

Impact of population aging on cancer incidence and mortality in South Korea: projection model-based study

Hoejun Kwon

**A Thesis Submitted to the Department of Cancer AI &
Digital Health in Partial Fulfillment of the Requirements
for the Degree of Master in Public Health**

August 2023

**National Cancer Center
Graduate School of Cancer Science and Policy**

Professor Hyunsoon Cho
Department of Cancer AI & Digital Health
National Cancer Center Graduated School of Cancer Science and Policy (PhD)

**We hereby approve the M.P.H. thesis of
Hoejun Kwon.**

August 2023

Jin Kyoung Oh



Chairman of Thesis Committee

Department of Cancer Control and Population Health

National Cancer Center Graduated School of Cancer Science and Policy (MPH, PhD)

Chang Mo Oh



Thesis Committee Member

Department of Preventive Medicine

Kyung Hee University College of Medicine (M.D, PhD)

Hyunsoon Cho



Thesis Committee Member

Department of Cancer AI & Digital Health

National Cancer Center Graduated School of Cancer Science and Policy (PhD)

**NATIONAL CANCER CENTER
GRADUATE SCHOOL OF CANCER SCIENCE AND POLICY**

ABSTRACT

Impact of population aging on cancer incidence and mortality in South Korea: projection model-based study

This study focuses on projecting cancer incidence and mortality rates in Korea, with an emphasis on the role of demographic factors. The study compares the differences of various models, including population structure-based and non-population structure-based models, in predicting cancer trends by gender and cancer type. The models used in the analysis include the NP, BAPC, JP, ARIMA, SSML. The study also utilizes APC-analysis to identify age-related cancers and quantify the contribution of aging factors to the increase in cancer incidence and mortality rates.

The study's results reveal variations in the rate of increase and decrease of cancer incidence and mortality rates by cancer type and demographic group across scenarios. The NP model consistently demonstrated the most realistic trends by gender and cancer type. The decomposition analysis consistently identified aging as a major factor contributing to the increase in cancer incidence or mortality. The study identifies certain limitations and challenges associated with predicting cancer incidence and mortality rates and underscores the importance of selecting appropriate models and validating them to obtain reliable

and accurate predictions.

Overall, this study highlights the significant role of aging and other demographic factors in cancer incidence and mortality rates and provides insights into the potential impact of these factors on future cancer trends in Korea. The findings have important implications for cancer prevention and control efforts and can help inform policy decisions aimed at reducing the burden of cancer in aging societies. The study's approach and methodology can also be applied to other settings to project cancer trends and quantify the contribution of demographic factors to cancer burden.

Copyright by

Hoejun Kwon

2023

Contents

ABSTRACT	i
List of Tables.....	ix
List of Figures	xii
1. INTRODUCTION	1
1.1 Background Information	1
1.2 Objectives & Methodology.....	4
1.3 Structure of the Thesis	8
2. LITERATURE REVIEW	9
2.1 Introduction.....	9
2.2 Decomposition Analysis	10
2.2.1 Das Gupta's Method	10
2.2.2 Bashir's Method.....	12
2.2.3 Cheng's Method	14
2.3 Population structure-based projection models	18
2.3.1 APC (Age-Period-Cohort) Model.....	17
2.3.2 NORDPRED (NP) Model.....	20

2.3.3 Bayesian (Bayesian Age–Period–Cohort) Model.....	22
2.4 Population structure-based projection models	24
2.4.1 JOINPOINT (JP) Model	24
2.4.2 Time Series Forecasting Model.....	26
2.4.2.1 ARIMA (Autoregressive Integrated Moving Average) Model.....	26
2.4.2.2 Space-State Model.....	28
2.4.2.2.1 Linear Space State Model (SSML)	29
2.5 Summary	30
3. SELECTION AGEING RELATED CANCER AND CURRENT TREND ANALYSIS IN 2000 TO 2019.....	32
3.1 Introduction.....	32
3.2 Selection of ageing-related cancer	32
3.2.1 Selection of ageing-related cancer with incidence.....	33
3.2.2 Selection of ageing-related cancer with mortality	39
3.2.2.2 Final Selection of ageing-related cancer with incidence and mortality	45
3.3 Validation selection of ageing-related cancer.....	50

3.3.1 Age-Period-Cohort Effect in 2000 to 2019.....	50
3.3.1.1 Validation male incidence and mortality rates	51
3.3.1.2 Validation female incidence and mortality rates	61
3.4 Summary	69
4. COMAPRISON BY PROJECITON MODELS WITH AND WITHOUT POPULATION STRUCTURE EFFECTS.....	70
4.1 Introduction	70
4.2 Population structure-based model.....	71
4.3 Non-population structure-based model	73
4.4 Projection results between population structure-based model and non-population structure-based model.....	74
4.4.1 Projection results for male incidence and mortality rates by models	74
4.4.2 Projection results for female incidence and mortality rates by models	
.....	80
4.5 Comparison by model characteristics based on outcomes by cancer type	85

4.6 Summary	90
5. FUTURE PREDICITONS ACCORDING TO POPULATION	
AGEING SCENARIOS...	92
5.1 Introduction	92
5.2 Scenarios setting	92
5.3 Applying decomposition analysis in future projection	94
5.4 Comparison of projection according to prospective population projection scenarios.....	97
5.4.1 Comparison male incidence and mortality by scenarios.....	97
5.4.2 Comparison female incidence and mortality by scenarios.....	103
5.5 Comparison of decomposition about projection according to prospective population projection scenarios	109
5.5.1 Comparison decompose male incidence and mortality by scenarios.....	109
5.5.2 Comparison decompose female incidence and mortality by scenarios.....	122
5.6 Summary	134
6. CONCLUSION...	135

BIBLIOGRAPHY.....	139
ACKNOWLEDGEMENT... ..	148

List of Tables

Table 1. Meaning of mathematical symbols in the decomposition formula	15
Table 2. Male and Interaction effect formula of each year	15
Table 3. Decomposition formula about main effect.....	16
Table 4. APC model on and Age-Period Array.....	19
Table 5-1. Cancer incidence and rank in all age by sex in 2019	35
Table 5-2. Cancer incidence and rank in all 65+ age by sex in 2019.....	36
Table 5-3. Cancer incidence rate ratio in age groups by sex in 2019.....	37
Table 5-4. JOINPOINT Average Annual Percent Change (AAPC) shifting last 10 years period in incidence by sex	38
Table 6-1. Cancer mortality and rank in all age by sex in 2019	41
Table 6-2. Cancer mortality and rank in all 65+ age by sex in 2019.....	42
Table 6-3. Cancer mortality rate ratio in age groups by sex in 2019	43
Table 6-4. JOINPOINT Average Annual Percent Change (AAPC) shifting last 10 years period in mortality by sex.....	44
Table7-1 Cancer selection for the model projection with adjustment on ageing population in incidence by sex.....	47

Table7-2 Cancer selection for the model projection with adjustment on ageing population in mortality by sex.	48
Table7-3 Cancer selection for the model projection with adjustment on ageing population in incidence and mortality by sex.	49
Table 8. Key scenarios for prospective population projections.....	94
Table 9-1. Male incidence rate of each cancer type by year by scenarios	101
Table 9-2. Male mortality rate of each cancer type by year by scenarios	102
Table 10-1. Female incidence rate of each cancer type by year by scenarios	107
Table 10-2. Female mortality rate of each cancer type by year by scenarios	108
Table 11-1. Male incidence case difference based on population aging risk	118
Table 11-2. Male mortality case difference based on population aging risk	119
Table 12-1. Net change by male incidence case.....	120
Table 12-2. Net change by male mortality case	121

Table 13-1. Female incidence case difference based on population aging risk	130
Table 13-2. Female mortality case difference based on population aging risk	131
Table 14-1. Net change by female incidence case.....	132
Table 14-2. Net change by female mortality case	133

List of Figures

Figure 1. Ageing related cancer selection framework	7
Figure 2-1. Age-Period-Cohort Effect for Male Incidence (Esophagus, Stomach, Colorectal Cancer).....	53
Figure 2-2. Age-Period-Cohort Effect for Male Incidence (Liver, Gallbladder, Pancreas Cancer)	54
Figure 2-3. Age-Period-Cohort Effect for Male Incidence (Lung, Prostate, Bladder Cancer).....	55
Figure 2-4. Age-Period-Cohort Effect for Male Incidence (Kidney Cancer, Non-Hodgkin Lymphoma)	56
Figure 3-1. Age-Period-Cohort Effect for Male Mortality (Esophagus, Stomach, Colorectal Cancer).....	57
Figure 3-2. Age-Period-Cohort Effect for Male Mortality (Liver, Gallbladder, Pancreas Cancer)	58
Figure 3-3. Age-Period-Cohort Effect for Male Mortality (Lung, Prostate, Bladder Cancer).....	59
Figure 3-4. Age-Period-Cohort Effect for Male Mortality (Kidney Cancer, Non-Hodgkin Lymphoma)	60

