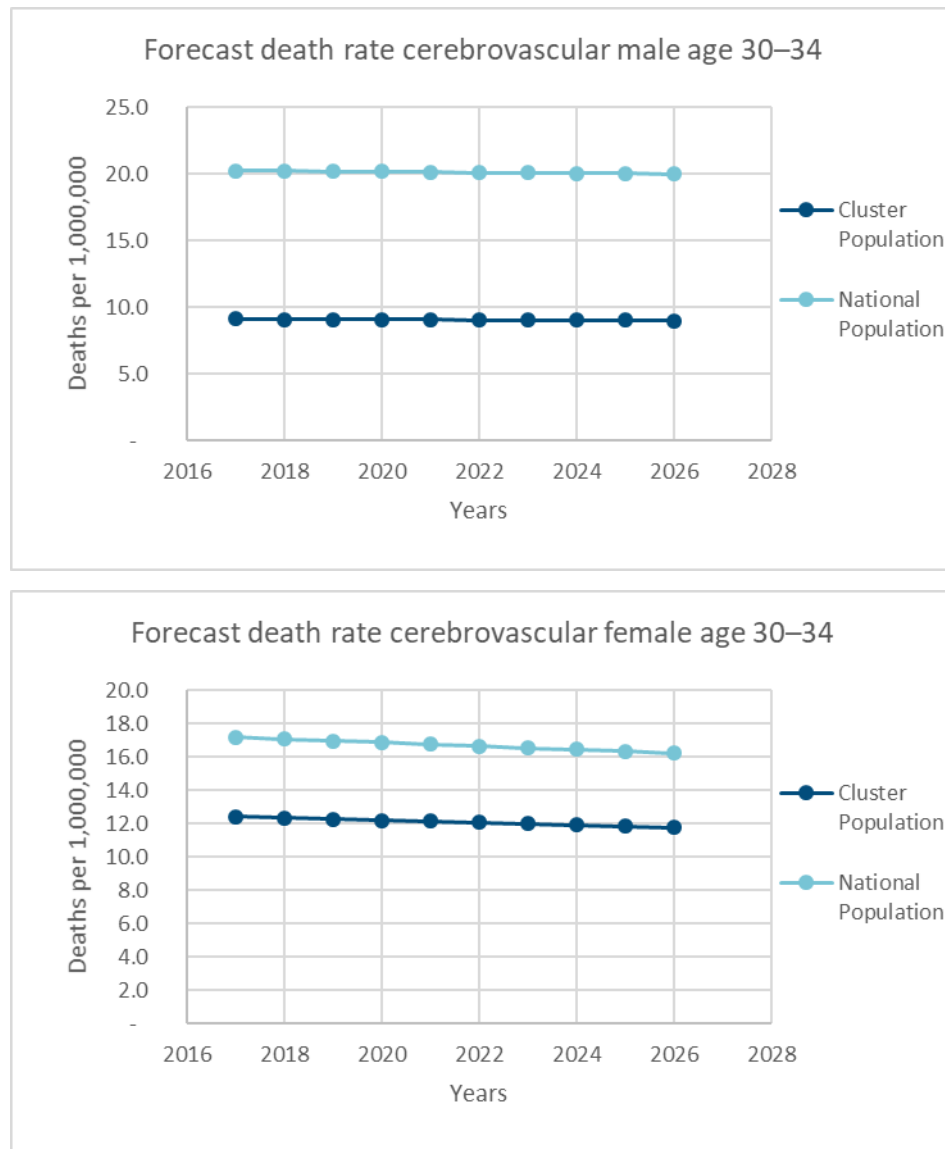# Appendix B: By-Cause Death Rates Forecasting

B.1 AT AGES 30–34

Figure 42
COMPARISON DEATH RATES FORECASTS FOR CARDIOVASCULAR / CLUSTER VS NATIONAL AT AGES 30–34
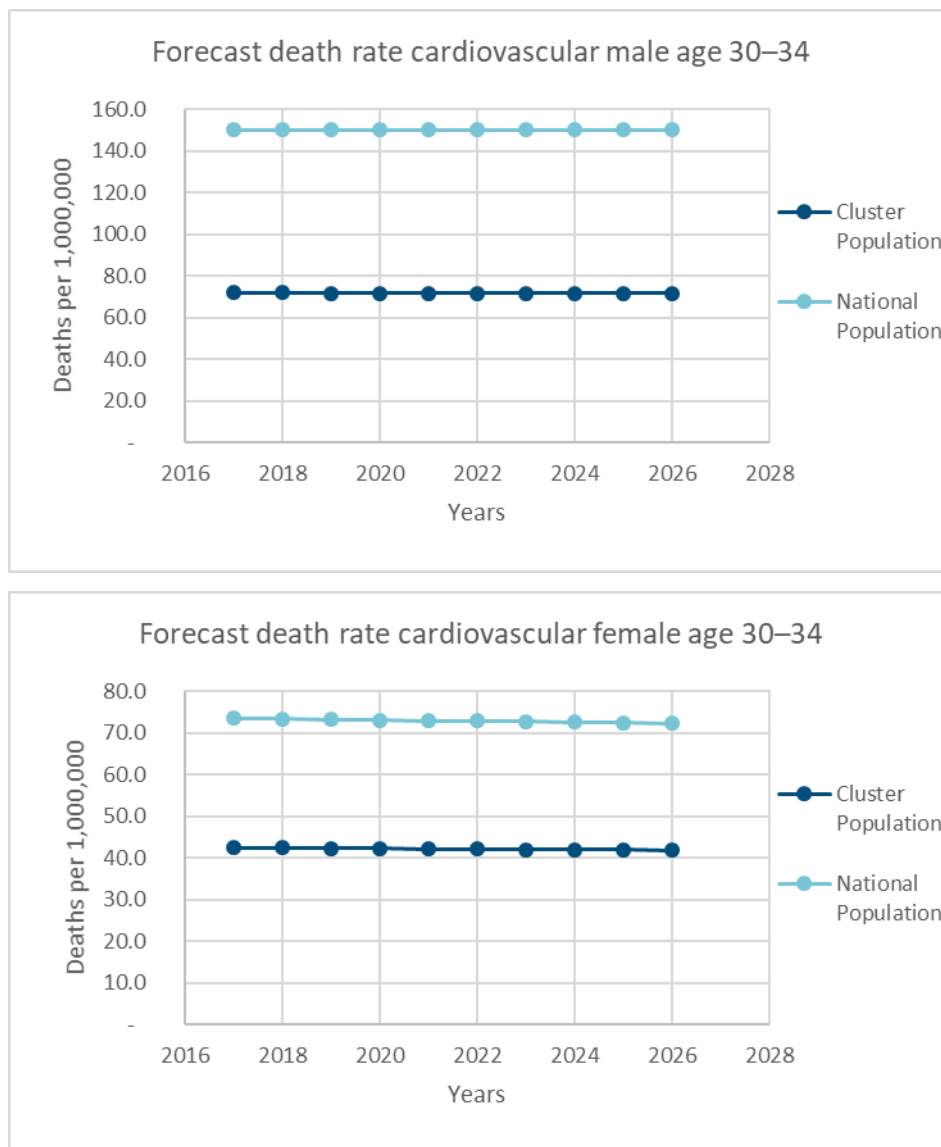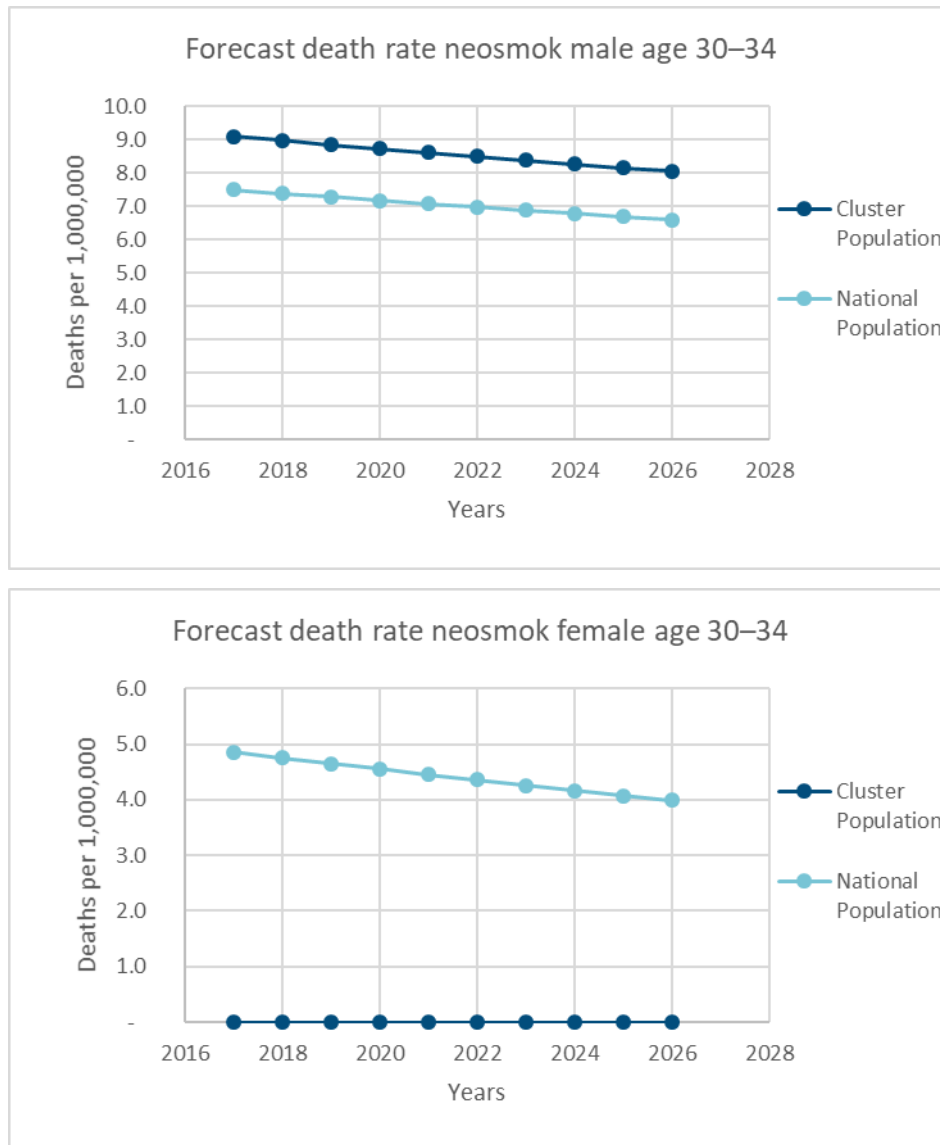

The death rates for cause cardiovascular at ages 30–34 are expected to remain stable. The death rate of the cluster population is below the death rate of the national population and the gap between the two populations is expected to remain stable.

Figure 43

COMPARISON DEATH RATES FORECASTS FOR CEREBROVASCULAR / CLUSTER VS NATIONAL AT AGES 30–34
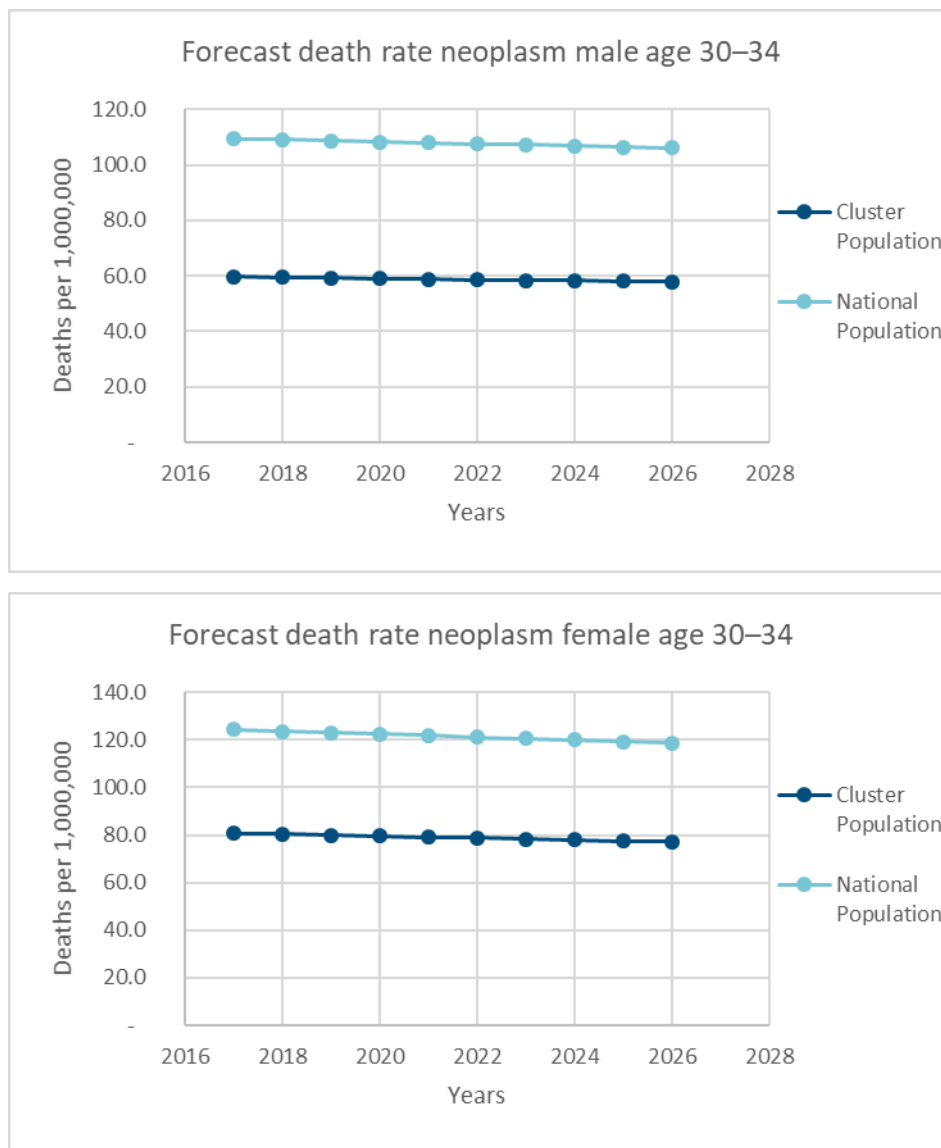

The death rates for cause cerebrovascular at ages 30–34 are expected to remain stable for males, and to decrease a little for females. This death rate is really low. The death rate of the cluster population is below the death rate of the national population and the difference between the two populations is expected to remain.

Figure 44

COMPARISON DEATH RATES FORECASTS FOR NEOSMOK / CLUSTER VS NATIONAL AT AGES 30–34

The death rates for cause neosmok (neoplasm induced by smoking) at ages 30–34 are very low. For males, the death rates of the cluster population are higher than the death rates of the national population and the differences between the two populations are expected to remain. The death rates are expected to decrease. For females, the death rates of the cluster population are almost zero (no deaths were observed in 2017 for the cluster population and few were observed for the national population).

Figure 45

COMPARISON DEATH RATES FORECASTS FOR NEOPLASM / CLUSTER VS NATIONAL AT AGES 30–34

The death rates for cause neoplasm at ages 30–34 are expected to remain stable or to decrease a little. The death rate of the cluster population is below the death rate of the national population and the difference between the two populations is expected to remain.

Figure 46
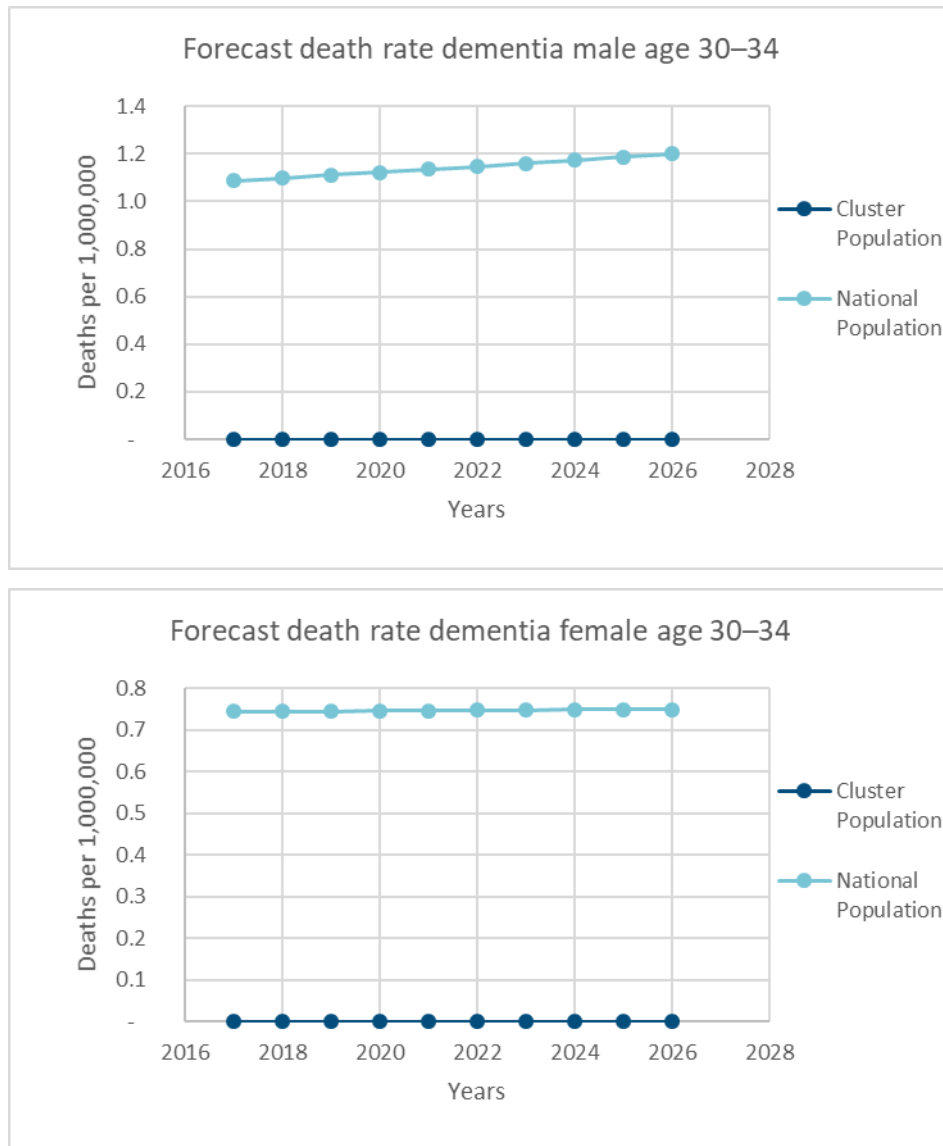COMPARISON DEATH RATES FORECASTS FOR DEMENTIA / CLUSTER VS NATIONAL AT AGES 30–34

The death rates of the cluster population for cause dementia at ages 30–34 are almost zero. Thus, it is difficult to forecast the evolution of this cause-of-death at such ages.