Figure 4-1. Age-Period-Cohort Effect for Female Incidence (Stomach, Colorectal, Liver Cancer)	63
Figure 4-2. Age-Period-Cohort Effect for Female Incidence (Gallbladder, Pancreas, Lung Cancer).....	64
Figure 4-3. Age-Period-Cohort Effect for Female Incidence (Cervix Uteri Cancer, Non-Hodgkin Lymphoma, Leukemia).....	65
Figure 5-1. Age-Period-Cohort Effect for Female Mortality (Stomach, Colorectal, Liver Cancer)	66
Figure 5-2. Age-Period-Cohort Effect for Female Mortality (Gallbladder, Pancreas, Lung Cancer).....	67
Figure 5-3. Age-Period-Cohort Effect for Female Mortality (Cervix Uteri Cancer, Non-Hodgkin Lymphoma, Leukemia).....	68
Figure 6-1. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Esophagus, Stomach, Colorectal, Colorectal & Anus, Liver Cancer).....	77
Figure 6-2. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-	

population structure-based models (Gallbladder, Pancreas, Lung, Prostate Cancer).....	78
Figure 6-3. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Kidney, Bladder Cancer, Non-Hodgkin Lymphoma)	79
Figure 7-1. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer).....	82
Figure 7-2. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Pancreas, Lung, Cervix uteri Cancer, Non-Hodgkin lymphoma).....	83
Figure 7-3. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Leukemia)	84
Figure 8-1. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Esophagus, Stomach, Colorectal, Colorectal & Anus, Liver Cancer).....	98

Figure 8-2. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer).....99

Figure 8-3. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Kidney, Bladder Cancer, Non-Hodgkin Lymphoma).....100

Figure 9-1. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer).....104

Figure 9-2. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Pancreas, Lung, Cervix Uteri, Cancer, Non-Hodgkin Lymphoma).....105

Figure 9-3. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Leukemia).....106

Figure 10-1. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Esophagus, Stomach, Colorectal, Colorectal & Anus, Liver Cancer)112

Figure 10-2. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer).....	113
Figure 10-3. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Kidney, Bladder Cancer, Non-Hodgkin lymphoma)	114
Figure 11-1. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Esophagus, Stomach, Colorectal, Colorectal & Anus, Liver Cancer)	115
Figure 11-2. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer).....	116
Figure 11-3. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer).....	117
Figure 12-1. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer)	124

Figure 12-2. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Pancreas, Lung, Cervix Uteri Cancer, Non-Hodgkin lymphoma)	125
Figure 12-3. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Leukemia)	126
Figure 13-1. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer)	127
Figure 13-2. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Pancreas, Lung, Cervix Uteri Cancer, Non-Hodgkin lymphoma)	128
Figure 13-3. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Leukemia)	129

CHAPTER I. INTRODUCTRION

1.1 Background Information

Cancer is an important health problem worldwide and is causing a persistent cancer burden. As of 2019, cancer was found to be the leading cause of death in Korea, accounting for 27.5% of all deaths [1] and cancer incidence also increased 3.6% compared to 2018, recording a high incidence of 257,170[2]. Increased cancer burden is known to have a particular effect on age dependence, which is supported by a high proportion of older people in cancer deaths and outbreaks [3]. This fact can be said to be greatly influenced by the demographic structure of Korea, where the aging rate of Korea is increasing. In Korea, the number of people aged 65 or older reached 14% or more in 2014, maintaining an aging society to date, which is the fastest aging rate in the world, and is expected to enter the super-aged society in the next few years [4]. Based on 2019 cancer incidence statistics, cancer incidence among people aged 65 years and older accounts for nearly half of all cancers diagnosed in South Korea.[2] Cancer poses significant economic costs in addition to the health burden. According to data from the Health Insurance Review and Assessment Service, between 2013 and 2019, the economic burden on cancer care in South Korea increased from \$7.1 billion a year to \$12.1 billion, with an average annual increase of 9.3% [5]. These facts demonstrate the importance of quantitatively and qualitatively measuring future cancer burdens. However, the

primary measurement of cancer burden based on the trend of incidence and mortality is necessary to assess the cancer burden resulting from other social and economic factors.

Cancer incidence and mortality rates are estimated using statistical models to estimate short-term or long-term future trends through past trends. In previous models used to predict the incidence and rate of cancer, simple approaches that assume a constant rate of increase were used [6, 7]. Generalized linear models (GLM) [8-10] based on age-period effects or generalized additive models (GAM) that include single or multivariate smoothers in specific linear predictors for age or period variables were also used.[11]. In recent years, the development of the model has been based on the age-periodic cohort (APC) model, which further integrates the cohort effect [12, 13], and furthermore, the age-shift-period-cohort (Nordpred) [14] model has been used. A model that predicts trends based on the joint point model has also been used as a representative model [15]. Not only the model constructed using the elements of the Lexis diagram, but also the time series prediction model, which can be said to be an act of predicting the future by understanding the past, has been continuously used [16]. These time series models have been widely disseminated and developed as essential importance of time series prediction in numerous practical fields such as business, economy, finance, science, and engineering [17,18,19], and various derivative models have also been developed.

In Korea, a variety of projection models have been employed to forecast cancer incidence and mortality rates. Prior to the Major use of APC models and JOINPOINT models, research primarily relied on a Poisson assumption to estimate incidence or mortality rates [20]. However, the APC and JOINPOINT models have demonstrated superior predictive power and reliability, leading to their widespread adoption. Consequently, long-term predictions in Korea have been made using APC-based models [21,22,23,24], while mid- and short-term predictions have been conducted with JOINPOINT models [25,26]. Despite the utilization of trend prediction models to assess the cancer burden in Korea, there remain limitations in their application to specific cancer types, and research comparing the accuracy of various models is still lacking. Furthermore, estimates derived solely from predictive models may not adequately reveal the extent to which increases in incidence or mortality rates can be attributed to aging. To elucidate the impact of aging on cancer incidence and mortality, it is crucial to disentangle the effects of aging from these rates. Decomposition analysis serves as a valuable method for addressing this challenge [27]. Previous Studies, researchers have sought to identify the effects of aging on cancer incidence and mortality rates by employing a combination of projection models and decomposition analysis [28,29,30]. As such, it is imperative to assess the actual influence of aging in the Korean context by utilizing a method that merges projection models and decomposition analysis for predicting cancer incidence and mortality rates. This approach will facilitate the separation of aging effects on incidence or mortality, thereby providing a more profound understanding of aging's impact on the cancer burden. Moreover,

previous studies might not fully account for alterations in cancer incidence and mortality rates due to potential changes in population structure, which could arise because of Korea's pace of aging in the future. These studies may also not adequately address uncertainties regarding future population changes. Consequently, it is essential to comprehensively grasp the relationship between aging and cancer burden in Korea by applying various scenarios of population projection estimates to diverse models.

1.2. Objectives & Methodology

The purpose of this study is to analyze the status of domestic cancer incidence and mortality rates using statistical methods and to measure the impact of aging through projection models and decomposition analysis. Based on observed data from 2000 to 2019, this study applies various models to perform long-term predictions up to 2040, focusing on identifying trends. Through decomposition analysis using model-specific prediction points, we can measure the impact of aging on incidence and mortality rates for both men and women.

Cancer incidence/mortality, population, and prospective population projection data were obtained from the National Statistical Portal of the National Statistical Office (KOSIS). This dataset provided information on personal characteristics, such as age (in 5-year) and sex, from 2000 to 2019. Although cancer incidence data is available up to 2020 and mortality data up to 2021, it has been reported that

data for 2020 and 2021 may be underestimated due to the impact of the COVID-19 pandemic [31]. As a result, this study utilizes pre-pandemic data to account for any potential biases during that period. In the case of domestic data, incidence and mortality data are classified or grouped by domain according to the characteristics of each cancer type. This study investigates incidence and mortality based on 24 cancer types commonly used in domestic cancer research. However, other cancer types (Re.C00-C96/C97) and other benign tumors (D00-D48) are excluded due to their unclear classification. And Breast cancer (C50) in males is an uncommon clinical manifestation; therefore, male breast cancer patients are excluded from the study [32].

In measuring the aging effect on each cancer type, it is assumed that not all cancer types are related to aging. Thus, we aimed to pre-select cancer types that were associated with aging. In this study, we adapted and modified the criteria used in previous research to select cancer types suitable for aging in the domestic context. The previous studies employed the ratio of age group-specific incidence rate differences as a basis for determining the predominance of cancer types occurring in the elderly population aged 65 or older [33]. A cancer type was considered age-related if its incidence or mortality rate in the elderly (65 years and older) was at least twice as high as that in other age groups, such as young (00-34 years, comparison1), middle-aged (35-49 years, comparison2), and older-aged (50-64 years, comparison3). However, due to advancements in domestic cancer management policies and treatments, the difference between the elderly and older-

aged populations may have narrowed. To account for this, we modified the criteria so that a difference of at least 1-fold between the older-aged and the elderly would suffice. Additionally, the concept of shifting was also taken into consideration. As mentioned in the same study previously, it was stated that the shifting phenomenon occurs when the incidence and mortality rates of those aged 65 and older continue to decrease, while the rates for other age groups increase. This phenomenon signifies a reversal of cancer incidence or mortality from older to younger age groups. However, the incidence and mortality rates of various types of cancer in the country do not consistently show decreasing or increasing trends, making it difficult to quantitatively assess. Therefore, the JOINPOINT model was employed to estimate the Annual Average Percent Change (AAPC), and conditions were set to identify cancer types that increased in those aged 65 and older but decreased in other age groups [34]. Moreover, we included a condition that required both the overall age-specific incidence/mortality rates and the incidence/mortality rates for those aged 65 and older to be within the top 10.