Figure 47

COMPARISON DEATH RATES FORECASTS FOR DIABETES / CLUSTER VS NATIONAL AT AGES 30–34
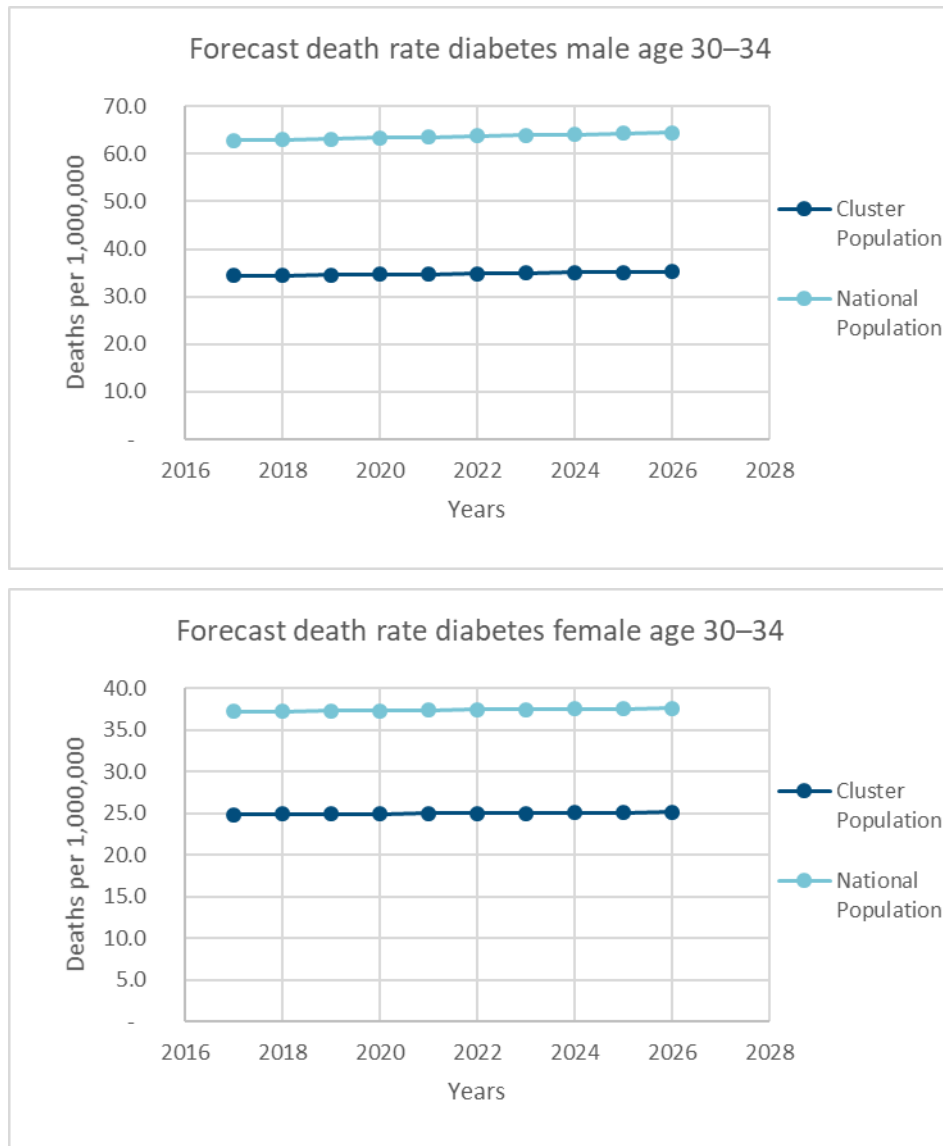
The death rates for cause diabetes at ages 30–34 are expected to remain stable. These death rates are lower for the cluster population than for the national population for both males and females, and these differences are expected to remain stable.

Figure 48
COMPARISON DEATH RATES FORECASTS FOR INFLUENZA / CLUSTER VS NATIONAL AT AGES 30–34

The death rates for cause influenza at ages 30–34 are expected to remain stable. These death rates are lower for the cluster population than for the national population for both males and females, and these differences are expected to remain stable.
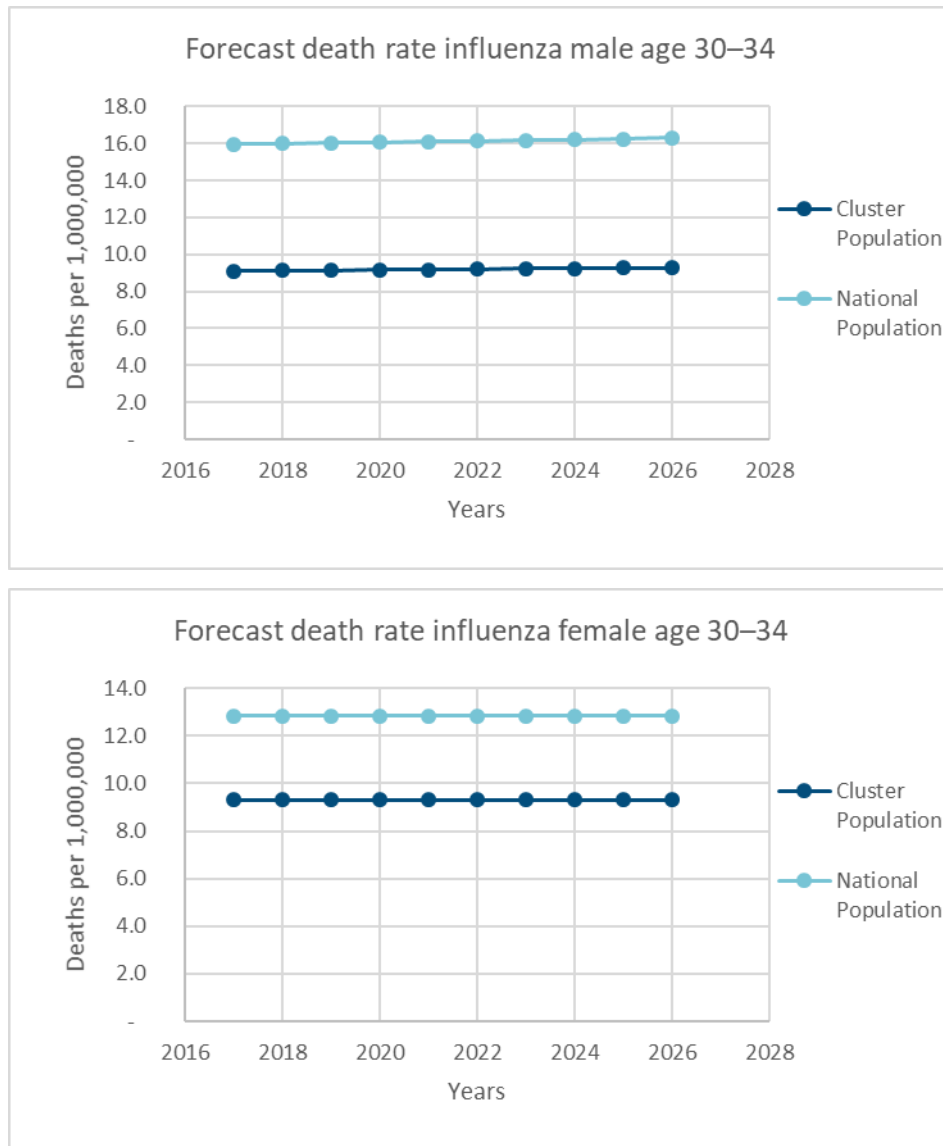
Figure 49

COMPARISON DEATH RATES FORECASTS FOR RESPIRATORY / CLUSTER VS NATIONAL AT AGES 30–34

The death rates for cause respiratory at ages 30–34 are expected to remain stable for males and to decrease for females. These death rates are lower for the cluster population than for the national population and the gaps are expected to remain stable.
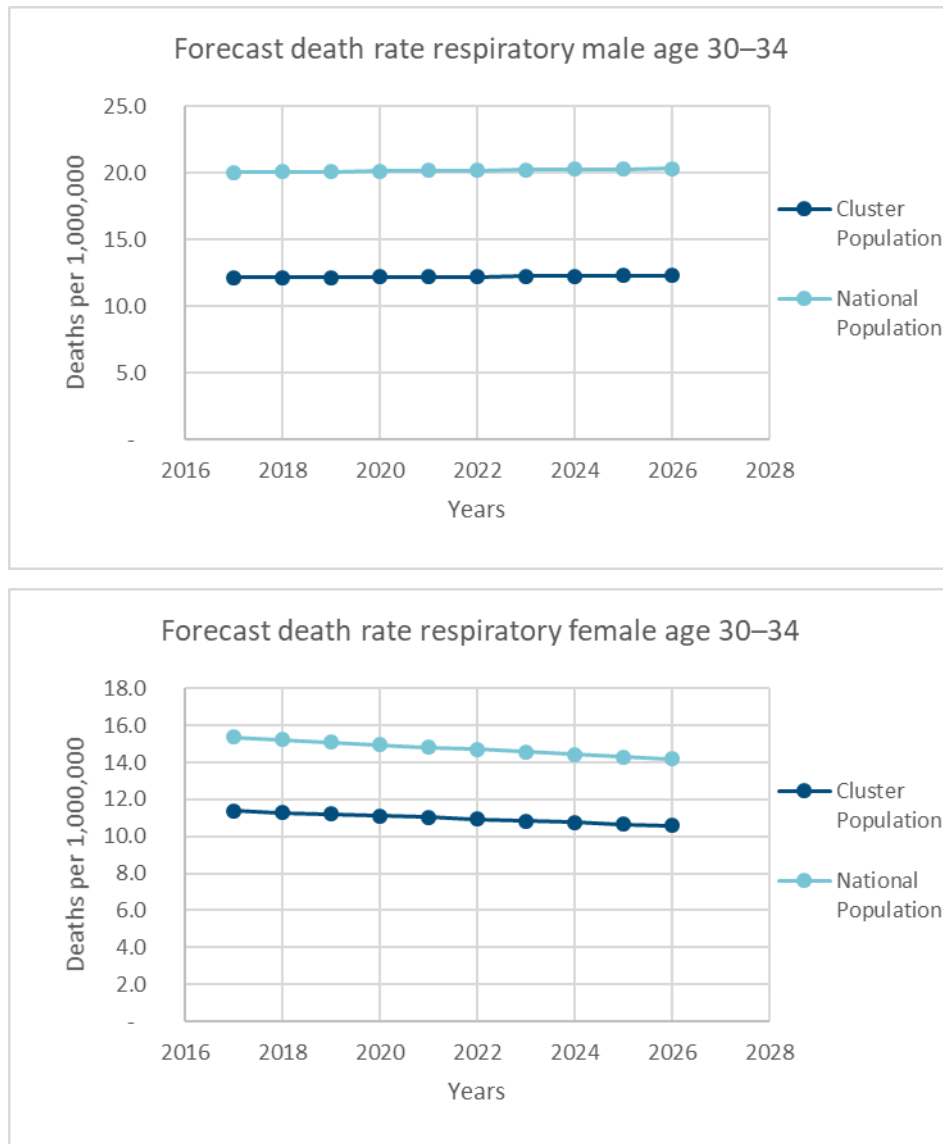
Figure 50

COMPARISON DEATH RATES FORECASTS FOR OTHER / CLUSTER VS NATIONAL AT AGES 30–34

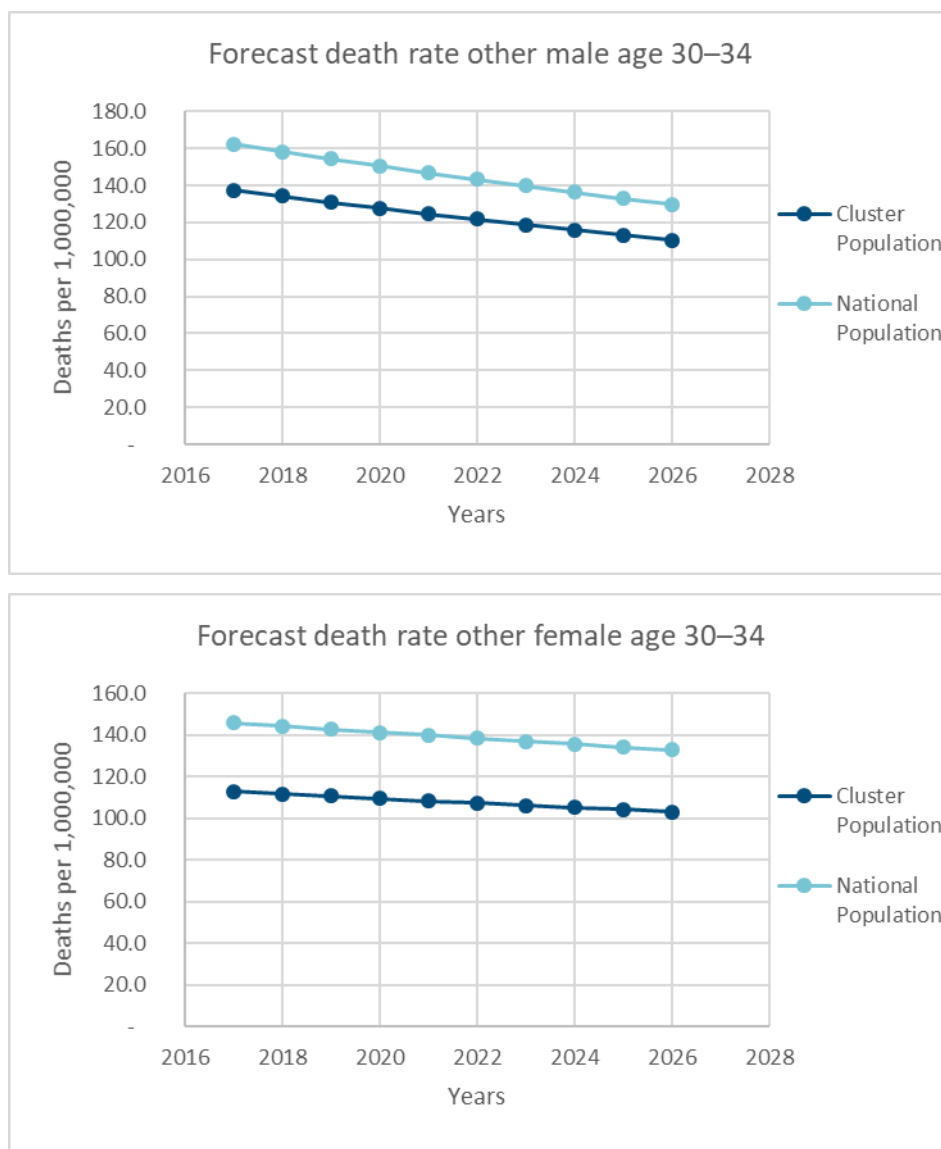The death rates for cause other at ages 30–34 are expected to decrease for both males and females. These death rates are lower for the cluster population than for the national population and the differences are expected to remain stable.

B.2 AT AGES 50–54

Figure 51
COMPARISON DEATH RATES FORECASTS FOR CEREBROVASCULAR / CLUSTER VS NATIONAL AT AGES 50–54

The death rates of the national population for cause cerebrovascular at ages 50–54 are expected to decrease for both males and females. However, for the cluster population the death rates are expected to decrease slower. Thus, the differences between the cluster and the national population are expected to narrow a little.

Figure 52

COMPARISON DEATH RATES FORECASTS FOR NEOSMOK / CLUSTER VS NATIONAL AT AGES 50–54
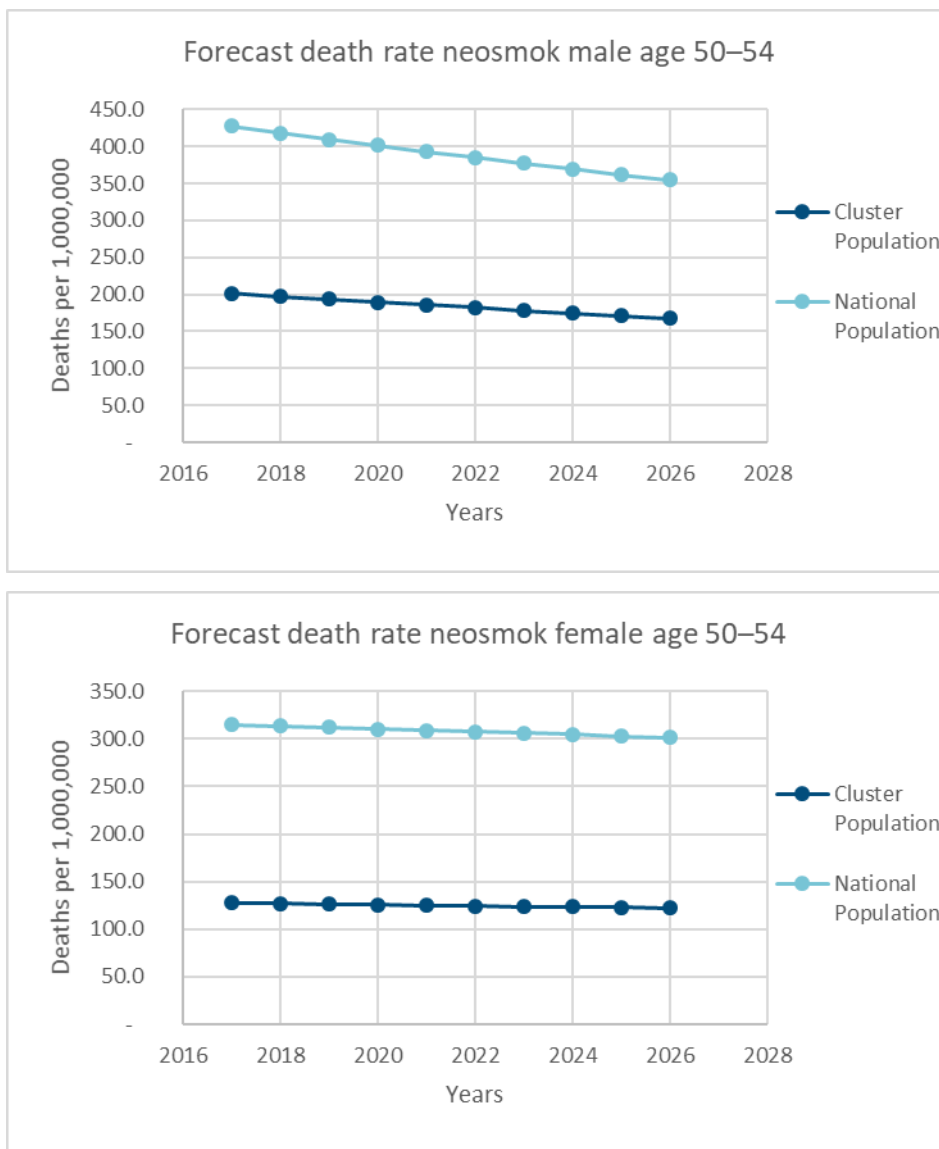
The death rates of the national population for cause neosmok at ages 50–54 are expected to decrease for both males and females and the decreases would be higher for males. For the cluster population the death rates are also expected to decrease, but slower. Thus, the differences between the cluster and the national population are expected to narrow a little.
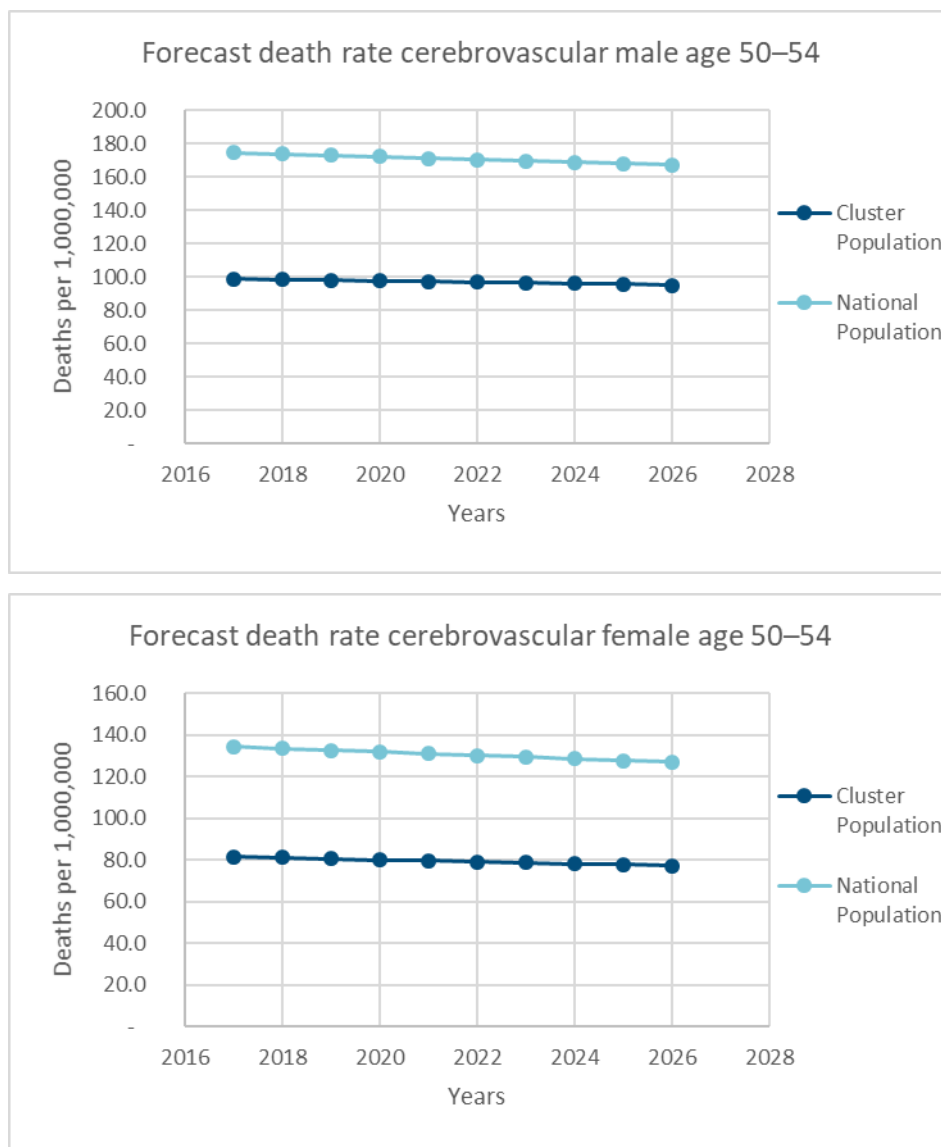
Figure 53

COMPARISON DEATH RATES FORECASTS FOR DEMENTIA / CLUSTER VS NATIONAL AT AGES 50–54
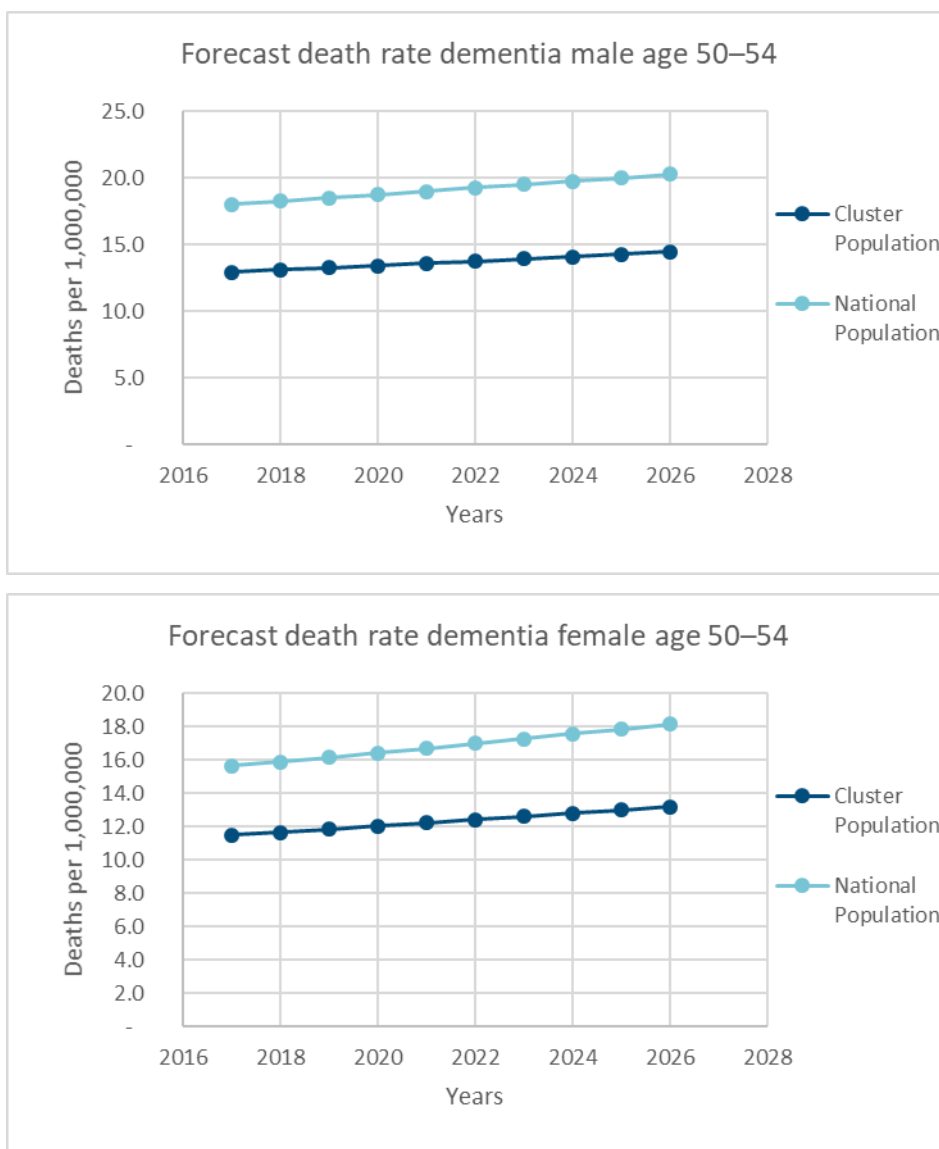
The death rates for cause dementia at ages 50–54 are expected to increase for both males and females. The differences between the cluster and the national population are expected to remain. The death rates at ages 50–54 remain still low.

Figure 54

COMPARISON DEATH RATES FORECASTS FOR DIABETES / CLUSTER VS NATIONAL AT AGES 50–54
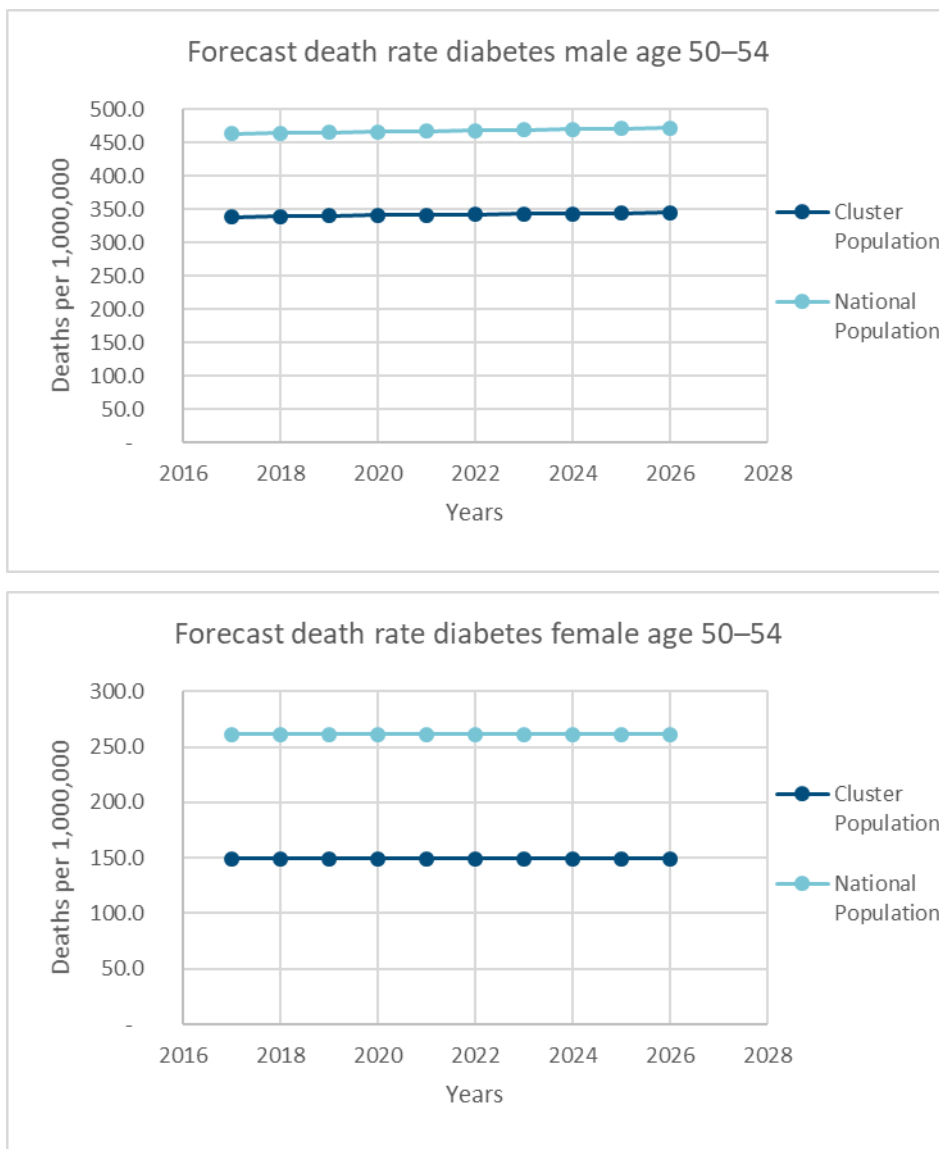
The death rates for cause diabetes at ages 50–54 are expected to remain quite stable for both males and females. The differences between the cluster and the national population are expected to be stable.

Figure 55

COMPARISON DEATH RATES FORECASTS FOR INFLUENZA / CLUSTER VS NATIONAL AT AGES 50–54

The death rates for cause influenza at ages 50–54 are expected to remain stable for both males and females. The differences between the cluster and the national population are also expected to be stable.

Figure 56
COMPARISON DEATH RATES FORECASTS FOR RESPIRATORY / CLUSTER VS NATIONAL AT AGES 50–54
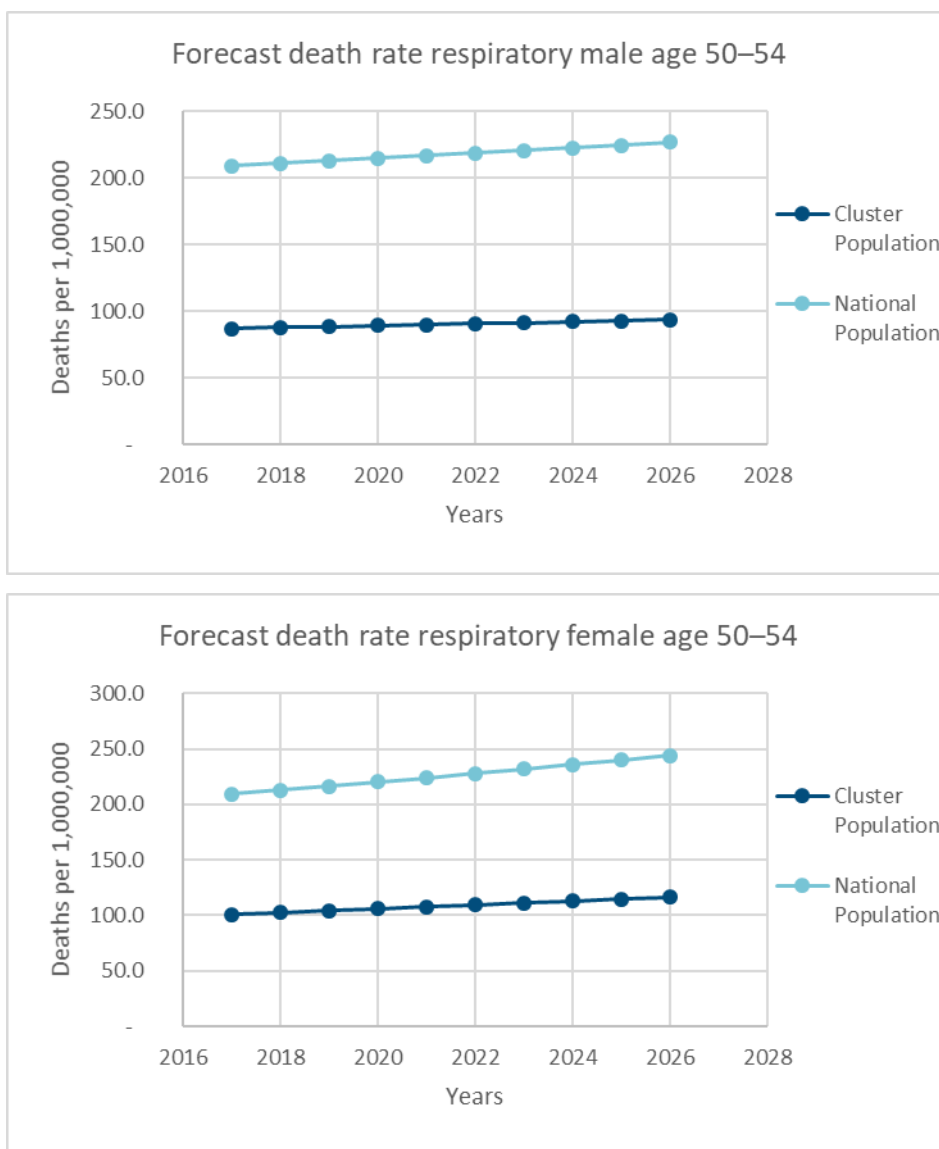
The death rates for cause respiratory at ages 50–54 are expected to increase for both males and females, but to increase slower for the cluster population than for the national population because of the difference of the initial level. Thus, the gap between the cluster and the national population would widen.

Figure 57
COMPARISON DEATH RATES FORECASTS FOR EXTERNAL / CLUSTER VS NATIONAL AT AGES 50–54

The death rates for cause external at ages 50–54 are expected to remain stable for both males and females, and for both the cluster and the national population.

Figure 58

COMPARISON DEATH RATES FORECASTS FOR OTHER / CLUSTER VS NATIONAL AT AGES 50–54

The death rates for cause other at ages 50–54 are expected to remain stable or to decrease a little for males, for both the cluster and the national population. For females, the death rates are expected to increase a little for both the cluster and the national population.