While the commonly accepted framework necessitates conducting analyses separately, employing solely the cancer species selected based on predefined criteria, it may not provide a comprehensive overview. From a research standpoint, the isolated consideration of incidence and mortality characteristics of cancer species could potentially lead to an overly constrained perspective. Analyzing the incidence or mortality in isolation can restrict the scope of inquiry, as each provides only a partial insight into the disease dynamics. Consequently, to attain a

holistic understanding, it becomes imperative to incorporate both these parameters while assessing cancer types. Therefore, the incidence and mortality rates determined by the above conditions were further adjusted to equalize the selected cancer types so that the incidence and mortality rates per cancer type could be compared to each other (Figure1).

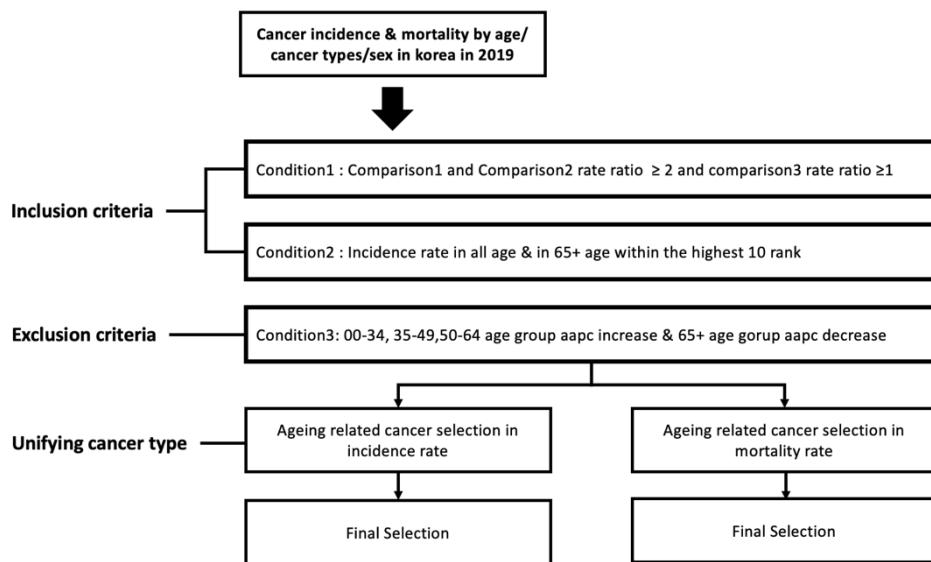


Figure 1. Ageing related cancer selection framework

1.3 Structure of the Thesis

Chapter 1 introduces the background, purpose, and overall overview of the study. Chapter 2 introduces decomposition analysis methods and projection models used for prediction. Chapter 3 presents the status of cancer incidence and death in Korea and presents the process and results of selection aging related cancers. Chapter 4 comparison between the long-term projection of population structure-based models and non-population structure-based models. Chapter 5 compares estimates across models for different population aging scenarios and uses decomposition to quantify the contribution of ageing factors. Chapter 6 presents model-specific results comparisons and suggestions and future studies. Each chapter provides a summary, except for the first and last chapters.

CHAPTER II. LITERATRE REVIEW

2.1 Introduction

This chapter presents a brief literature review of decomposition analysis and cancer prediction models, including major publications in other journals. Decomposition analysis has been developed to quantitatively assess cancer incidence or mortality. Its primary purpose is to evaluate the impact of changes in population structure on incidence or mortality rates.

The decomposition analysis is divided into three factors: cancer risk, age distribution, and population structure, which help compare the differences in incidence or mortality. By setting a reference year and a comparison year, it is possible to determine the difference contribution of each factor to incidence or mortality. Projection models are essential tools used to estimate the future burden of cancer in a population. These models rely on current data on cancer incidence, mortality, and survival rates, and can provide valuable information for public health policies and resource allocation.

Despite their limitations, cancer projection models remain a crucial area of research and can contribute to the development of effective cancer prevention and treatment strategies. Cancer prediction models consider various factors that can influence cancer incidence and mortality, such as demographic changes, advancements in screening and diagnostic methods, and shifts in lifestyle and

environmental factors. There are several different types of cancer prediction models, including population-based, disease-specific, and individual-level models. Each model type has its own strengths and limitations, and the choice of model may vary depending on the specific research question being addressed.

2.2. Decomposition Analysis

2.2.1. Das Gupta's Method

Das Gupta's Standard Rate Decomposition is a widely used method in demography for analyzing differences in age-specific mortality rates or other demographic rates between two populations or time periods [35]. This technique enables researchers to decompose the differences in rates into three distinct components: the composition effect, the age-specific rate effect, and the interaction effect. The composition effect reflects the differences in age distribution between the two populations or time periods being compared, while the age-specific rate effect accounts for the differences in age-specific rates between the two groups. Lastly, the interaction effect represents the combined impact of both composition and age-specific rate differences. When the incidence mortality rate according to the population group is r , the factors can be expressed as follows.

$$r(x_1 \dots x_k) = \sum_{i=1}^k x_i$$

Utilizing superscript 'a' to denote the initial population and superscript 'b' for the subsequent population, the mathematical representation of the contributions made

by two factors, $F(x_1)$ and $F(x_2)$, to the disparity between a and b within the framework of these two factors can be articulated as follows.

$$\begin{cases} F(x_1) = \frac{1}{2} (x_2^a + x_2^b)(x_1^a - x_1^b) \\ F(x_2) = \frac{1}{2} (x_1^a + x_1^b)(x_2^a - x_2^b) \end{cases}$$

Intuitively, the contribution of a factor is derived from its conditional impact on the mean values of other factors. Consequently, the relative contribution of x_1 is given by $\frac{F(x_1)}{F(x_1) + F(x_2)}$, while the relative contribution of x_2 is represented by $\frac{F(x_2)}{F(x_1) + F(x_2)}$. This methodology is generally unambiguous when only a limited number of factors are involved. Nonetheless, as the value of k increases, the calculations become increasingly complex due to the necessity of computing all potential counterfactuals (2^k) and subsequently aggregating the results. Formally, the contribution of the i th factor to the rate can be expressed as follows.

$$F(x_i) = \sum_{j=1}^{k-1} \frac{R(j-1, i)}{k \binom{k-1}{j-1}} (x_i^a - x_i^b)$$

Wherein $R(j, i)$ represents the summation of all conceivable values of the product of $k - 1$ factors (excluding x_i), originating from j factors in population and the remaining factors in population b. The quantity of possible values can be substantial when multiple factors are involved, as the number of permutations increases more rapidly than k .

This decomposition technique considers all possible rate substitutions between populations, avoiding path dependency. Das Gupta's method weighs each path, with the importance of a factor decreasing as other factors change. The method intuitively presents results and applies to various research areas, such as sociodemographic disparities, income inequality, and health outcomes.

2.2.2. Bashir's Method

Bashir's decomposition is method for analyzing differences in incidence or mortality rates between two populations, considering the contributions of risk factors and demographic factors [36]. This method allows researchers to distinguish the contributions of risk factors and demographic factors to the observed differences in rates. Calculate age-specific rates for both populations: The first step is to compute the age-specific incidence or mortality rates for each population, typically in age groups (e.g., 5-year age intervals).

Let us consider two groups, Group 1, and Group 2, with cases or deaths C_1 and C_2 , and populations P_1 and P_2 , respectively. We aim to analyze the relative difference between C_1 and C_2 , which can be expressed as:

$$\frac{C_2}{C_1} - 1 = \frac{R_2}{R_1} \frac{P_2}{P_1} - 1 = \left(\frac{R_2}{R_1} - 1 \right) + \left(\frac{P_2}{P_1} - 1 \right) \times \frac{R_2}{R_1}$$

Note that the second term above meaning that the change in the population size

generates change cases which rely on the changes in the crude incidence/mortality rate. We further decompose this quantity into components due to differences in risk and population structure. Let us define the following:

- 1) λ_{ix} rate in age group x for group i
- 2) ω_{ix} proportion of population in age group x for group i

We aim to analyze the difference in crude rates between Group 1 and Group 2, using Group 1 as the baseline for comparison. The crude rates for Group 1 and Group 2 are represented by R_1 and R_2 respectively. Our objective is to examine the relative difference in crude rates,

$$\frac{R_2 - R_1}{R_1}$$

In terms of risk and population structure Using above condition 1 and 2, R_3 will be $\Sigma_x \lambda_{1x} \omega_{2x}$ that is the rate in group 1 applied to the population proportion in group 2.

$$\frac{\Sigma_x (\lambda_{2x} \omega_{2x} - \lambda_{1x} \omega_{1x})}{\Sigma_x \lambda_{1x} \omega_{1x}} = \frac{\Sigma_x \omega_{2x} (\lambda_{2x} - \lambda_{1x})}{\Sigma_x \lambda_{1x} \omega_{1x}} + \frac{\Sigma_x \lambda_{1x} (\omega_{2x} - \omega_{1x})}{\Sigma_x \lambda_{1x} \omega_{1x}}$$

The first component on the right-hand side of above equation represents the proportion of the difference in crude rates between Group 1 and Group 2

attributable to differences in population structure, while the second component represents the proportion attributable to differences in risk.