B.3 AT AGES 80–84

Figure 59

COMPARISON DEATH RATES FORECASTS FOR CEREBROVASCULAR / CLUSTER VS NATIONAL AT AGES 80–84

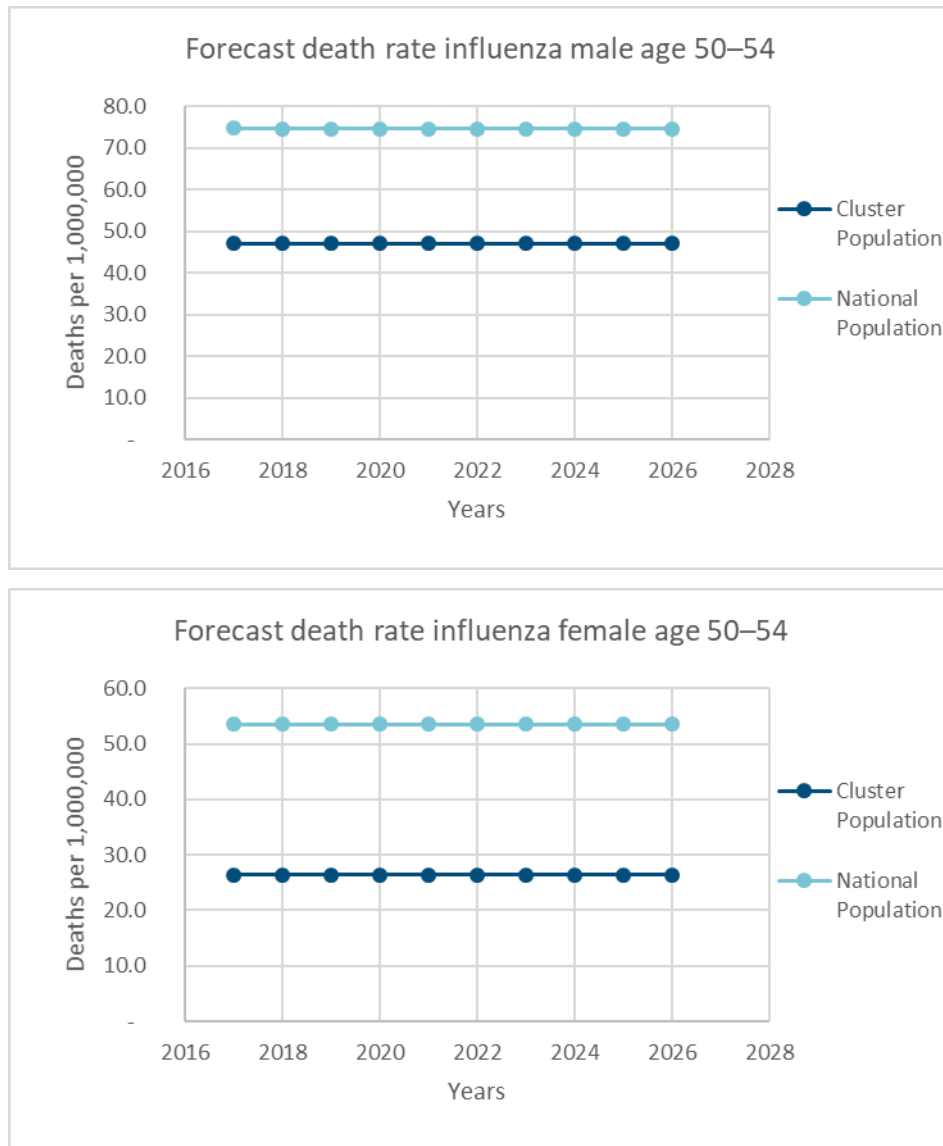The death rates for cause cerebrovascular at 80–84 are expected to decrease for both males and females. The gap between the cluster and the national population is expected to remain stable.

Figure 60
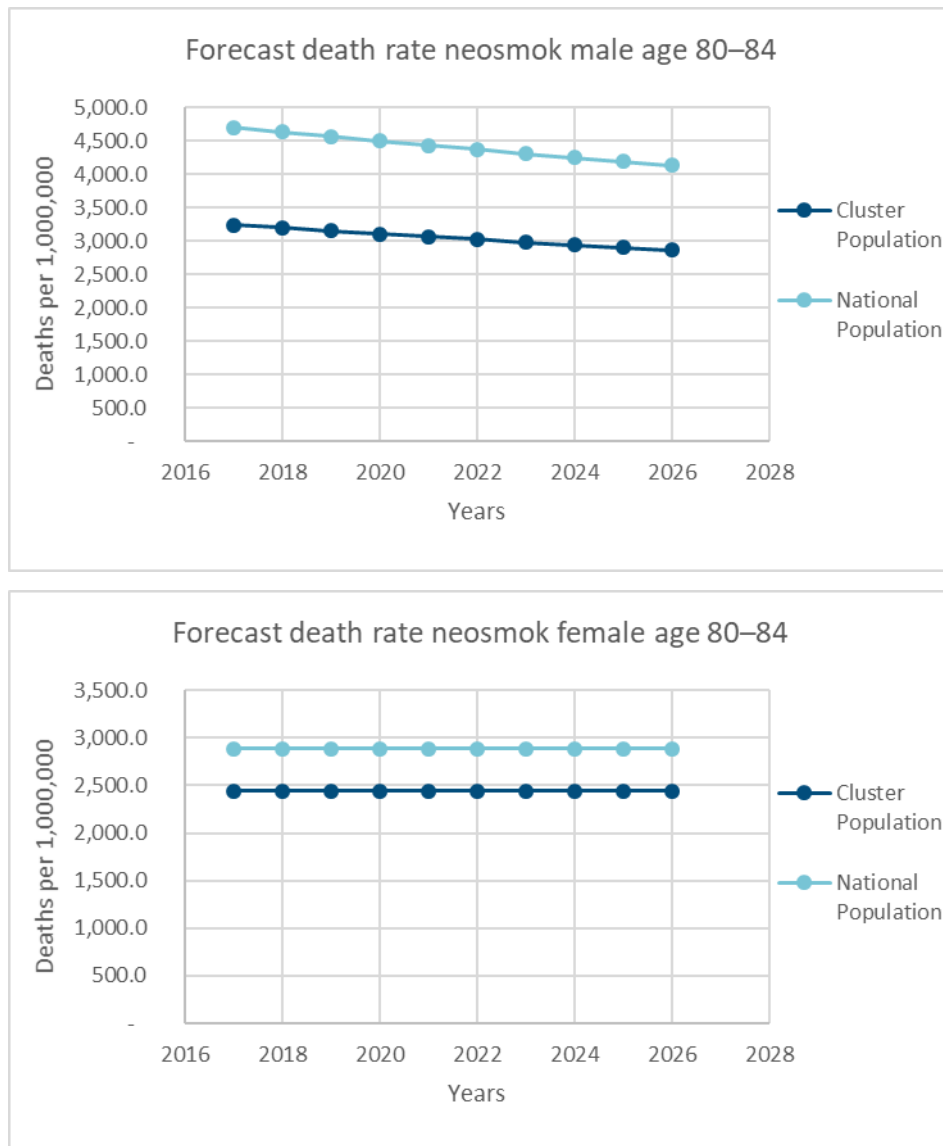
The death rates for cause neosmok at 80–84 are expected to decrease for males but slower for the cluster population than for the national population, which means that the differences would narrow. For females, the death rates and the gap between the cluster and the national population are expected to remain stable.

Figure 61
COMPARISON DEATH RATES FORECASTS FOR DIABETES / CLUSTER VS NATIONAL AT AGES 80–84

The death rates for cause diabetes at 80–84 are expected to decrease a little for both males and females and for both the cluster and the national population. The gap between the cluster and the national population is expected to remain stable.

Figure 62

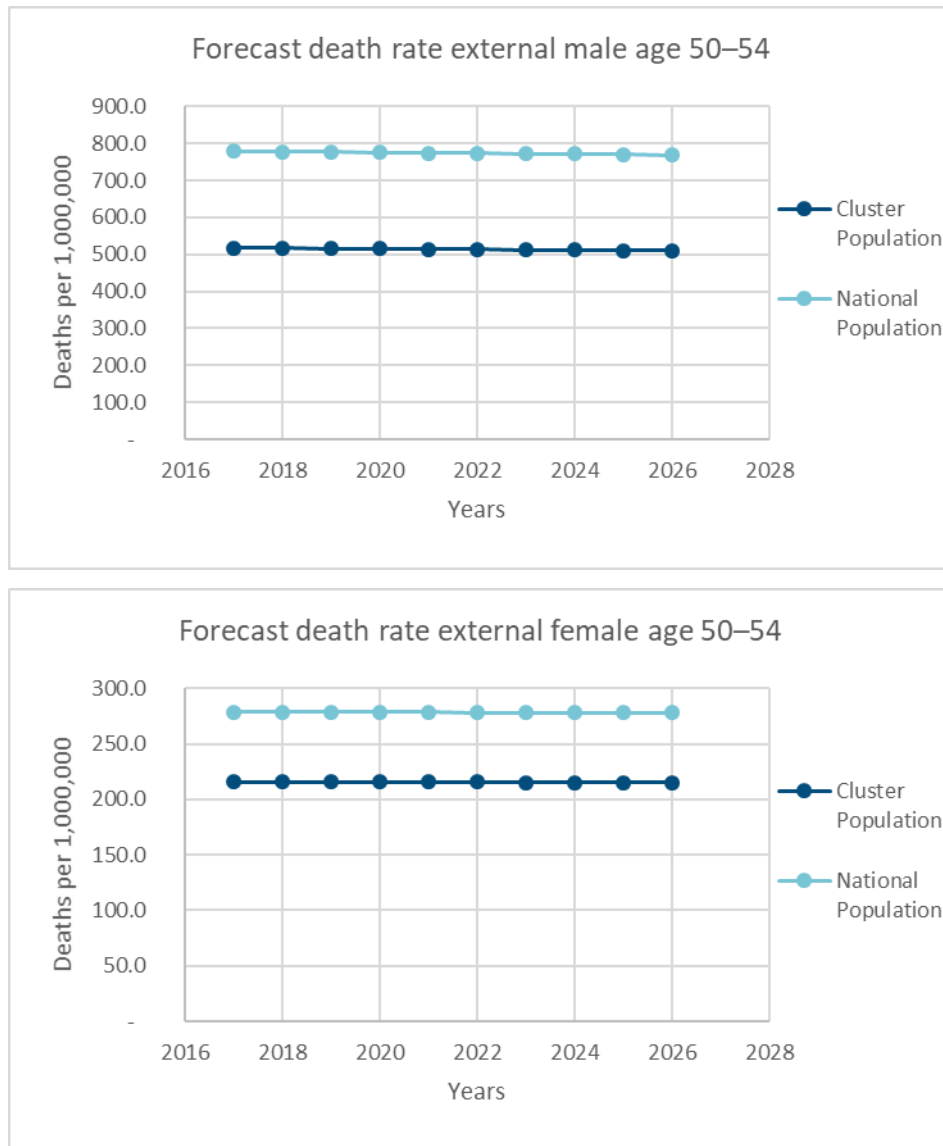COMPARISON DEATH RATES FORECASTS FOR INFLUENZA / CLUSTER VS NATIONAL AT AGES 80–84

The death rates for cause influenza at 80–84 are expected to decrease for males for both the cluster and the national population. For females, the death rates would remain stable. The gap between the cluster and the national population is expected to remain stable.

Figure 63

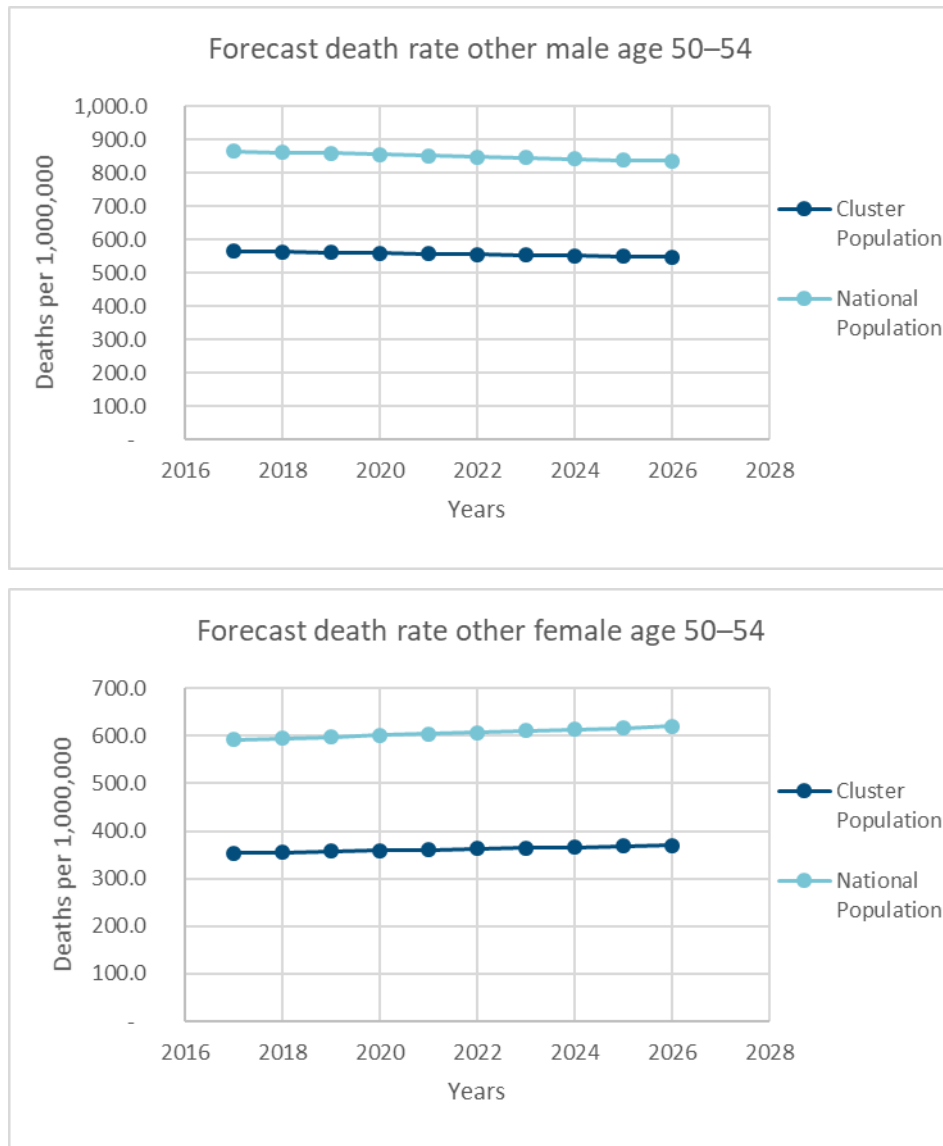COMPARISON DEATH RATES FORECASTS FOR RESPIRATORY / CLUSTER VS NATIONAL AT AGES 80–84

The death rates for cause respiratory at 80–84 are expected to decrease for males but a little slower for the cluster population than for the national population. For females, the death rates would remain stable. The gap between the cluster and the national population is expected to remain stable for females and to narrow a little for males.

Figure 64

COMPARISON DEATH RATES FORECASTS FOR DRUG ABUSE / CLUSTER VS NATIONAL AT AGES 80–84
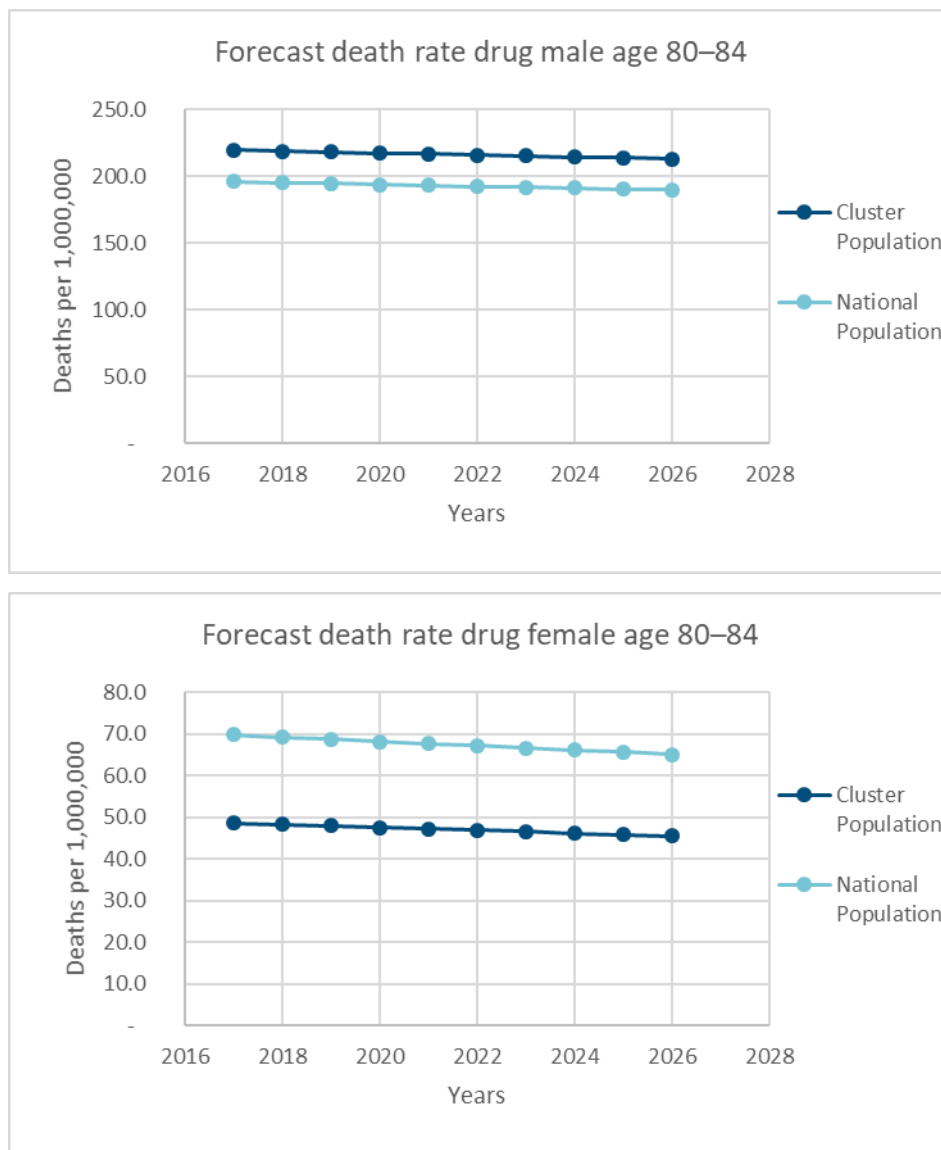
The death rates for cause drug at 80–84 are expected to decrease for both males and females. Note that for males at 80–84, the cause drug is a cause for which the death rates would be higher for the cluster population than the national population. The differences between the cluster and the national population would remain stable.

Figure 65

COMPARISON DEATH RATES FORECASTS FOR EXTERNAL / CLUSTER VS NATIONAL AT AGES 80–84

Forecast death rate external male age 80–84

Forecast death rate external female age 80–84
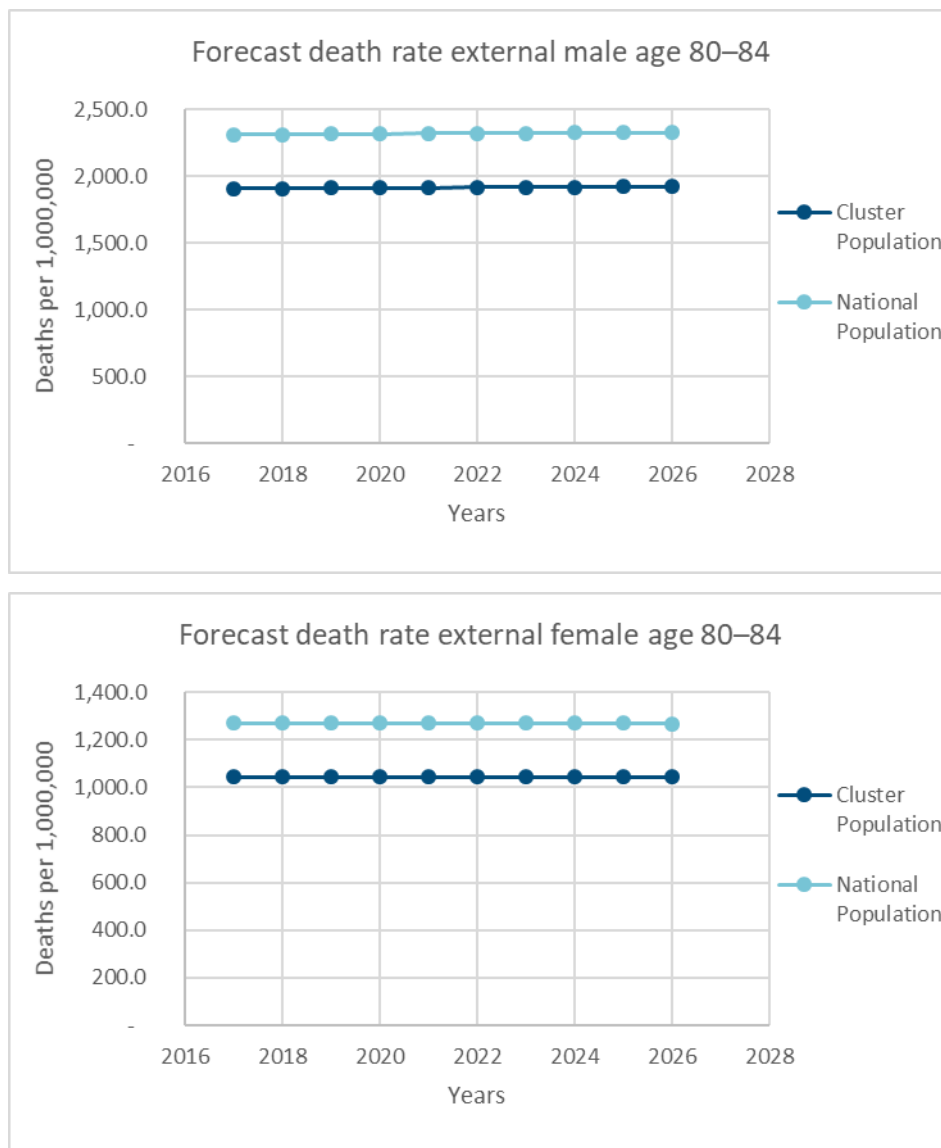
The death rates for cause external at 80–84 are expected to remain stable for both males and females, and for both the cluster and the national population.
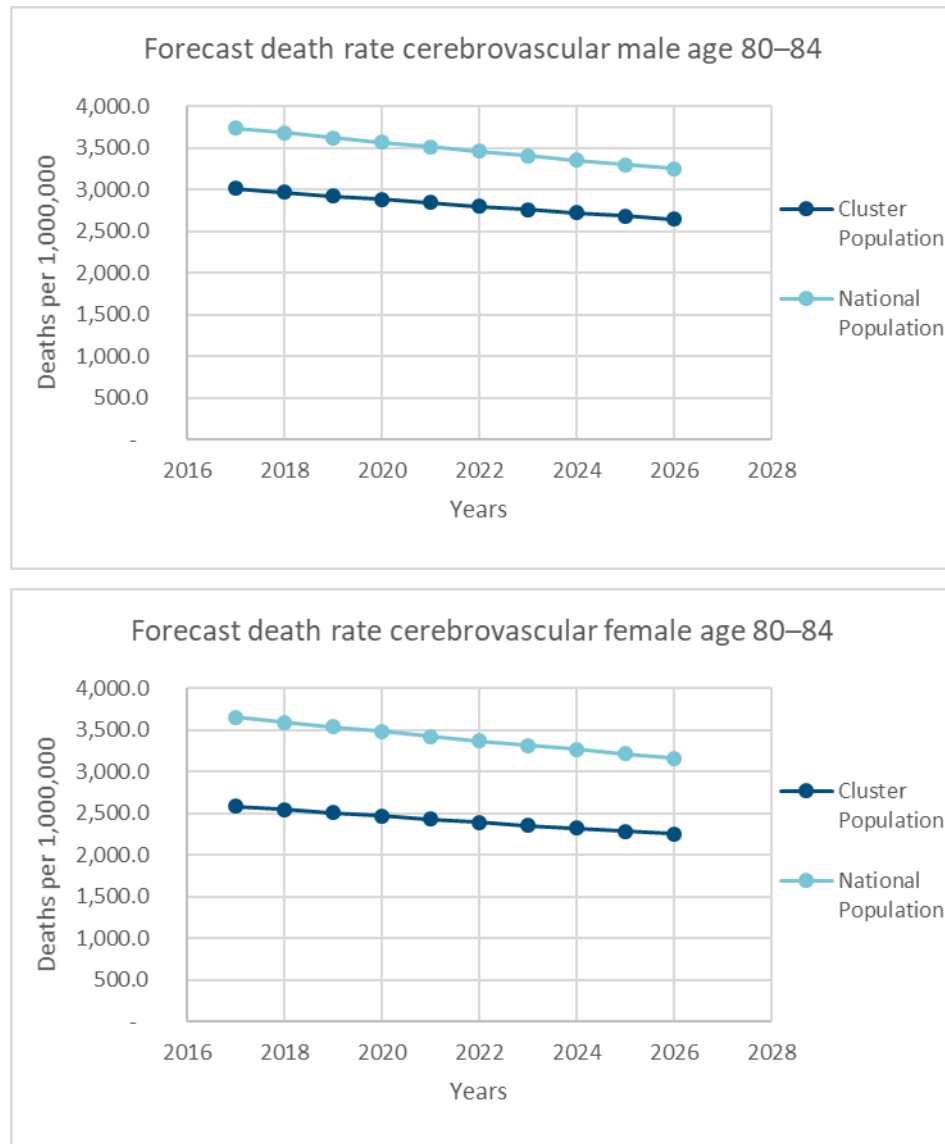
Figure 66
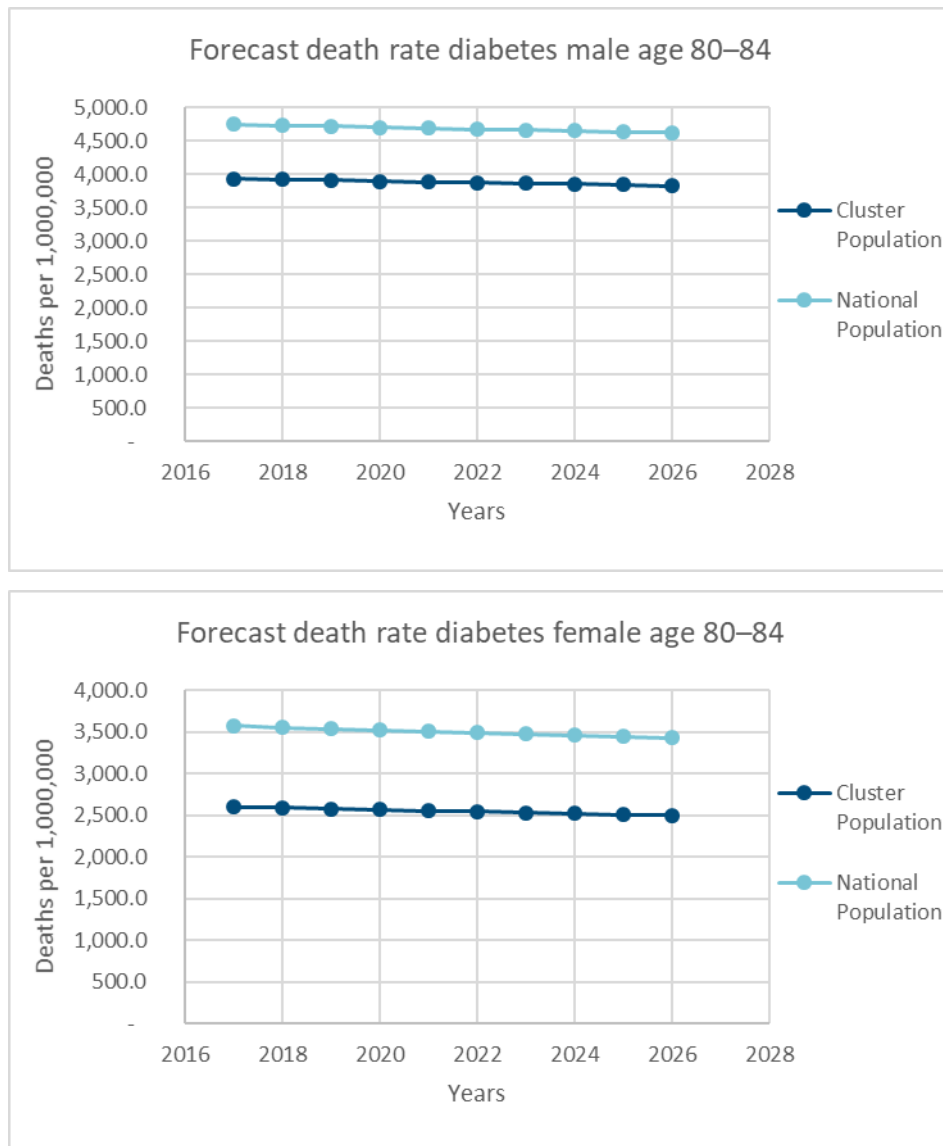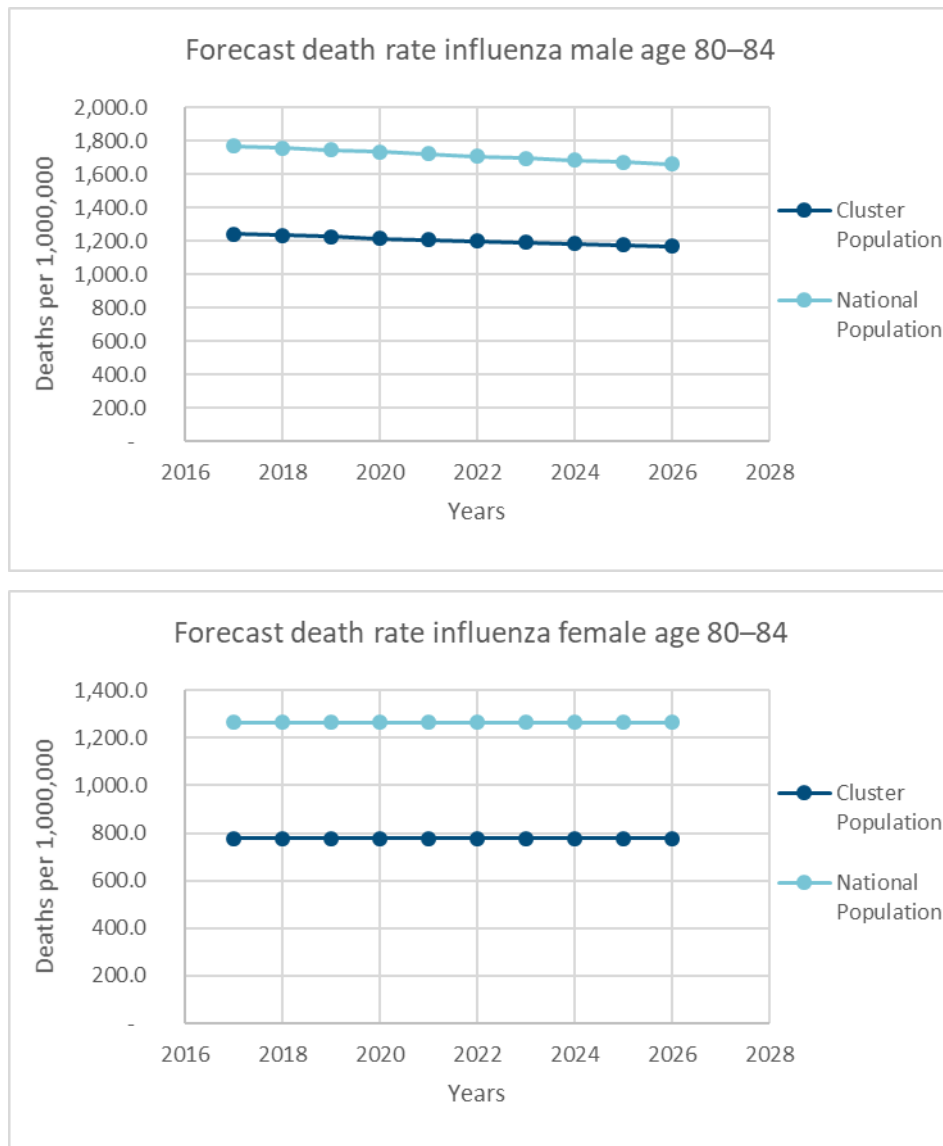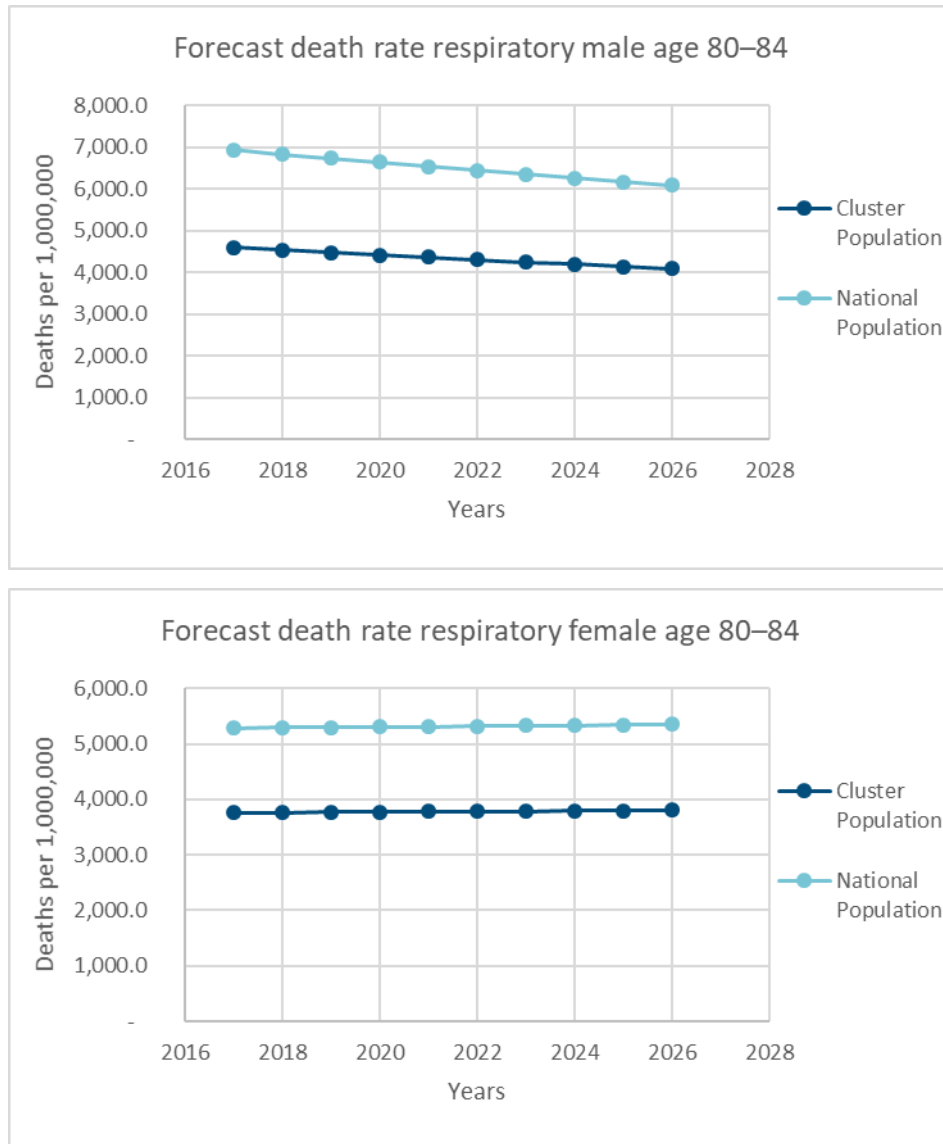COMPARISON DEATH RATES FORECASTS FOR OTHER / CLUSTER VS NATIONAL AT AGES 80–84
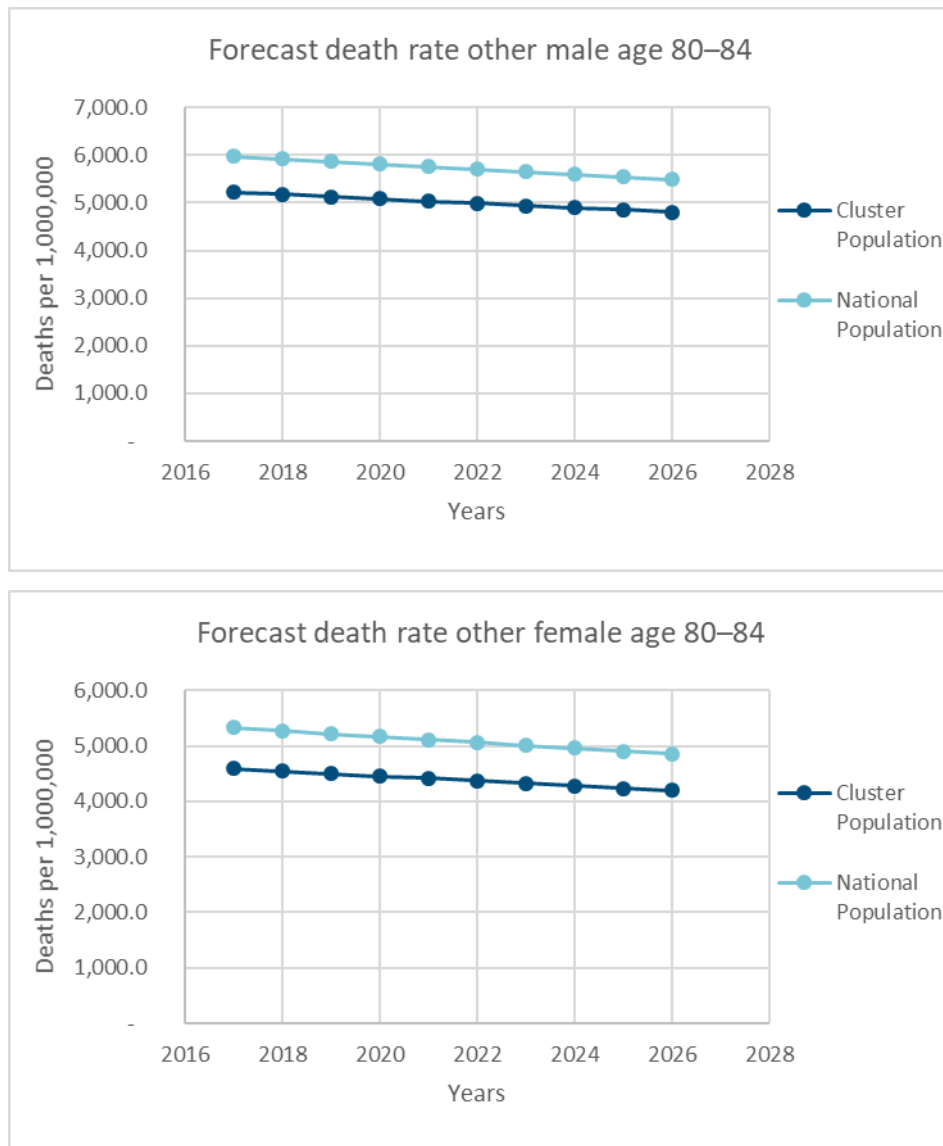
The death rates for cause other at 80–84 are expected to decrease for both males and females, and for both the cluster and the national population. The differences between the cluster and the national population are expected to remain stable.