By applying this decomposition method, researchers can gain insights into the relative importance of demographic and risk factors in explaining differences in incidence or mortality rates between two populations. This can be helpful for public health professionals and policymakers when designing interventions or evaluating the effectiveness of existing policies. It's worth mentioning that while the paper's method is inspired by Das Gupta's Standard Rate Decomposition method, it specifically focuses on the contributions of risk factors and demographic factors to the differences in rates, making it a more specialized approach for addressing these specific factors in epidemiological studies.

2.2.3 Cheng's Method

Decomposition methods for absolute numbers aim to attribute differences or changes in total deaths to various components or factors, such as population size, age structure, and mortality rates [28]. The novel decomposition method adopted in this study has been reported to demonstrate robustness in relation to the choice of decomposition order for the three factors and the selection of reference group, as compared to the two most employed decomposition methods.

Table 1. Meaning of mathematical symbols in the decomposition formula

Age group	A (j=1)				B (j=2)			
	Case	Population	Rate	Age structure	Case	Population	Rate	Age structure
0-4	c_{11}	n_{11}	r_{11}	s_{11}	c_{12}	n_{12}	r_{12}	s_{12}
5-9	c_{21}	n_{21}	r_{21}	s_{21}	c_{22}	n_{22}	r_{22}	s_{22}
10-14	c_{31}	n_{31}	R_{31}	s_{31}	c_{32}	n_{32}	r_{32}	s_{32}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
85+	c_{181}	n_{181}	R_{181}	s_{181}	c_{182}	n_{182}	R_{182}	s_{182}
Total	C_1	N_1	R_1	$S_1=1$	C_2	N_2	R_2	$S_2=1$

Using M_p , M_a and M_r to represent the main effects of the changes in population size, in age structure and in rates, and I_{pa} , I_{pr} , I_{ac} and I_{par} to represent their two-way and three-way interactions, respectively.

Table 2. Main and Interaction effect formula of each year

	A Year	B Year
$M_p(M'_p)$	$\sum_{i=1}^{18} (N_2 - N_1) s_{i1} r_{i1}$	$\sum_{i=1}^p (N_1 - N_2) s_{i2} r_{i2}$
$M_a(M'_a)$	$\sum_{i=1}^{18} N_1 (s_{i2} - s_{i1}) r_{i1}$	$\sum_{i=1}^p N_2 (s_{i1} - s_{i2}) r_{i2}$
$M_r(M'_r)$	$\sum_{i=1}^{18} N_1 s_{i1} (m_{i2} - r_{i1})$	$\sum_{i=1}^p N_2 s_{i2} (r_{i1} - r_{i2})$
$I_{pa}(I'_p a)$	$\sum_{i=1}^{18} (N_2 - N_1) (s_{i2} - s_{i1}) r_{i1}$	$\sum_{i=1}^p (N_1 - N_2) (s_{i1} - s_{i2}) r_{i2}$
$I_{pr}(I'_p r)$	$\sum_{i=1}^{18} (N_2 - N_1) s_{i1} (r_{i2} - r_{i1})$	$\sum_{i=1}^p (N_1 - N_2) s_{i2} (r_{i1} - r_{i2})$
$I_{ar}(I'_p a)$	$\sum_{i=1}^{18} N_1 (s_{i2} - s_{i1}) (r_{i2} - r_{i1})$	$\sum_{i=1}^p N_2 (s_{i1} - s_{i2}) (r_{i1} - r_{i2})$
$I_{par}(I'_p a r)$	$\sum_{i=1}^{18} (N_2 - N_1) (s_{i2} - s_{i1}) (r_{i2} - r_{i1})$	$\sum_{i=1}^p (N_1 - N_2) (s_{i1} - s_{i2}) (r_{i1} - r_{i2})$

Let a%, b%, and c% represent the two-way interaction proportions between population size and age structure, population size and rate change, and age structure and ate change allocated to the first factor, respectively. Consequently, (100-a) %, (100-b) %, and (100-c) % of the three two-way interactions are

allocated to the second factor (Table2).

Let $d1\%$, $d2\%$, and $(100-d1-d2)\%$ represent the proportions of the three-way interaction allocated to population size, age structure, and rate change, respectively. Utilizing A (A'), R (R'), and P (P') to denote the number of deaths attributed to age structure, rate change, and population size as defined by the method when using year, A and B , the contributions of the three factors can be calculated as follows:

Table 3. Decomposition formula about main effect

Factor		Formula
A Year	P	$M_p + a\%I_{pa} + b\%I_{pr} + d_1\%I_{par}$
	A	$M_a + (100 - a)\%I_{pa} + c\%I_{ar} + d_2\%I_{par}$
	R	$M_r + (100 - b)\%I_{pr} + (100 - c)\%I_{ar} + (100 - d_1 - d_2)\%I_{par}$
B Year	P'	$M'_p + a\%I'_{pa} + b\%I'_{pr} + d_1\%I'_{par}$
	A'	$M'_a + (100 - a)\%I'_{pa} + c\%I'_{ar} + d_2\%I'_{par}$
	R'	$M'_r + (100 - b)\%I'_{pr} + (100 - c)\%I'_{ar} + (100 - d_1 - d_2)\%I'_{par}$

The results of decomposition should remain invariant in absolute value when the reference population changes. Thus, a set of three equations can be constructed as follows:

$$\begin{cases} P \equiv -P' \\ A \equiv -A' \\ M \equiv -M' \end{cases}$$

Based on the above assumption, it is possible to construct a single expression for each effect for each factor.

$$\begin{aligned}\sum_{i=1}^p (N_2 - N_1) [(s_{i1}r_{i1} - s_{i2}r_{i2})(100 - a - b)\% + (s_{i2}r_{i1} - s_{i1}r_{i2})(a - b)\%] &\equiv 0 \\ \sum_{i=1}^p (s_{i2} - s_{i1}) [(N_2m_{i1} - N_1r_{i2})(100 - a - c)\% + (N_1r_{i1} - N_2r_{i2})(a - c)\%] &\equiv 0 \\ \sum_{i=1}^p (r_{i2} - r_{i1}) [(N_2s_{i2} - N_1s_{i1})(100 - b - c)\% + (N_1s_{i2} - N_2s_{i1})(b - c)\%] &\equiv 0\end{aligned}$$

It is evident that the three equations presented cannot be true under all circumstances unless the variables a, b, and c are all equal to 50. However, it should be noted that there are no specific requirements for the values of the variables d1 and d2 in relation to these equations.

Furthermore, it is important to consider that there is no established theoretical guidance on how to allocate the three-way interaction of three factors. As a result, it is recommended to divide the interaction equally among the three factors to ensure a fair and balanced allocation ($d1=d2=\frac{1}{3} \times 100$.)

$$A = M_a + \frac{1}{2}I_{ar} + \frac{1}{2}I_{pa} + \frac{1}{3}I_{par}$$

$$P = M_p + \frac{1}{2}I_{pr} + \frac{1}{2}I_{pa} + \frac{1}{3}I_{par}$$

$$R = M_r + \frac{1}{2}I_{pr} + \frac{1}{2}I_{ar} + \frac{1}{3}I_{par}$$

2.3. Population structure-based projection models

2.3.1. APC (Age–period–cohort) model

The APC model is a statistical model that distinguishes the effects of age, period, and cohort on cancer incidence and mortality. The model assumes that cancer incidence and mortality are determined by the interaction of these three factors and that the relative contribution can be estimated through statistical analysis.

This model has been used to predict cancer incidence and mortality in various populations and periods and to investigate underlying trends and patterns [37,38].

The APC model is a linear regression model that separates the effects of age, period, and cohort on cancer incidence and mortality. A general formula for the APC model can be expressed as follows:

$$Y_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k + \varepsilon_{ijk}$$

A regression model with an outcome variable Y_{ijk} to be explained, which is modeled as a function of several predictor variables. Specifically, the model includes an intercept term μ , as well as separate age effects α_i , period effects π_j , and cohort effects γ_k . The error term ε_{ijk} represents the variation in the outcome variable that is not explained by the predictor variables and is assumed to follow a normal distribution with mean zero and variance σ^2 .

To prevent overparameterization in our analysis, we have implemented the commonly used "usual constraints" whereby the parameters are required to sum to zero. By doing so, we can ensure that our model remains parsimonious and avoids overfitting the data.