Appendix C: Partial Life Expectancies Forecasts

The partial life expectancies forecasts at birth until 85 for males and females are detailed below. The partial life expectancy at birth is not clear as it is expected to stagnate or even to decrease. This is explained by the capture of the high historical increase of deaths for the cause drug and the low level of improvement of mortality rates at young ages for the other causes. This decline may not happen in practice if deaths because of drug do not continue to increase as fast as it used to be in the past (see Appendix E related to CDC death forecasts up to June 2019 for drug overdose). The partial life expectancy is 2.5 years higher for the male cluster population than it is for the male national population. It is also 2 years higher for the female cluster population than it is for the female national population. This confirms that the cluster (and richer) population has a higher life expectancy than the national population, and this difference is expected to remain for the future years.

Figure 67
PARTIAL LIFE EXPECTANCY FORECASTS AT BIRTH UNTIL 85
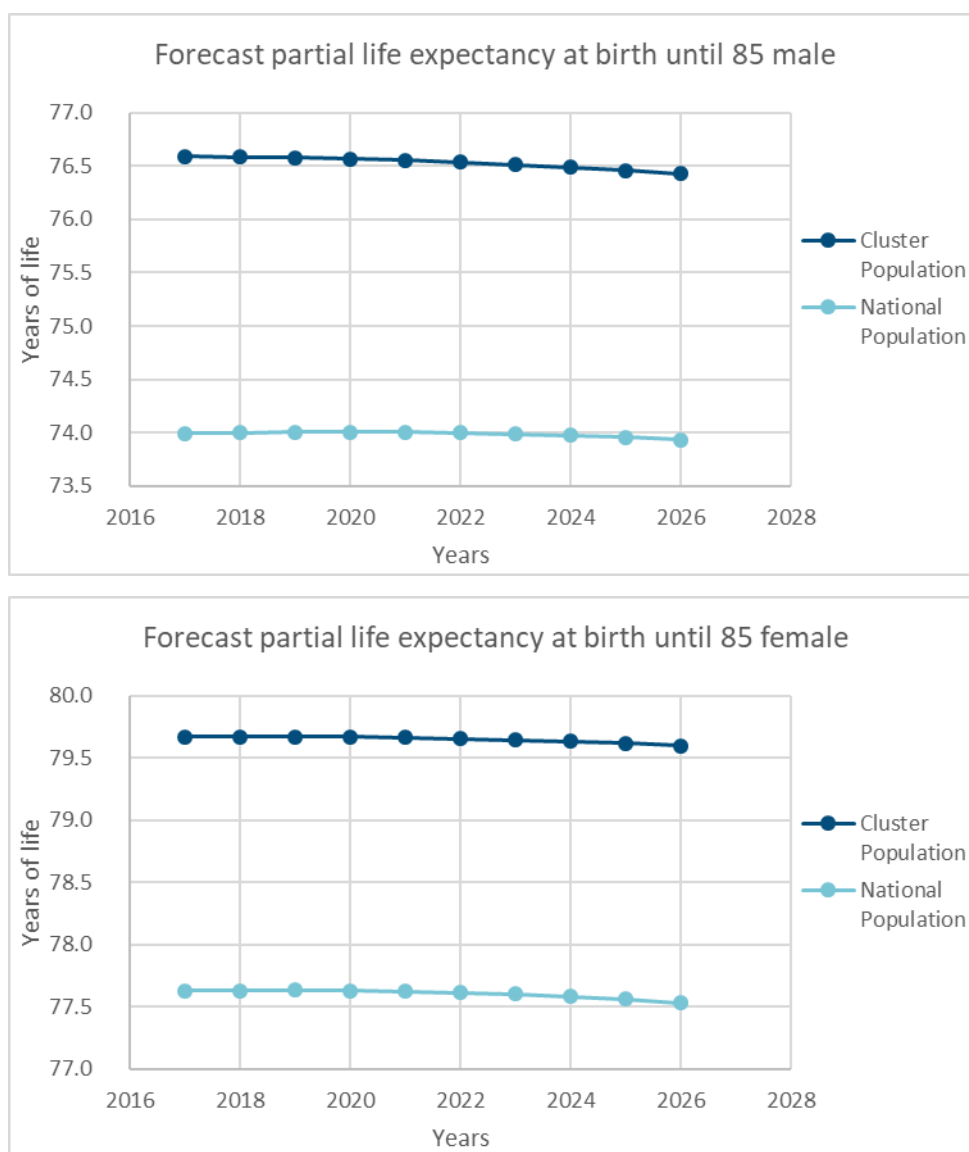
Appendix D: Description of Socioeconomic Variables

Source: BEA Regional Economic Accounts

110 – Per capita personal income: personal income of a given area divided by the resident population of the area. The personal income consists of the income that persons receive in return for their provision of labor, land, and capital used in current production as well as other income, such as personal current transfer receipts.

120 – Per capita net earnings: Net earnings by place of residence of a given area divided by the resident population of the area. Net earnings consists of earnings by place of work less contributions for government social insurance plus the adjustment for residence.

130 - Per capita personal current transfer receipts: receipts of persons from government and business for which no current services are performed. Current transfer receipts from government include Social Security benefits, medical benefits, veterans' benefits, and unemployment insurance benefits. Current transfer receipts from business include liability payments for personal injury and corporate gifts to nonprofit institutions.

140 – Per capita income maintenance benefits: income maintenance benefits of a given area divided by the resident population of the area. Income maintenance benefits consists largely of Supplemental Security Income (SSI) benefits, Earned Income Tax Credit (EITC), Additional Child Tax Credit, Supplemental Nutrition Assistance Program (SNAP) benefits, family assistance, and other income maintenance benefits, including general assistance.

150 – Per capita unemployment insurance compensation: unemployment insurance compensation of a given area divided by the resident population of the area. Unemployment insurance compensation is made up of the following: State unemployment compensation are benefits consisting mainly of the payments received by individuals under state-administered unemployment insurance (UI) programs, but they include the special benefits authorized by federal legislation for periods of high unemployment. The provisions that govern the eligibility, timing, and amount of benefit payments vary among the states, but the provisions that govern the coverage and financing are uniform nationally. Unemployment compensation of Federal civilian employees are benefits, which are received by former federal civilian employees under a federal program administered by the state employment security agencies acting as agents for the U.S. Government. Unemployment compensation of railroad employees are benefits which are received by railroad workers who are unemployed because of sickness or because work is unavailable in the railroad industry and in related industries, such as carrier affiliates. This UI program is administered by the Railroad Retirement Board (RRB) under a federal formula that is applicable throughout the Nation. Unemployment compensation of veterans are benefits, which are received by unemployed veterans who have recently separated from military service and who are not eligible for military retirement benefits. Trade adjustment assistance are benefits received by workers who are unemployed because of the adverse economic effects of international trade arrangements.

160 – Per capita retirement and other: retirement and other of a given area divided by the resident population of the area. Retirement and other consists of personal current transfer receipts excluding unemployment insurance compensation and income maintenance benefits.

170 – Per capita dividends, interest, and rent: dividends, interest, and rent of a given area divided by the resident population of the area. It consists of personal dividend income, personal interest income, and rental income of persons with capital consumption adjustment.

250 – Wage and salary employment (% Total employment – number of jobs): Wage and salary employment, also referred to as wage and salary jobs, measures the average annual number of full-time and part-time jobs in each area by place of work. All jobs for which wages and salaries are paid are counted. Although compensation paid to jurors, expert legal witnesses, prisoners, and justices of the peace (for marriage fees), is counted in wages and salaries, these activities are not counted as jobs in wage and salary employment. Corporate directorships are counted as self-employment. This number of jobs is divided by the total number of jobs (both full-time and part-time, including wage and salary jobs, sole proprietorships, and individual general partners, but not unpaid family workers nor volunteers).

270 – Farm proprietors employment (% Proprietors employment): consists of the part of sole proprietors and non-corporate partners in the farm industry among the total number of both farm proprietors employment and nonfarm proprietors employment.

2100 – Retirement and disability insurance benefits – Per capita: Retirement and Disability Insurance Benefits consist of old-age, survivors, and disability (OASDI) benefits; railroad retirement and disability benefits; Federal and state workers' compensation; temporary disability benefits; black lung benefits; and Pension Benefit Guaranty benefits.

2200 – Medical benefits – Per capita: Medical benefits include:
- Medicare benefits - These benefits are Federal Government payments made through intermediaries to beneficiaries for the care provided to individuals under the Medicare program.
- Public assistance medical care - These medical benefits are received by low-income individuals. These payments consist mainly of the payments made through intermediaries to the vendors for care provided to individuals under the federally assisted, state-administered Medicaid program and Children's Health Insurance Program (CHIP) and under the general assistance medical programs of state and local governments.
- Military medical insurance benefits - These benefits are vendor payments made under the TriCare Management Program, formerly called the Civilian Health and Medical Plan of the Uniformed Services program, for the medical care of dependents of active duty military personnel and of retired military personnel and their dependents at nonmilitary medical facilities.

2600 – Education and training assistance – per capita: Education and training assistance consists of:
- Federal fellowships - These benefits consist of the payments to outstanding science students who receive National Science Foundation (NSF) grants, the subsistence payments to the cadets at the six state maritime academies, and the payments for all other Federal fellowships.
- Higher education student assistance- These benefits consist of the Federal payments, called Pell Grants, for an undergraduate education for students with low incomes.
- Job Corps payments - These benefits are primarily the allowances for living expenses received by economically disadvantaged individuals who are between the ages of 16 and 21 and who are enrolled in the designated vocational and educational training programs. These benefits also include the adjustment allowances received by trainees upon the successful completion of their training.
- Interest payments on guaranteed student loans - These payments are made by the Department of Education to commercial lending institutions on behalf of the individuals who receive low-interest, deferred-payment loans from these institutions in order to pay the expenses of higher education.
- State educational assistance - These benefits consist of educational assistance provided by states to individuals for tuition and other educational expenses not including loans. The national and state estimates are based on data for state government expenditures for "other education assistance and subsidies" from the Census Bureau's annual State Government Finances.

Appendix E: CDC Death Forecasts Up To June 2019 For Drug Overdose

The two following figures come from Centers for Disease Control and Prevention (CDC) website and are based on data available for analysis on 01/05/2020. The data are incomplete after 2017. Estimates for 2018 and 2019 are based on provisional data whereas estimates for 2015–2017 are based on final data.

Figure 68

12 MONTH-ENDING PROVISIONAL COUNTS OF DRUG OVERDOSE DEATHS: UNITED STATES
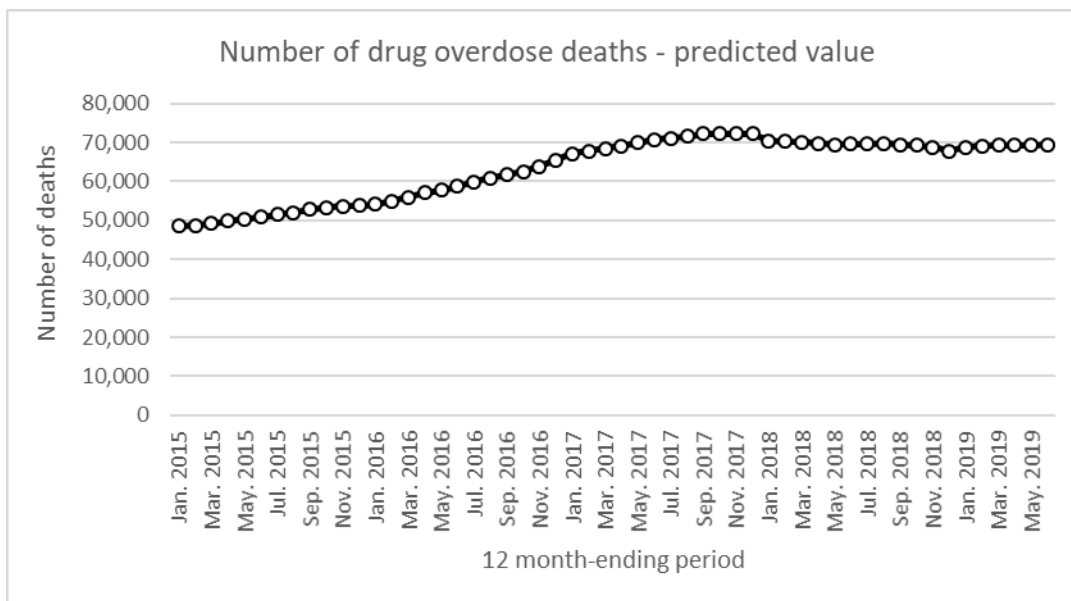
According to the CDC, the number of drug overdose deaths has increased between 2015 and 2017. After December 2017, a slight decrease and a stagnation of the number of deaths because of drug overdose is predicted at around 70,000 deaths for the United States. However, this change would differ from a state to another, for example with a decrease in deaths in Maine but an increase in deaths in California can be observed, as shown in the figure below.

Figure 69

PERCENT CHANGE IN PREDICTED 12 MONTH-ENDING COUNT OF DRUG OVERDOSE DEATHS, BY JURISDICTION, JUNE 2018 TO JUNE 2019
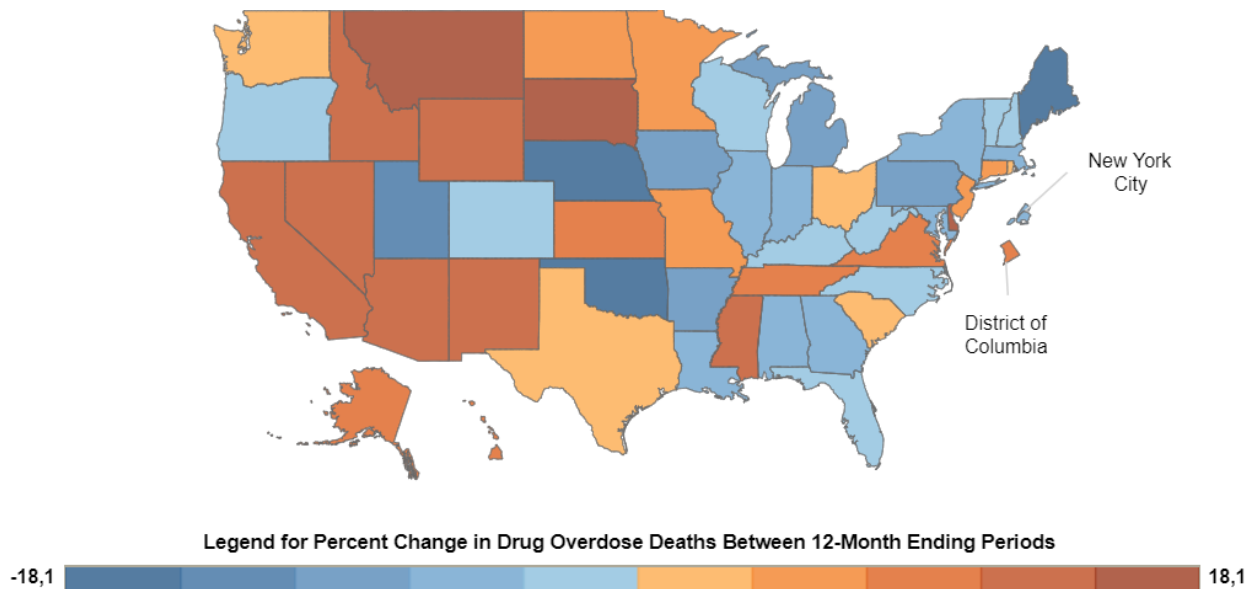
Notes from the CDC on this study are given below:

Reported provisional counts for 12-month ending periods are the number of deaths received and processed for the 12-month period ending in the month indicated. Drug overdose deaths are often initially reported with no cause-of-death (pending investigation) because they require lengthy investigation, including toxicology testing. Reported provisional counts may not include all deaths that occurred during a given time period. Therefore, they should not be considered comparable with final data and are subject to change. Predicted provisional counts represent estimates of the number of deaths adjusted for incomplete reporting. Deaths are classified by the reporting jurisdiction in which the death occurred. Percent change refers to the relative difference between the reported or predicted provisional numbers of deaths due to drug overdose occurring in the 12-month period ending in the month indicated compared with the 12-month period ending in the same month of the previous year. Drug overdose deaths are identified using ICD–10 underlying cause-of-death codes: X40–X44, X60–X64, X85, and Y10–Y14.

Appendix F: Cause-Of-Deaths Classification

Table 11 shows the mapping between the specific causes-of-death used in this study and HCD/GBD lists. The next Table 12 provides the ICD 10 codes corresponding to the HCD intermediate list classification.

Table 11
CAUSES MAPPING

Cause of Death	GBD Classification	HCD Classification
Cardiovascular diseases (A)	B.2.1 Rheumatic heart disease B.2.2 Ischemic heart disease B.2.5 Non-rheumatic valvular heart disease B.2.6 Cardiomyopathy and myocarditis B.2.7 Atrial fibrillation and flutter B.2.8 Aortic aneurysm B.2.9 Peripheral artery disease B.2.10 Endocarditis B.2.11 Other cardiovascular and circulatory diseases	Rheumatic heart diseases Essential hypertension Acute myocardial infarction Atherosclerotic cardiovascular and heart diseases Other IHD Pulmonary heart diseases Non rheumatic valve disorders Cardiac arrest Heart failure Other heart diseases Diseases of arteries, arterioles and capillaries Other circulatory diseases
Cerebrovascular diseases (B)	B.2.3. Stroke	Intracranial hemorrhage Cerebral infarction, occlusion, and stenosis Other Cerebrovascular Sequelae of cerebrovascular disease
Neoplasms directly induced by smoking (NeoSmok, C)	B.1.1 Lip and oral cavity cancer B.1.2 Nasopharynx cancer B.1.3 Other pharynx cancer B.1.10 Larynx cancer B.1.11 Tracheal, bronchus, and lung cancer	Malignant Neoplasm of lip, oral cavity and pharynx Malignant neoplasm of larynx Malignant Neoplasm of trachea, bronchus and lung
Neoplasms not directly induced by smoking (D)	B.1.4 Esophageal cancer B.1.5 Stomach cancer B.1.6 Colon and rectum cancer B.1.7 Liver cancer B.1.8 Gallbladder and biliary tract cancer B.1.9 Pancreatic cancer B.1.12 Malignant skin melanoma B.1.13 Non-melanoma skin cancer B.1.14 Breast cancer B.1.15 Cervical cancer B.1.16 Uterine cancer B.1.17 Ovarian cancer B.1.18 Prostate cancer B.1.19 Testicular cancer B.1.20 Kidney cancer B.1.21 Bladder cancer B.1.22 Brain and nervous system cancer B.1.23 Thyroid cancer B.1.24 Mesothelioma B.1.25 Hodgkin lymphoma B.1.26 Non-Hodgkin lymphoma B.1.27 Multiple myeloma B.1.28 Leukemia B.1.29 Other malignant Neoplasm B.1.30 Other Neoplasm	Malignant neoplasm of esophagus Malignant neoplasm of stomach Malignant Neoplasm of colon Malignant neoplasm of rectum and anus Malignant Neoplasm of liver and intrahepatic bile ducts Malignant neoplasm of pancreas Other malignant neoplasm of digestive system Malignant neoplasm of skin Malignant neoplasm of breast Malignant neoplasm of cervix uteri Malignant Neoplasm of uterus Malignant neoplasm of ovary Malignant neoplasm of prostate Malignant neoplasm of other genital organs Malignant neoplasm of bladder Malignant Neoplasm of kidney and other urinary organ Malignant Neoplasm of meninges, brain and other parts of central nervous system Leukemia

		Other malignant Neoplasm of lymphoid, hematopoietic and related tissue Malignant Neoplasm of independent (primary) multiple sites Other cancer In situ Neoplasm, benign Neoplasm and Neoplasm of uncertain or unknown behavior
Dementia (E)	B.5.1 Alzheimer's disease and other dementias B.5.2 Parkinson's disease	Dementia, vascular, senile or unspecified Parkinson's disease and other extrapyramidal and movement disorders Alzheimer's disease and other degenerative diseases of the nervous system
Diabetes (F)	B.2.4 Hypertensive heart disease B.8.1 Diabetes mellitus B.8.2 Chronic kidney disease	Diabetes mellitus Hypertensive disease (heart, kidney and secondary) Renal failure
Influenza (G)	A.2.2 Lower respiratory infections A.2.3 Upper respiratory infections	Influenza Pneumonia Other acute respiratory infections
Respiratory diseases (H)	A.2.1 Tuberculosis B.3.1 Chronic obstructive pulmonary disease B.3.2 Pneumoconiosis B.3.3 Asthma B.3.4 Interstitial lung disease and pulmonary sarcoidosis B.3.5 Other chronic respiratory diseases	Asthma Other chronic obstructive pulmonary disease Pneumonitis due to solids and liquids Pneumoconiosis and chemical effects Other Respiratory, principally affecting the interstitium Other diseases of the respiratory system
Drug abuse (I)	B.2.4 Hypertensive heart disease B.8.1 Diabetes mellitus B.8.2 Chronic kidney disease	Alcohol abuse Drug abuse Alcoholic cirrhosis of liver Accidental poisoning by alcohol Accidental poisoning by other substance
External causes (J)	C.1.1 Road injuries C.1.2 Other transport injuries C.2.1 Falls C.2.2 Drowning C.2.3 Fire, heat, and hot substances C.2.4 Poisonings C.2.5 Exposure to mechanical forces C.2.6 Adverse effects of medical treatment C.2.7 Animal contact C.2.8 Foreign body C.2.9 Environmental heat and cold exposure C.2.10 Exposure to forces of nature C.2.11 Other unintentional injuries C.3.1 Self-harm C.3.2 Interpersonal violence C.3.3 Conflict and terrorism C.3.4 Executions and police conflict	Transport accidents Accidental falls Accidental drowning and submersion Accidental exposure to smoke, fire and flames Other accidental threats to breathing Suicide and self-inflicted injury Assault Event of undetermined intent Complications of medical and surgical care Other accidents and late effects of accidents (remainder)
Other (K)	A.1.1 HIV/AIDS A.1.2 Sexually transmitted infections excluding HIV A.2.4 Otitis media A.3.1 Diarrheal diseases A.3.2 Typhoid and paratyphoid	Other specified intestinal infections Diarrhea and gastroenteritis of presumed infectious origin TBC Septicemia Other bacterial diseases

A.3.3 Invasive Non-typhoidal Salmonella (iNTS) A.3.5 Other intestinal infectious diseases A.4.1 Malaria A.4.2 Chagas disease A.4.3 Leishmaniasis A.4.4 African trypanosomiasis A.4.5 Schistosomiasis A.4.6 Cysticercosis A.4.7 Cystic echinococcosis A.4.8 Lymphatic filariasis A.4.9 Onchocerciasis A.4.10 Trachoma A.4.11 Dengue A.4.12 Yellow fever A.4.13 Rabies A.4.14 Intestinal nematode infections A.4.15 Food-borne trematodiases A.4.16 Leprosy A.4.17 Ebola A.4.18 Zika virus A.4.19 Guinea worm disease A.4.20 Other neglected tropical diseases A.5.1 Meningitis A.5.2 Encephalitis A.5.3 Diphtheria A.5.4 Whooping cough A.5.5 Tetanus A.5.6 Measles A.5.7 Varicella and herpes zoster A.5.8 Acute hepatitis A.5.9 Other unspecified infectious diseases A.6.1 Maternal disorders A.6.2 Neonatal disorders A.7.1 Protein-energy malnutrition A.7.2 Iodine deficiency A.7.3 Vitamin A deficiency A.7.4 Dietary iron deficiency A.7.5 Other nutritional deficiencies B.4.2 Upper digestive system diseases B.4.3 Appendicitis B.4.4 Paralytic ileus and intestinal obstruction B.4.5 Inguinal, femoral, and abdominal hernia B.4.6 Inflammatory bowel disease B.4.7 Vascular intestinal disorders B.4.8 Gallbladder and biliary diseases B.4.9 Pancreatitis B.4.10 Other digestive diseases B.5.3 Epilepsy B.5.4 Multiple sclerosis B.5.5 Motor neuron disease B.5.6 Headache disorders B.5.7 Other neurological disorders B.6.1 Schizophrenia B.6.2 Depressive disorders B.6.3 Bipolar disorder B.6.4 Anxiety disorders B.6.5 Eating disorders	HIV disease Viral hepatitis Other viral diseases Other and unspecified infectious and parasitic disease Malnutrition Other endocrinologic and metabolic diseases Blood diseases Other mental disorders Systemic atrophies and demyelinating diseases of the central nervous system Epilepsy Other diseases of nervous system Gastric and duodenal ulcer Hernia Enteritis, colitis and other intestinal diseases Other cirrhoses of liver Other diseases of liver Cholelithiasis and other disorders of biliary tracts Diseases of pancreas Other digestive diseases Diseases of skin and subcutaneous tissue Diseases of the musculoskeletal system and connective tissue Renal tubulo-interstitial diseases Other diseases of urinary system Diseases of genital organs Complications of pregnancy, childbirth, and puerperium Certain conditions originating in the perinatal period Congenital malformations, deformations, and chromosomal abnormalities Sudden infant death syndrome (SIDS)

B.6.6 Autism spectrum disorders B.6.7 Attention-deficit/hyperactivity disorder B.6.8 Conduct disorder B.6.9 Idiopathic developmental intellectual disability B.6.10 Other mental disorders B.8.3 Acute glomerulonephritis B.9.1 Dermatitis B.9.2 Psoriasis B.9.3 Bacterial skin diseases B.9.4 Scabies B.9.5 Fungal skin diseases B.9.6 Viral skin diseases B.9.7 Acne vulgaris B.9.8 Alopecia areata B.9.9 Pruritus B.9.10 Urticaria B.9.11 Decubitus ulcer B.9.12 Other skin and subcutaneous diseases B.10.1 Blindness and vision impairment B.10.2 Age-related and other hearing loss B.10.3 Other sense organ diseases B.11.1 Rheumatoid arthritis B.11.2 Osteoarthritis B.11.3 Low back pain B.11.4 Neck pain B.11.5 Gout B.11.6 Other musculoskeletal disorders B.12.1 Congenital birth defects B.12.2 Urinary diseases and male infertility B.12.3 Gynecological diseases B.12.4 Hemoglobinopathies and hemolytic anemias B.12.5 Endocrine, metabolic, blood, and immune disorders B.12.6 Oral disorders B.12.7 Sudden infant death syndrome	

Table 12

HCD INTERMEDIATE LIST

No.	Title	Category codes according to ICD10
0	All causes	A00–Y98
1	Other specified intestinal infections	A00–A08
2	Diarrhea and gastroenteritis of presumed infectious origin	A09
3	TBC	A15–A19, B90
4	Septicemia	A40–A41
5	Other bacterial diseases	A20–A28, A30–A39 A42–A44, A46, A48–A49
6	HIV disease	B20–B24
7	Viral hepatitis	B15–B19
8	Other viral diseases	A80–A89, B00–B09, B25–B34
9	Other and unspecified infectious and parasitic diseases	A50–A75, A77–A79, A90–A99, B35–B60, B64–B89, B91, B92, B94–B97, B99
10	Malignant neoplasms of lip, oral cavity and pharynx	C00–C14

11	Malignant neoplasm of esophagus	C15
12	Malignant neoplasm of stomach	C16
13	Malignant neoplasms of colon	C18
14	Malignant neoplasm of rectum and anus	C19–C21
15	Malignant neoplasms of liver and intrahepatic bile ducts	C22
16	Malignant neoplasm of pancreas	C25
17	Other malignant neoplasm of digestive system	C17, C23–C24, C26
18	Malignant neoplasm of larynx	C32
19	Malignant neoplasms of trachea, bronchus and lung	C33–C34
20	Malignant neoplasm of skin	C43, C44
21	Malignant neoplasm of breast	C50
22	Malignant neoplasm of cervix uteri	C53
23	Malignant neoplasms of uterus	C54–C55
24	Malignant neoplasm of ovary	C56
25	Malignant neoplasm of prostate	C61
26	Malignant neoplasm of other genital organs	C51, C52, C57, C58, C60, C62, C63
27	Malignant neoplasm of bladder	C67
28	Malignant neoplasms of kidney and other urinary organ	C64–C66, C68
29	Malignant neoplasms of meninges, brain and other parts of central nervous system	C70–C72
30	Leukemia	C91–C95
31	Other malignant neoplasms of lymphoid, hematopoietic and related tissue	Other malignant neoplasms of lymphoid, hematopoietic and related tissue
32	Malignant neoplasms of independent (primary) multiple sites	C97
33	Other cancer	C30–C31,C37–C41, C45–C49, C69, C73–C80
34	In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior	D00–D48
35	Diabetes mellitus	E10–E14
36	Malnutrition	E40–E46
37	Other endocrinologic and metabolic diseases	E00–E07, E15–E16, E20–E35, E50–E68, E70–E90
38	Blood diseases	D50–D89
39	Dementia, vascular, senile or unspecified	F01, F03
40	Alcohol abuse	F10
41	Drug abuse	F11–F19
42	Other mental disorders	F04–F09, F20–F99
43	Systemic atrophies and demyelinating diseases of the central nervous system	G10–G12, G35–G37
44	Parkinson's disease and other extrapyramidal and movement disorders	G20–G25
45	Alzheimer's disease and other degenerative diseases of the nervous system	G30, G31
46	Epilepsy	G40–G41
47	Other diseases of nervous system	G00–G09, G43–G44, G47–G83, G90–G99, H00–H95
48	Rheumatic heart diseases	I00–I09
49	Essential hypertension	I10
50	Hypertensive disease (heart, kidney and secondary)	I11–I15
51	Acute myocardial infarction	I21–I23
52	Atherosclerotic cardiovascular and heart diseases	I25.0, I25.1
53	Other IHD	I20, I24, I25.2 to .9
54	Pulmonary heart diseases	I26–I28
55	Non rheumatic valve disorders	I34–I38
56	Cardiac arrest	I46
57	Heart failure	I50
58	Other heart diseases	I30–I33, I40–I45, I47–I49, I51
59	Intracranial hemorrhage	I60–I62
60	Cerebral infarction, occlusion, and stenosis	I63, I65, I66
61	Other cerebrovascular diseases	G45, I64, I67
62	Sequelae of cerebrovascular disease	I69
63	Diseases of arteries, arterioles and capillaries	I70–I78
64	Other circulatory diseases	I80–I99
65	Influenza	J09–J11
66	Pneumonia	J12–J18
67	Other acute respiratory infections	J00–J06, J20–J22, U04
68	Asthma	J45–J46

| 69	Other chronic obstructive pulmonary disease	J40–J44, J47
70	Pneumonitis due to solids and liquids	J69
71	Pneumoconiosis and chemical effects	J60–J68, J70
72	Other respiratory diseases, principally affecting the interstitium	J80–J84
73	Other diseases of the respiratory system	J30–J39, J85–J98
74	Gastric and duodenal ulcer	K25–K28
75	Hernia	K40–K46
76	Enteritis, colitis and other intestinal diseases	K35–K38, K50–K63
77	Alcoholic cirrhoses of liver	K70
78	Other cirrhoses of liver	K74
79	Other diseases of liver	K71–K73, K75, K76
80	Cholelithiasis and other disorders of biliary tracts	K80–K83
81	Diseases of pancreas	K85–K86
82	Other digestive diseases	K00–K22, K29–K31, K65–K66, K90–K92
83	Diseases of skin and subcutaneous tissue	L00–L98
84	Diseases of the musculoskeletal system and connective tissue	M00–M99
85	Renal tubulo-interstitial diseases	N00–N15
86	Renal failure	N17–N19
87	Other diseases of urinary system	N20–N36, N39
88	Diseases of genital organs	N40–N99
89	Complications of pregnancy, childbirth, and puerperium	O00–O99
90	Certain conditions originating in the perinatal period	P00–P96
91	Congenital malformations, deformations, and chromosomal abnormalities	Q00–Q99
92	Sudden infant death syndrome (SIDS)	R95
93	Transport accidents	V01–V99
94	Accidental falls	W00–W19
95	Accidental drowning and submersion	W65–W74
96	Accidental exposure to smoke, fire and flames	X00–X09
97	Accidental poisoning by alcohol	X45
98	Accidental poisoning by other substance	X40–X44, X46–X49
99	Other accidental threats to breathing	W75–W84
100	Suicide and self-inflicted injury	X60–X84
101	Assault	X85–Y09, Y35, Y36
102	Event of undetermined intent	Y10–Y34
103	Complications of medical and surgical care	Y40–Y84
104	Other accidents and late effects of accidents (remainder)	W20–W64, W85–W99, X10–X39, X50–X59, Y85–Y91, Y95–Y98

Appendix G: The Tool

A tool has been created to forecast death rates for both the national and the cluster population. It relies on the same assumptions as the by-cause model presented in this report but it also offers the option for a user to input external opinions about the future pattern of causes-of-deaths. Two ways exist for doing so: modifying the complete structure of the forecast for a cause or targeting an evolution of the death rates at a given horizon. These opinions affect directly the deaths forecasts of the baseline population (national population) and modify indirectly the deaths forecasts of the cluster population through the relational model.

Two examples of expert opinion have been preloaded in two external files, one for drug, the other for neoplasms. In the drug example, the complete structure of the forecast has been changed by imposing flat rates for all ages and years (adjustment "yes" in the tab "Main" and values at 100% in the tab "Adjustment Drug", which means a flat evolution of death rates over the years for all ages), in order to model a stagnation of mortality related to that cause instead of an increase. This also implies indirectly a trend at 0 for the cluster population, based on the relational model. The result is a flat evolution of the death rates of the cause drug for the national population, an indirect flat evolution of the death rates for the cluster population and the same cause through the relational model, and an increase in life expectancy for both the national and the cluster populations as shown below.