Table 4. APC Model on an Age-Period Array

Age Groups	Periods Groups		
	j=1	...	j=k
i=1	$\mu + \alpha_1 + \pi_1 + \gamma_3 + \varepsilon_{1,1,3}$		$\mu + \alpha_1 + \pi_k + \gamma_{3+(k-1)} + \varepsilon_{1,1,3+(k-1)}$
i=2	$\mu + \alpha_2 + \pi_1 + \gamma_2 + \varepsilon_{2,1,3}$		$\mu + \alpha_2 + \pi_k + \gamma_{2+(k-1)} + \varepsilon_{2,1,3+(k-1)}$
i=3=I	$\mu + \alpha_3 + \pi_1 + \gamma_1 + \varepsilon_{3,1,1}$		$\mu + \alpha_3 + \pi_k + \gamma_{1+(k-1)} + \varepsilon_{3,1,1+(k-1)}$

To facilitate the presentation of the APC model, it is convenient to use matrix notation. By doing so, we can represent the model in a more compact form and simplify the calculations involved. the equation as follows:

$$y = X\beta + \varepsilon$$

In the frequentist statistical tradition, estimates of the model parameters can be obtained by maximizing the likelihood function. Specifically, for Equation 8, the likelihood function represents the probability of the observed data given a set of parameters β , a scalar σ^2 , and the input variables X (Wang, Yue, and Faraway 2018: 39-40).

$$p(y|\beta, \sigma^2, X) = L(\beta, \sigma^2 | y, X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{(I \times J)} \times e^{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)}$$

In some cases, the maximum likelihood estimates (MLE) of the model parameters may coincide with those obtained through ordinary least squares (OLS). Specifically, for above Equation, the MLE and OLS estimates of the parameter vector β can be shown to be equal, such that $\hat{\beta}_{MLE} = \hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$, where the superscript -1 denotes a regular inverse.

One of the advantages of the APC model is its ability to isolate the effects of age, period, and cohort, which can provide insight into the underlying mechanisms driving cancer incidence and mortality trends. APC models can also be used to predict cancer incidence and mortality in various scenarios, which can inform cancer control and prevention policies. However, APC models have several limitations, such as assumptions of linearity that may not be maintained in some cases and the possibility of overfitting when dealing with small sample sizes [39,40,41].

2.3.2 NORDPRED (NP) model

The NP approach has been shown to provide more accurate projections for long-term trends in cancer incidence rates compared to standard APC models using log-link functions [14]. The power-link function used in NP helps to lower the exponential growth in projected rates over time, making the projections more realistic.

The power-link function approximates the log-link function based on Box-Cox power transformation theory, in which $\lambda \rightarrow 0$ $\lim x^\lambda = \log(x)$. The NP model is defined as

$$case_{ap} \sim Poisson(\mu_{ap})$$

$$R_{ap} (or \frac{\mu_{ap}}{n_{ap}}) = (A_a + D \cdot P + P_p + C_c)^5$$

where R_{ap} is the incidence rate for age group α in calendar period p , μ_{ap} is the mean count of cases in age group α and calendar period p , and n_{ap} is the corresponding population size. The non-linear components of age, period, and cohort are denoted as A_a , P_p , and C_c , respectively, while D represents the common linear drift parameter of period and cohort. The cohort component is calculated as $c = A + p - a$, where A is the number of age groups. To ensure an adequate fit of the model to the data, the NP software (R package) uses a goodness-of-fit test to determine the number of 5-year periods on which the projections should be based [42]. The software also provides options for projecting future periods, including the choice between using the average trend or the slope for the last 10 years of observed values as the drift component, and the use of a "cut trend" option to attenuate the impact of current trends in future periods.

Overall, the NP approach provides a flexible and powerful method for projecting cancer incidence rates and has been shown to produce more realistic long-term projections compared to standard APC models. It is widely used in cancer research and can help inform public health policies and resource allocation [14].

2.3.3 BAPC (Bayesian Age–period–cohort) Model

The Bayesian Age-Period-Cohort (BAPC) model is a robust extension of classical APC models that enables uncertainty quantification and Bayesian inference. In cancer dynamics, BAPC models have been used to predict cancer incidence and mortality and to investigate underlying trends and patterns. This statement model is used to project future cancer incidence rates by extrapolating the estimated effects into the future.

In the classical APC Poisson model, Bray incorporated a second-order autoregressive prior model to smooth the age, period, and cohort effects and to extrapolate the period and cohort effects into the future [43,44]. The model can be expressed as follows:

$$case_{ap} \sim Poisson(\mu_{ap})$$

$$\log\left(\frac{\mu_{ap}}{n_{ap}}\right) = A_a + P_p + C_c$$

Above formula refers to the prior distributions used in the NP model for the age effects. It notes that if we compute N-period projections based on P-period observed data, we will have a total of $C = A + P - 1$ cohorts, where A is the number of age groups and P is the number of observed periods. In the NP model, an individual cohort c is calculated as $c = A + p - a$. The statement then goes on to specify the prior distributions for the A age effects.

$$A_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_A})$$

$$A_2 | A_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_A})$$

$$A_a | A_{1, \dots, a-1} \sim \text{normal}(2A_{a-1} - A_{a-2}, \frac{1}{\tau_A})$$

$$3 \leq a \leq A$$

For the P + N period effects:

$$P_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_P})$$

$$P_2 | P_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_P})$$

$$P_a | P_{1, \dots, a-1} \sim \text{normal}(2P_{a-1} - P_{a-2}, \frac{1}{\tau_P})$$

$$3 \leq p \leq P + N$$

For the C + N cohort effects:

$$C_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_C})$$

$$C_2 | C_1 \sim \text{normal}(0, 10000 \frac{1}{\tau_C})$$

$$C_a | C_{1, \dots, a-1} \sim \text{normal}(2C_{a-1} - C_{a-2}, \frac{1}{\tau_C})$$

$$3 \leq c \leq C + N$$

The variance parameters τ_A , τ_P , and τ_C , which determine the smoothness of age, period, and cohort effects respectively, are given the same log-gamma prior (Initial values of gamma distribution may vary).

$$\tau \sim \text{loggamma}(1, 0.005)$$

Estimated rates are derived by combining the simulated age, period and cohort effects based on

$$R_{ap} = \exp(A_a + P_p + C_c)$$

Bayesian frameworks can also allow model selection and comparison and incorporate external data sources such as prior knowledge of risk factors or genetic markers. However, BAPC models can be computationally intensive and require advanced statistical software and computational resources and selection of pre-distribution and hyperparameters can also affect the model results and requires careful consideration [45,46,47,48].

2.4. Non-population structure-based projection models

2.4.1. JOINPOINT (JP) Model

JP regression is a statistical method used to analyze trends in time series data that may have different rates of change over time, such as cancer incidence or mortality rates. The JP model fits a series of straight lines to the data, allowing for changes in the slope of the line at one or more points in time. The number and location of JP are determined by the data and the model selection criteria [15].

In cancer projection, JP regression can be used to identify changes in the trend of cancer incidence or mortality rates over time and to estimate the annual percent change (APC) in rates within each segment of the trend. The APC is an estimate of the average rate of change in the cancer rate over a specific time period, such as the past decade or the projected future.

This model allows us to detect significant changes in the rate of increase or decrease in certain outcome variables over time, which can provide valuable insights into trends and patterns in population-level data. The general formula for the JP model can be expressed as follows:

$$\log(y_i) = \begin{cases} \beta_0 + \beta_1 x_i + \epsilon_i, & x_i < \tau_1 \\ \beta_0 + \beta_1 x_i + \delta_1(x_i - \tau_1) + \epsilon_i, & \tau_1 \leq x_i < \tau_2 \\ \vdots & \vdots \\ \beta_0 + \beta_1 x_i + \delta_{k-1}(x_i - \tau_{k-1}) + \epsilon_i, & \tau_{k-1} \leq x_i < \tau_k \\ \beta_0 + \beta_1 x_i + \delta_k(x_i - \tau_k) + \epsilon_i, & \tau_k < x_i \end{cases}$$

where where rate is the cancer incidence or mortality rate, x is the year of observation $\beta_0, \beta_1, \dots, \beta_k$ are regression are regression coefficients (with $\delta_1, \dots, \delta_k$ being slope differences given time period τ) and $\epsilon_i \sim N(0, \sigma^2)$. The number and location of the JOINPOINTS, x_1, \dots, x_i , are estimated using a grid search algorithm, and the slopes $\sigma_1, \dots, \sigma_k$ are estimated using a maximum likelihood method. The JP model assumes linearity, and error ϵ_i and independent and normally distributed.

The model is fitted using a segment regression approach, and the number of coupling points and the location of coupling points are determined through a goodness-of-fit test. Several methods exist in determining these parts, and the permutation test method and the mBIC method are mainly used. JP models are often used in cancer projection studies to estimate future incidence or mortality based on trends observed in the past, and this model can also be used to identify significant periods of change in cancer incidence that can provide insight into the effects of changes in public health interventions or risk factors [49,50,51,52].

2.4.2. Time Series Forecasting Model

A time series model is a type of statistical model used to analyze time data collected at regular intervals. In cancer projection studies, time series models can be used to predict future cancer incidence or mortality based on historical trends. When using time series models for predictive purposes, various methods are used, starting with several derivatives of ARIMA (Autoregressive Integrated Moving Average), such as exponential smoothing, state space, and GPR (Gaussian Process Regression) ... etc. The models used in this study were largely divided into three types, and we would like to introduce ARIMA, VAR (Vector Autoregression), and Space-state methods [53,54,55,56,57,58].

2.4.2.1 ARIMA (Autoregressive Integrated Moving Average) Model

The autoregressive integrated moving average (ARIMA) model is a widely used

time series prediction model that captures dependencies between observations in a time series. It is a generalization of simple autoregressive (AR) and moving average (MA) models. The ARIMA model assumes that the future value of the time series depends linearly on the historical value and the error (or residual) of the model [59]. Specifically, it consists of three components:

Automatic Regression (AR) component: This component models the dependency between a linear combination of observations and historical values. The order of AR components represented by p specifies the number of historical observations to be included in the linear combination.

Integration (I) component: This component models the anomalies of the time series by differentiating the time series until it is stationary. The order of differences expressed by d specifies the number of times the column is differentiated.

Moving Average (MA) component: This component models the dependency between observations and a linear combination of historical errors. The order of MA components represented by q specifies the number of historical errors to be included in the linear combination.

ARIMA models are represented by ARIMA (p, d, q), where p, d, and q are the order of AR, I, and MA components, respectively. Models are generally suitable for time series by estimating parameters through maximum likelihood estimation or Bayesian inference. The ARIMA (p,d,q) model can be defined as follows:

$$y(t) = c + \varphi_1 y(t-1) + \dots + \varphi_p y(t-p) + \varepsilon(t) + \theta_1 \varepsilon(t-1) + \dots + \theta_q \varepsilon(t-q)$$

where $y(t)$ is the incidence or mortality rate at time t , c is a constant term, $\varphi_1, \dots, \varphi_p$ are the autoregressive coefficients of order p , $\varepsilon(t)$ is the white noise error term at time t , and $\theta_1, \dots, \theta_q$ are the moving average coefficients of order q . The model also requires the specification of the order of differencing d , which determines the number of times the time series needs to be differenced to achieve stationarity. The notation for an ARIMA(p,d,q) model after d differencing is ARMA(p,q). The model parameters $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$ and the constant term c are estimated using maximum likelihood estimation, which involves finding the values of the parameters that maximize the likelihood of observing the data given the model.

2.4.2.2 Space-State Model

A state space model is a class of statistical models used to analyze time series data representing trends and/or seasonal patterns. In cancer projection studies, state space models can be used to predict future cancer incidence or mortality based on

historical trends and to estimate and adjust unobserved factors that may affect outcomes of interest. Overall, state space models provide a flexible framework for analyzing and predicting time data, including cancer incidence or mortality. In general, these models are mainly predicted through the construction of linear state space models [60].

2.4.2.2.1 Linear Space State Model (SSML)

Linear Space State Models (SSML) are a popular class of models for cancer projection. They are a type of time series model that can be used to forecast future cancer incidence or mortality rates [61,62]. In SSML, the state variable is assumed to evolve linearly over time, and it is related to the observed data through a linear measurement equation. SSML can be fitted to historical cancer data to estimate the underlying trend and seasonality, and to make predictions for future time periods.

Measurement equation: $Y_t = F_t \theta_t + v_t, v_t \sim N(0, R_t)$

State equation: $\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, Q_t)$

where Y_t is the unobserved incidence or mortality rate at time t , F_t is the design matrix including age and period effects, θ_t is the state vector including age, period, and cohort effects for unobserved time trend, and v_t is the un-observation error at time t with variance R_t , θ_{t-1} is the vector at observation time point, G_t is the state transition matrix from time t to $t-1$ and w_t is the process error at time t with

variance Q_t .

Both R_t and Q_t are typically assumed to follow a normal distribution with mean μ_0 and variance σ^2 . The linear space state model and the quadratic space state model share the same expression, and only differences in parameters exist. The state vector θ_t includes the linear coefficients, where α_t is the intercept, β_t is the linear coefficient and the transition matrix G_t can be written as:

$$\text{linear } \theta_t = (\alpha_t, \beta_t), \quad \text{linear } G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

It should be noted that linear spatial state models have limitations, especially when dealing with complex nonlinear relationships. When the incidence or mortality rate changes over time with a linear pattern, it is appropriate for the use of the model, and irregular patterns can make it difficult to fit the model, so it is preferred to check the current trend of cancer observations.

2.5. Summary

In conclusion, this study emphasizes the necessity of constructing a projection model through an empirical method for predicting cancer occurrence and establishing appropriate health policies in Korea. The review of literature on previous models suggests that time series analysis and APC-based models are useful tools for predicting cancer incidence and mortality, but both have limitations and challenges. Moreover, the research on comparing different models based on the type and age of cancer is insufficient, and there may be different distributions

for the same cancer. Furthermore, available data may have demographic or sociological characteristics that affect the accuracy of prediction. Thus, empirical methods are required to select the appropriate model for each cancer type based on the available data, and it is necessary to compare multiple scenarios and models to make reasonable predictions.

The literature also reveals various methods for selecting decomposition analysis methods. Three previous decomposition analysis methods measure the degree of risk based on aging, population structure change, age structure change, and three factors based on population structure base. Cheng's method combines the other two methods' advantages and produces robust results with a small calculation formula. Therefore, in this study, we will conduct decomposition analysis based on Cheng's method. In the earlier discussion, decomposition analysis was based on demographic changes, which made it possible to perform the analysis only if the estimates were divided by age group. However, the time series and JP models focus on estimating the ASR from all cases or deaths without considering age groups in the prediction. Therefore, it is difficult to perform a decomposition analysis for these models, so only model comparisons of the predicted values are performed.

CHAPTER III. SELECTION AGEING RELATED CANCER AND CURRENT TREND ANALYSIS IN 2000 TO 2019

3.1. Introduction

This chapter outlines the process of selecting cancers related to aging and confirms their validity. The initial screening involves the application of a previous framework, and subsequently, the impact of aging on cancer incidence and death is quantified using decomposition analysis. Additionally, APC-analysis is employed to verify each age, period, and cohort effect.

3.2. Selection of ageing-related cancer

To select for age-associated cancers for a given cancer type, we decided to do so under four conditions based on the framework mentioned in Chapter 1. First, for the selection, we looked at the incidence and mortality rates of each cancer type in 2019, the last year for which data is available, for all age groups and for those aged 65 and older and selected only those cancer types that ranked in the top 10 in both conditions [33]. The reason for considering all age groups rather than just those aged 65 and older when investigating age-related cancers is that there are currently more people under 65 ages than over 65 ages, so differences in absolute numbers of cancer cases or deaths cannot be overlooked. In addition, the population under the age of 65 is a population that may be at latent risk of developing or dying from cancer if a long-term trend develops [63,64,65].

3.2.1. Selection of ageing-related cancers with incidence

To select age-related cancers, it is necessary to first establish inclusion criteria by isolating specific cancer types. This can be achieved by comparing the incidence rates across all age groups and specifically in the over age 65 age group, using the most recent data available, which in this case is from 2019.

Table 5-1 and Table 5-2 presents the incidence rates of each cancer type by gender in both the all-age and over 65 age groups. For both men and women, several types of cancer were found to have high incidence rates, including stomach, colorectal, liver, biliary tract, pancreas, and lung cancers. Additionally, among men, urinary system-related cancers, including prostate cancer, were highly ranked. Among women, female-specific cancers such as breast, cervix, uterus, and ovary cancers were also highly ranked. Notably, only breast and ovarian cancers had higher incidence rates than other female cancers in the over age 65 age group. Additionally, thyroid cancer had higher incidence rates among women.

Table 5-3 presents the ratio of the incidence rate for individuals aged 65 and older to the incidence rate for each age group (00-34, 35-49, and 50-64), stratified by sex and cancer type. However, the age group of 50-64 may share similar biological risk factors and external factors associated with aging with the over age 65 group, unlike younger age groups. To address this issue, we adjusted the ratio by a factor of 1 or more for the 50-64 age group and by a factor of 2 or more for the other age groups, to identify cancers that may be linked to aging. Our findings

indicate that most cancers were more dominated in individuals over age 65, except for female-specific cancers and thyroid cancer, which exhibited a high dominant among women. Among men, testis cancer was excluded from the analysis, while Hodgkin's lymphoma was excluded for both sexes.

In our analysis, we also considered age-specific trends that vary over time. If the incidence rate of a cancer showed an increasing trend in the under-65 age group, we considered the cancer to not be associated with aging. As of 2019, we examined cancers where the annual average percent change in cancer incidence by age group over the past decade (2010-2019) indicated a decreasing trend in the over age 65 age group and an increasing trend in the remaining age groups.

Among the cancers analyzed in Table 5-4, only LOCP and testis cancers occurred in males, while the remaining cancers exhibited the same decreasing or increasing trend. Consequently, we excluded lip and mouth and testis cancers from the table and included them in the criteria to finally identify cancers that are associated with aging.