Figure 70
TOTAL LIFE EXPECTANCY FORECASTS AT BIRTH WITH FLAT NATIONAL DRUG RATES
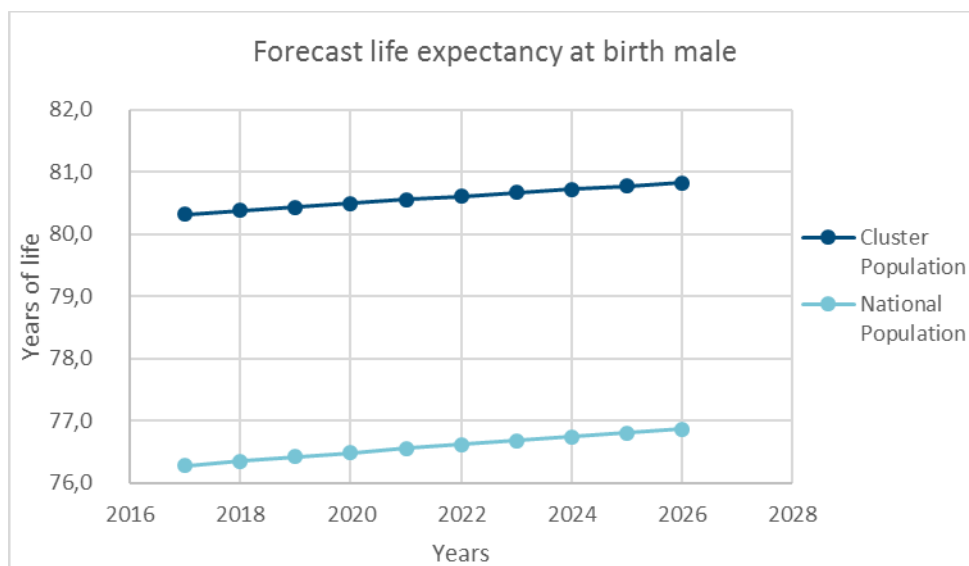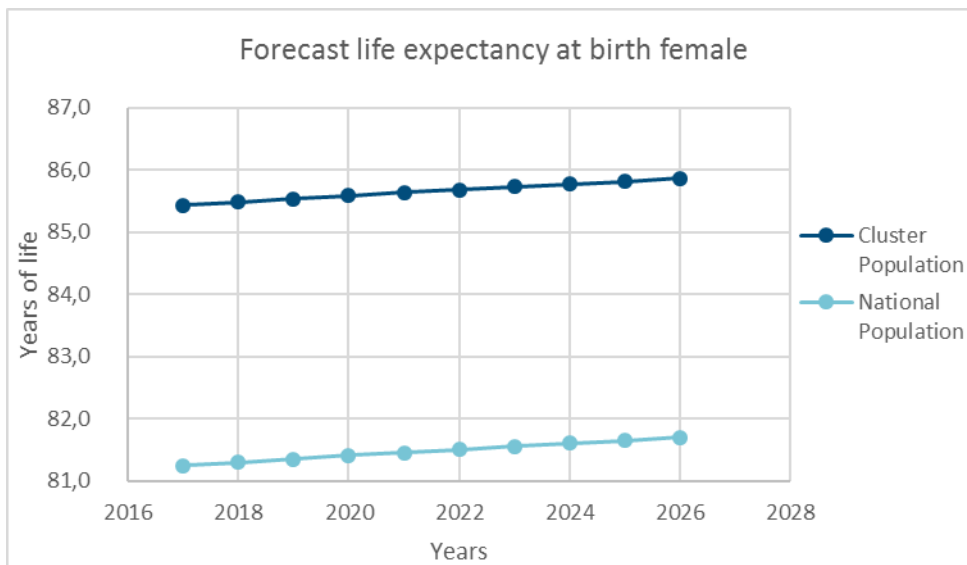
Forecast life expectancy at birth female

The total life expectancy at birth is expected to increase instead of stagnate for both males and females. It increases by 0.5 years between 2017 and 2026.

In the neoplasms example, we have targeted an evolution of the death rates of −20% at 10 years' horizon for ages 80–84 (type of trend "manual", values at −20%, horizon 10, and ages 80–84 in the tab "Main", which means a death rate of −20% ten years later for the cause neoplasms and ages 80–84). This implies of change of trend for the cause neoplasms and thus affects all ages and years for this cause. The result is a decrease of the death rates of the cause neoplasms for the national population at all ages, an indirect decrease of the death rates for the cluster population and the same cause through the relational model, and an increase in life expectancy for both the national and the cluster populations.

Figure 71

TOTAL LIFE EXPECTANCY FORECASTS AT BIRTH WITH A DECREASE OF −20% OF NATIONAL NEOPLASMS RATES AT AGES 80–84 AND AT 10 YEARS HORIZON
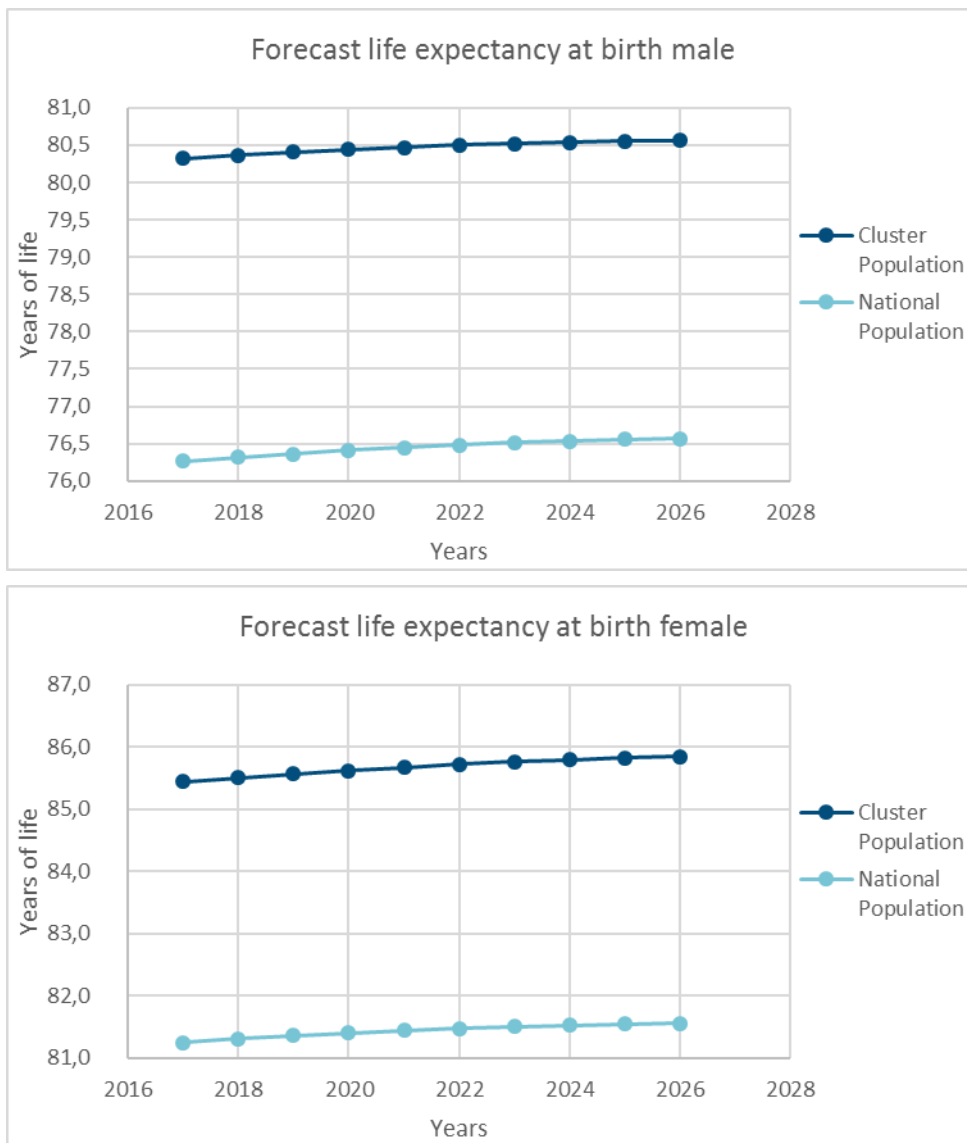

The total life expectancy at birth is expected to increase instead of stagnate for both males and females. It increases by 0.3 years between 2017 and 2026.

For more details about the tool, see the User Guide joint with the tool.

References

Alai, Daniel H., Séverine, Arnold, Madhavi, Bajekal, and Andrés M., Villegas. 2018. Mind the Gap: A Study of Cause-Specific Mortality by Socioeconomic Circumstances. *North American Actuarial Journal*, 22, no. 2:161–181.

American Cancer Society. 2018. Cancer Facts & Figures 2018. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html (accessed on October 17, 2019).

Arnold, Séverine, Alexandre, Boumezoued, Héloïse Labit, Hardy, Nicole, El Karoui. (2018). Cause-of-Death Mortality: What Can Be Learned from Population Dynamics?. *Insurance: Mathematics and Economics*, 78:301–315.

Beck, Thorsten, and Webb, Ian. 2003. Economic, Demographic, and Institutional Determinants of Life Insurance Consumption across Countries, *The World Bank Economic Review*, 17, no. 1:51–88.

Bergeron-Boucher, Marie-Pier, Vladimir, Canudas-Romo, Jim, Oeppen, and James W., Vaupel. 2017. Coherent Forecasts of Mortality with Compositional Data Analysis. *Demographic Research*, 37:527–566.

Brouhns, Natacha, Michael, Denuit, and Jeroen K, Vermunt. 2002. A Poisson Log-Bilinear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and economics*, 31, no. 3:373–393.

Cairns, Andrew J.G., David, Blake, Kevin, Dowd, Guy D., Coughlan, David, Epstein, and Marwa, Khalaf-Allah. 2011. Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics and Economics*, 48, no. 3:355–367.

Cairns, Andrew J., David, Blake, Kevin, Dowd, Guy, Coughlan, David, Epstein, et al. 2009. A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States. *North American Actuarial Journal*, 13, no. 1:1–35.

Canudas-Romo, Vladimir, Eva, DuGoff, Albert W., Wu, Saifuddin, Ahmed, and Gerard, Anderson. 2016. Life Expectancy in 2040: What Do Clinical Experts Expect?. *North American Actuarial Journal*, 20, no. 3:276–285.

Case, A., & Deaton, A. 2016. Reply to Schmid, Snyder, and Gelman and Auerbach: Correlates of the Increase in White Non-Hispanic Midlife Mortality in the 21st century. *Proceedings of the National Academy of Sciences*, 113, no. 7:E818–E819.

Case, Anne, and Angus, Deaton. 2015. Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century. *Proceedings of the National Academy of Sciences*, 112, no. 49:15078–15083.

Caselli, G., Vallin, J., and Marsili, M. 2019. How Useful are the Causes-of-death When Extrapolating Mortality Trends. An Update. *Old and New Perspectives on Mortality Forecasting,* 237–259. Switzerland: Springer Open.

Centers for Disease Control and Prevention (CDC). National Vital Statistics System. United States. 12 Month-ending Provisional Number of Drug Overdose Deaths. www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm (accessed January 2020).

Chetty, Raj, Stepner, Michael, Abraham, Sarah, et al. 2016. The Association Between Income and Life Expectancy in the United States, 2001–2014. *JAMA*, 315, no. 16:1750–1766.

Czado, Claudia, Antoine, Delwarde, and Michel, Denuit. 2005. Bayesian Poisson Log-bilinear Mortality Projections. *Insurance: Mathematics and Economics*, 36, no. 3:260–284.

Dimitrova, Dimitrina, Steven, Haberman, and Vlademir K., Kaishev. 2013. Dependent Competing Risks: Cause Elimination and Its Impact on Survival. *Insurance: Mathematics and Economics*, 53, no. 2:464–477.

Dowd, Kevin, Andrew J.G., Cairns, David, Blake, Guy, Coughlan, David, Epstein, et al. 2010. Evaluating the Goodness of Fit of Stochastic Mortality Models. *Insurance: Mathematics and Economics*, 47, no. 3:255–265.

Feigin, Valery L., Gregory A., Roth, Mohsen, Naghavi, Priya, Parmar, Rita, Krishnamurthi, et al. 2016. Global Burden of Stroke and Risk Factors in 188 Countries, During 1990–2013: A Systematic Analysis for the Global Burden of Disease Study 2013. *The Lancet Neurology*, 15, no. 9:913–924.

Foreman, Kyle J., Neal, Marquez, Andrew, Dolgert, Kai, Fukutaki, Nancy, Fullman, et al. 2018. Forecasting Life Expectancy, Years of Life Lost, and All-Cause and Cause-Specific Mortality for 250 Causes-of-death: Reference and Alternative Scenarios for 2016–40 for 195 Countries and Territories. *The Lancet*, 392, no. 10159:2052–2090.

Frees, Edward W., and Sun, Yunjee. 2010. Household Life Insurance Demand -a Multivariate Two-Part Model. *North American Actuarial Journal*, 14, no. 3:338–354.

Gelman, Andrew, and Jonathan, Auerbach. 2016. Age-Aggregation Bias in Mortality Trends. *Proceedings of the National Academy of Sciences*, 113, no. 7:E816–E817.

Glushko, Viktoriya, and Séverine, Arnold. 2018. Common Factor Decomposition of Cause-Specific Mortality Rates Using the Cointegration Analysis. Presented at the Longevity 14 Conference, Amsterdam, September 2018.

Heo, Wookjae, Grable, John E., Chatterjee, Swan. 2013. Life insurance consumption as a function of wealth change. *Financial Services Review*, 22, no.4:389–404.

Human Cause-of-Death Database. French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany). www.causeofdeath.org (accessed January 2019).

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org or www.humanmortality.de (accessed January 2019).

Institute for Health Metrics and Evaluation (IHME). GBD Results tool. Seattle, WA: IHME, University of Washington, 2017. Available from http://ghdx.healthdata.org/gbd-results-tool (accessed January 2019).

Kjærgaard, Søren, Yunus E., Ergemen, Malene, Kallestrup-Lamb, Jim, Oeppen, and Rune, Lindahl-Jacobsen. 2018. Forecasting Causes-of-death Using Compositional Data Analysis: The Case of Cancer Deaths. Presented at the Fourteenth International Longevity Risk and Capital Markets Solutions Conference (Longevity 14), Amsterdam, September 2018.

Koehler, Anne B., Ralph, Snyder, J. Keith, Ord, and Adrian, Beaumont. 2010. Forecasting Compositional Time Series with Exponential Smoothing Methods (No. 20/10). Monash University, Department of Econometrics and Business Statistics.

Lee, Ronald D., and Lawrence, Carter. 1992. Modeling and Forecasting US Mortality. *Journal of the American Statistical Association*, 87, no. 419:659–671.

Li, Han, Hong, Li, Yang, Lu, Y., and Anastasios, Panagiotelis. 2019. A Forecast Reconciliation Approach to Cause-of-Death Mortality Modeling. *Insurance: Mathematics and Economics*, 86, no. 5:122–133

Lourés, Cristian Redondo, and Cairns, Andrew JG. Mortality In The US By Education Level. 2019. http://www.macs.hw.ac.uk/~andrewc/ARCresources/RedondoCairns2019.pdf

Mokdad, A. H., Ballestros, K., Echko, M., et al. 2018. The state of US health, 1990–2016: burden of diseases, injuries, and risk factors among US states. *Jama*, 319, no.14:1444–1472.

Murray, Christopher J., and Alan, Lopez. 1997. Alternative Projections of Mortality and Disability by Cause 1990–2020: Global Burden of Disease Study. *The Lancet*, 349, no. 9064:1498–1504.

Nelson, Sue, Laurie, Whitsel, Olga, Khavjou, Diane, Phelps, and Alyssa, Leib, A. 2016. Projections of Cardiovascular Disease Prevalence and Costs. RTI International. https://healthmetrics.heart.org/wp-content/uploads/2017/10/Projections-of-Cardiovascular-Disease.pdf (accessed October 17, 2019).

Oeppen, Jim. 2008. Coherent Forecasting of Multiple-Decrement Life Tables: A Test Using Japanese Cause-of-death Data. Paper presented at the European Population Conference 2008, July 9–12, 2008, Barcelona, Spain.

Piveteau, Samuel, and Julien, Tomas. 2018. Mortality Forecasting by Cause-of-death and Basis Risk Modelling With Compositional Data. Presented at the Longevity 14 Conference, Amsterdam, September 2018.

Priyadarshana, W.J.R.M., and Georgy, Sofronov. 2015. Multiple Break-Points Detection in Array CGH Data Via the Cross-Entropy Method. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12, 2:487–498.

Rahib, L.ola, Benjamin, Smith, Rhonda, Aizenberg, Allison, Rosenzweig, Julie, Fleshman, et al. 2014. Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Research*, 74:11:2913–2921.

Richards, S.J. 2009. Selected Issues in Modelling Mortality by Cause and in Small Populations. *British Actuarial Journal*, 15, no. S1:267–283.

Roth, Gregory, Degu, Abate, Kalkidan, Hassan Abate, Solomon, Abay, Cristiana, Abbafati, et al. 2018. Global, Regional, and National Age-Sex-Specific Mortality for 282 Causes-of-death in 195 Countries and Territories, 1980–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392, no. 10159:1736–1788.

Salhi, Yahia, and Pierre-Emmanuel, Thérond. 2018. Age-Specific Adjustment of Graduated Mortality. *ASTIN Bulletin: The Journal of the IAA*, 48, 2:543–569.

Salhi, Yahia, Pierre-Emmanuel, Thérond, and Julien, Tomas. 2016. A Credibility Approach of the Makeham Mortality Law. *European Actuarial Journal*, 6, no. 1:61–96.

Society of Actuaries. 2019. Economic Impact of Non-Medical Opioid Use in the United States. Research report.

Society of Actuaries. 2019. U.S. Population Mortality Observations – Updated with 2017 Experience. https://www.soa.org/resources/research-reports/2018/population-mortality-observations/ (accessed October 17, 2019).

Society of Actuaries. 2019. Modeling and Forecasting Cause-of-Death Mortality. Research report.

Sung, Hyuna, Rebecca, Siegel, Phillip, Rosenberg, and Ahmedin, Jemal. 2019. Emerging Cancer Trends Among Young Adults in the USA: Analysis of a Population-Based Cancer Registry. *The Lancet Public Health*, 4: no. 3,137–147

Villegas, Andrés M., Vladimir K. Kaishev, Pietro Millossovich. 2018. StMoMo: An R package for stochastic mortality modeling. Journal of Statistical Software 84(3).

Weir, Hannah K., Trevor, Thompson, Ashwini, Soman, Bjorn, Møller, Steven, Leadbetter, et al. 2015. Peer Reviewed: Meeting the Healthy People 2020 Objectives to Reduce Cancer Mortality. *Preventing Chronic Disease*, 12.

Weir, Hannah K., Robert, Anderson, Sallyann, Coleman King, Ashwini, Soman, Trevor, Thompson, et al. 2016. Peer Reviewed: Heart Disease and Cancer Deaths—Trends and Projections in the United States, 1969–2020. *Preventing Chronic Disease,* 13.

Wilmoth, John R. 1995. Are Mortality Projections Always More Pessimistic When Disaggregated by Cause-of-death?. *Mathematical Population Studies*, 5, no. 4:293–319.

World Health Organization. 2004. *International statistical classification of diseases and related health problems (Vol. 1).* World Health Organization, Geneva.

About The Society of Actuaries

With roots dating back to 1889, the Society of Actuaries (SOA) is the world's largest actuarial professional organization with more than 31,000 members. Through research and education, the SOA's mission is to advance actuarial knowledge and to enhance the ability of actuaries to provide expert advice and relevant solutions for financial, business and societal challenges. The SOA's vision is for actuaries to be the leading professionals in the measurement and management of risk.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org