Table 5-1. Cancer incidence and ranking in all age by sex in 2019

Cancer Sites	ICD-10	Male		Female	
		Incidence Rate	Rank	Incidence Rate	Rank
LOCP	C00-C14	11.2	12	4.3	15
Esophagus	C15	10	13	1.1	19
Stomach	C16	77.1	2	37.4	4
Colorectal	C18-C20	67.3	3	46	3
Liver	C22	45.1	5	15.7	6
Gallbladder	C23-C24	15.2	10	13.3	8
Pancreas	C25	16.2	7	15.1	7
Larynx	C32	4.5	15	0.3	21
Lung	C33-C34	79.5	1	37	5
Breast	C50	-	-	94.9	1
Cervix uteri	C53	-	-	12.6	9
Corpus uteri	C54	-	-	12.6	9
Ovary	C56	-	-	11.2	11
Prostate	C61	65.8	4	-	-
Testis	C62	1.3	18	-	-
Kidney	C64	16.2	7	7.2	13
Bladder	C67	15.6	9	3.5	16
CNS	C70-C72	4.2	16	3.5	16
Thyroid	C73	29.5	6	89.2	2
Hodgkin lymphoma	C81	0.7	19	0.5	20
Non-Hodgkin lymphoma	C82-C86, C96	12.1	11	8.7	12
Multiple myeloma	C90	3.9	17	3.2	18
Leukemia	C91-C95	8.2	14	6.2	14

LOCP: Lip, oral cavity, and pharynx; CNS: Center Nervous System

Table 5-2. Cancer incidence and ranking in 65+ age by sex in 2019

Cancer Sites	ICD-10	Male		Female	
		Incidence Rate	Rank	Incidence Rate	Rank
LOCP	C00-C14	34.5	12	9.3	17
Esophagus	C15	44.8	9	3.1	19
Stomach	C16	294.2	3	108.6	3
Colorectal	C18-C20	253.5	4	143.9	1
Liver	C22	155.7	5	55.5	7
Gallbladder	C23-C24	81.8	6	56.7	6
Pancreas	C25	72.5	8	57.0	5
Larynx	C32	19.2	15	0.8	20
Lung	C33-C34	424.7	1	124.7	2
Breast	C50	-	-	89.2	4
Cervix uteri	C53	-	-	16.1	12
Corpus uteri	C54	-	-	12.7	15
Ovary	C56	-	-	16.6	10
Prostate	C61	373.7	2	-	-
Testis	C62	0.3	19	-	-
Kidney	C64	44.3	10	16.2	11
Bladder	C67	81.2	7	14.4	13
CNS	C70-C72	11.4	17	7.4	18
Thyroid	C73	22.9	13	49.3	8
Hodgkin lymphoma	C81	1.3	18	0.7	21
Non-Hodgkin lymphoma	C82-C86, C96	40.9	11	23.1	9
Multiple myeloma	C90	17.7	16	11.0	16
Leukemia	C91-C95	22.3	14	12.8	14

LOCP: Lip, oral cavity, and pharynx; CNS: Center Nervous System

Table 5-3. Cancer incidence rate ratio in age groups by sex

Cancer Sites	ICD-10	Comparison1		Comparison2		Comparison3		Selection	
		Male	Female	Male	Female	Male	Female	Male	Female
LOC	C00-C14	24.2	7.3	7.0	3.6	1.7	1.5	Yes	Yes
Esophagus	C15	2239.5	35.7	49.1	7.0	2.8	1.7	Yes	Yes
Stomach	C16	110.4	26.4	12.9	6.4	2.3	2.2	Yes	Yes
Colorectal	C18-C20	48.2	33.6	12.1	8.3	2.3	2.5	Yes	Yes
Liver	C22	79.7	59.5	10.1	16.0	1.9	3.1	Yes	Yes
Gallbladder	C23-C24	430.5	261.4	49.7	37.7	5.3	5.8	Yes	Yes
Pancreas	C25	103.5	54.8	21.0	23.9	3.3	3.7	Yes	Yes
Larynx	C32	480.0	19.4	48.0	25.2	2.5	2.3	Yes	Yes
Lung	C33-C34	339.2	82.1	53.2	14.7	5.1	2.6	Yes	Yes
Breast	C50	-	3.6	-	0.7	-	0.5	-	NO
Cervix uteri	C53	-	2.5	-	1.2	-	0.9	-	NO
Corpus uteri	C54	-	3.3	-	1.2	-	0.5	-	NO
Ovary	C56	-	4.3	-	1.8	-	0.9	-	NO
Prostate	C61	3114.4	-	238.5	-	6.0	-	Yes	Yes
Testis	C62	0.2	-	0.5	-	0.7	-	NO	-
Kidney	C64	13.9	11.6	4.2	4.2	1.6	1.4	Yes	-
Bladder	C67	269.7	221.4	39.1	26.4	4.8	5.1	Yes	Yes
CNS	C70-C72	5.1	3.9	5.0	3.7	2.2	1.7	Yes	Yes
Thyroid	C73	1.0	0.7	0.7	0.5	0.6	0.4	NO	NO
Hodgkin lymphoma	C81	1.9	1.0	3.7	4.3	1.6	2.9	NO	NO
Non-Hodgkin lymphoma	C82-C86, C96	12.0	9.6	7.2	5.8	2.5	2.0	Yes	Yes
Multiple myeloma	C90	160.8	169.6	17.0	20.2	3.5	2.6	Yes	Yes
Leukemia	C91-C95	4.6	3.7	5.0	3.7	2.4	1.6	Yes	Yes

LOC: Lip, oral cavity, and pharynx; CNS: Center Nervous System

†If the estimated mortality case is too small due to the low number of deaths in the comparative group, it is marked as '—'

Table 5-4. JOINPOINT Average Annual Percent Change (AAPC) shifting last 10 years period in incidence by sex

Cancer Sites	ICD-10	Male			Female			
		00-34	35-49	50-64	65+	00-34	35-49	50-64
LOC	C00-C14	3.35	1.16	0.13	-0.51	3.2	2.9	1.8
Esophagus†	C15	-	-2.69	-1.58	-2.83	-9.0	5.3	3.7
Stomach†	C16	-9.19	-6.45	-4.84	-4.72	-	-4.6	-3.6
Colorectal	C18-C20	4.11	-2.10	-4.60	-3.89	0.3	-0.2	-4.1
Liver	C22	-7.99	-6.76	-4.89	-2.46	-2.5	-3.3	-5.2
Gallbladder	C23-C24	-2.59	-2.45	-1.59	1.03	-3.6	-2.1	-2.2
Pancreas	C25	11.38	2.34	0.27	1.06	16.8	3.5	2.4
Larynx†	C32	-	-5.31	-3.78	-3.44	-	-4.3	-2.2
Lung	C33-C34	-0.04	-1.74	-2.30	-0.60	1.5	1.2	2.5
Breast	C50	-	-	-	-	1.9	3.7	3.2
Cervix uteri	C53	-	-	-	-	-2.5	-2.6	-4.6
Corpus uteri	C54	-	-	-	-	7.7	6.1	3.5
Ovary	C56	-	-	-	-	1.7	2.7	2.3
Prostate	C61	-0.13	1.79	2.92	3.44	-	-	-
Testis	C62	5.90	6.21	4.07	-0.11	-	-	-
Kidney	C64	5.95	5.34	1.56	0.99	4.0	4.8	3.6
Bladder	C67	-3.67	-2.12	-2.14	-0.18	-1.4	-1.5	0.7
CNS	C70-C72	0.16	-0.20	-0.21	0.24	0.4	-0.1	-0.4
Thyroid	C73	4.40	0.73	-3.68	-3.99	-0.4	-4.6	-8.3
Hodgkin lymphoma	C81	5.28	1.98	0.24	1.37	5.6	5.5	0.6
Non-Hodgkin lymphoma	C82-C86, C96	3.00	3.65	0.97	2.34	3.2	2.6	1.6
Multiple myeloma	C90	-2.24	0.54	1.33	3.48	-0.7	1.4	1.5
Leukemia	C91-C95	1.02	1.05	0.72	1.16	0.6	0.3	0.9

LOC: Lip, oral cavity, and pharynx; CNS: Center Nervous System

†If the estimated mortality case is too small due to the low number of deaths in the comparative group, it is marked as '—'

3.2.2 Selection of ageing-related cancers with mortality

In our analysis of mortality rates, Table 6-1 and Table 6-2 displays the mortality rates and rankings for both sexes and cancer types across all age groups and the 65+ age group. Across all age groups, selected cancer about the mortality rates were consistent with the selected cancer about incidence rates, with stomach, colorectal and anus, liver, and pancreas cancers exhibiting high mortality rates.

Additionally, gender-specific cancers such as prostate, breast, cervical, and ovarian also had high mortality rates. For some cancers, such as esophagus cancer, non-Hodgkin's lymphoma, and leukemia, there were different rankings for men and women. The over age 65 group exhibited the same selection of cancers as all age groups, except for breast cancer.

Table 6-3 presents the rate ratios of mortality compared to the mortality rate at over age 65 by age group. Except for breast cancer, all cancers demonstrated a dominance of deaths at over age 65.

Table 6-4 shows the 10-year trend for mortality rates in AAPC. Unlike the incidence rates, only one cancer showed a shifting trend in females, which was identified as Other cervical cancers. No other cancers exhibited a shifting trend.

Table 6-5 displays the results for the selected cancers when the previous conditions were combined. For mortality, breast cancer in women was excluded

because it did not meet the criteria, like the results for incidence. Consequently, esophagus, stomach, colorectal and anus, liver, pancreas, prostate, non-Hodgkin's lymphoma, and leukemia were identified as age-related cancers in men, while stomach, colorectal, liver, pancreas, lung, cervical, non-Hodgkin's lymphoma, and leukemia were identified as ageing-related cancers in women.

Table 6-1. Cancer mortality rate ratio in age groups by sex

Cancer Sites	ICD-10	Male		Female	
		Mortality Rate	Rank	Mortality Rate	Rank
LOCP	C00-C14	3.9	12	1.1	17
Esophagus	C15	5.5	8	0.5	18
Stomach	C16	19.2	4	10.2	5
Colorectal & Anus	C18-C21	19.6	3	14.9	2
Liver	C22	30.2	2	10.7	4
Gallbladder	C23-C24	10.0	6	9.2	7
Pancreas	C25	13.3	5	11.3	3
Larynx	C32	1.1	16	0.1	20
Lung	C33-C34	53.1	1	18.6	1
Malignant melanoma of skin	C43	0.5	17	0.5	18
Breast	C50	-	-	10	6
Cervix uteri	C53	-	-	3.4	9
Other of uterus	C54-C55	-	-	1.7	13
Ovary	C56	-	-	4.7	8
Prostate	C61	7.9	7	-	-
Kidney	C64	2.6	14	1.2	16
Bladder	C67	4.5	9	1.5	15
CNS	C70-C72	2.9	13	2.5	12
Non-Hodgkin lymphoma	C82-C86	4.5	9	3.2	10
Multiple myeloma	C90	2	15	1.7	13
Leukemia	C91-C95	4.4	11	2.9	11

LOCP: Lip, oral cavity, and pharynx; CNS: Center Nervous System

Table 6-2. Cancer mortality and ranking in 65+ age by sex

Cancer Sites	ICD-10	Male		Female	
		Mortality Rate	Rank	Mortality Rate	Rank
LOCP	C00-C14	16.1	12	4.2	17
Esophagus	C15	27.4	9	1.6	19
Stomach	C16	94.2	4	38.3	6
Colorectal & Anus	C18-C21	101.1	3	62.8	2
Liver	C22	122.4	2	45.0	4
Gallbladder	C23-C24	81.8	6	58.2	5
Pancreas	C25	66.4	5	48.1	3
Larynx	C32	6.5	16	0.5	20
Lung	C33-C34	316.8	1	79.4	1
Malignant melanoma of skin	C43	2.3	17	1.9	18
Breast	C50	-	-	17.9	7
Cervix uteri	C53	-	-	9.0	11
Other of uterus	C54-C55	-	-	4.5	16
Ovary	C56	-	-	12.5	9
Prostate	C61	54.4	7	-	-
Kidney	C64	44.3	13	12.6	15
Bladder	C67	28.9	8	7.2	14
CNS	C70-C72	10.5	15	7.3	12
Non-Hodgkin lymphoma	C82-C86	22.9	10	12.8	8
Multiple myeloma	C90	11.7	14	7.3	13
Leukemia	C91-C95	19.2	11	9.4	10

LOCP: Lip, oral cavity, and pharynx; CNS: Center Nervous System

Table 6-3. Cancer mortality rate ratio in age groups by sex

Cancer Sites	ICD-10	Comparison1		Comparison2		Comparison3		Selection	
		Male	Female	Male	Female	Male	Female	Male	Female
LOC	C00-C14	134.3	97.7	16.1	8.5	2.7	4.2	Yes	Yes
Esophagus†	C15	-	54.7	7.7	3.8	2.8	Yes	Yes	
Stomach	C16	392.4	153.9	22.4	9.3	4.2	4.6	Yes	Yes
Colorectal & Anus	C18-C21	359.7	341.1	28.1	20.1	4.8	6.3	Yes	Yes
Liver	C22	582.9	244.8	13.6	31.3	2.6	5.7	Yes	Yes
Gallbladder†	C23-C24	-	83.1	66.2	7.2	9.4	Yes	Yes	
Pancreas	C25	1328.3	2187.6	31.6	39.8	4.2	5.5	Yes	Yes
Larynx†	C32	-	-	-	-	6.7	16.6	Yes	Yes
Lung	C33-C34	3519.5	913.2	85.6	33.6	8.2	5.9	Yes	Yes
Malignant melanoma of skin	C43	231.1	44.4	23.1	19.3	3.3	4.0	Yes	Yes
Breast	C50	-	43.6	-	2.0	-	0.95	-	NO
Cervix uteri	C53	-	34.6	-	3.2	-	2.1	-	Yes
Other of uterus	C54-C55	-	82.7	-	8.4	-	1.5	-	Yes
Ovary	C56	-	72.1	-	5.2	-	1.7	-	Yes
Prostate	C61	2718.8	-	271.9	-	25.6	-	Yes	Yes
Kidney	C64	421.3	468.8	21.1	31.1	4.2	12.0	Yes	Yes
Bladder	C67	1445.0	651.4	144.5	143.3	12.4	16.7	Yes	-
CNS	C70-C72	13.5	13.2	7.0	5.2	3.0	2.7	Yes	-
Non-Hodgkin lymphoma	C82-C86	78.6	91.0	25.4	17.2	5.0	5.0	Yes	Yes
Multiple myeloma†	C90	1172.5	-	117.3	62.8	6.7	6.5	Yes	Yes
Leukemia	C91-C95	17.6	18.9	10.7	8.2	4.7	3.1	Yes	Yes

LOC: Lip, oral cavity, and pharynx; CNS: Center Nervous System

†If the estimated mortality case is too small due to the low number of deaths in the comparative group, it is marked as '—'

Table 6-4. JOINPOINT Average Annual Percent Change (AAPC) shifting last 10 years period in mortality by sex

Cancer Sites	Male						Female		
	ICD-10	00-34	35-49	50-64	65+	00-34	35-49	50-64	65+
LOC	C00-C14	-0.42	-2.66	-1.62	-2.59	-2.21	0.05	-2.11	0.13
Esophagus†	C15	-	-4.09	-2.65	-4.45	-	6.24	5.23	-4.74
Stomach	C16	-7.54	-7.50	-6.93	-6.84	-8.14	-4.94	-4.76	-5.95
Colorectal & Anus	C18-C21	-2.83	-2.41	-2.37	-3.32	-13.14	-1.61	-2.57	-2.02
Liver	C22	-11.26	-8.42	-6.56	-2.61	-13.00	-6.28	-7.03	-3.24
Gallbladder	C23-C24	-	-	-2.61	1.02	-	-4.73	-2.96	-0.44
Pancreas	C25	-2.74	-0.73	-1.05	0.13	-2.68	-1.01	0.62	1.68
Larynx†	C32	-	-11.82	-7.13	-6.50	-4.66	-	3.91	-12.01
Lung	C33-C34	-4.29	-4.73	-4.72	-1.89	-	-3.90	-3.89	-2.23
Malignant melanoma of skin	C43	-2.80	-0.12	-4.87	-1.11	-2.33	-1.65	-0.25	3.07
Breast	C50	-	-	-	-	0.88	0.24	-0.09	1.96
Cervix uteri	C53	-	-	-	-	1.07	-0.02	-4.57	-5.58
Other of uterus ‡	C54-C55	-	-	-	-	0.79	1.57	2.21	-0.16
Ovary†	C56	-	-	-	-	-	0.83	1.19	0.60
Prostate†	C61	-	-2.72	-0.99	0.80	-	-	-	-
Kidney	C64	-	-	-2.14	-1.94	-0.36	-0.27	-0.96	-0.35
Bladder†	C67	-	-2.72	-2.75	-0.67	-	-0.27	-2.70	-0.74
CNS	C70-C72	-1.46	-1.83	-1.51	0.73	-1.54	-1.10	-0.65	-1.93
Non-Hodgkin lymphoma	C82-C86	-3.33	-2.85	-2.06	0.82	-3.57	-2.23	-1.47	2.14
Multiple myeloma†	C90	-	-4.48	-5.27	-0.28	-4.39	-3.37	-3.07	2.23
Leukemia†	C91-C95	-3.32	-2.61	-2.32	0.33	-	-3.25	-1.69	0.76

LOC: Lip, oral cavity, and pharynx; CNS: Center Nervous System

†If the estimated mortality case is too small due to the low number of deaths in the comparative group, it is marked as †

3.2.3 Final selection of ageing-related cancers considers to both incidence and mortality

Table 7-1 summarizes our selection process for identifying age-related cancers. Initially, we selected only those cancers that ranked in the top 10 for incidence across all selected age groups, including over age 65. Next, we further screened for age-related cancers by combining the presence or absence of shifting with cancers that had a dominant incidence over age 65. The values in the table represent the answers to the conditional statements for males and females, respectively.

Among the cancers that were dominant in incidence over age 65, the first screening condition was satisfied for men, but breast and thyroid cancers were excluded for women. Concerning shifting, cancers occurring in men were not included because they were already outside the top 10 in the incidence ranking, indicating that not all cancers are shifting.

Finally, for the cancers that were dominant but not shifting based on the first two conditions, the final selection included stomach, colorectal, liver, biliary tract, pancreas, lung, prostate, kidney, for men. In contrast, breast and thyroid cancers were excluded from the first two screening conditions for women, resulting in the final selection of stomach, colorectal, liver, biliary tract, pancreas, and lung cancers as age-related cancers for women.

Table 7-2 summarizes the results for the selected cancers when the previous conditions were combined. The description of the selection process is the same as for carcinoma selection in the previous incidence. For mortality, breast cancer in women was excluded because it did not meet the criteria, like the results for incidence. Consequently, esophagus, stomach, colorectal and anus, liver, gallbladder, pancreas, prostate, non-Hodgkin's lymphoma, were identified as age-related cancers in men, while stomach, colorectal, liver, gallbladder, pancreas, lung, cervical, non-Hodgkin's lymphoma, and leukemia were identified as ageing-related cancers in women.

Table 7-3 summarizes the cancers that were selected for aggregation by cancer for both incidence and mortality based on the cancers selected in the previous incidence and mortality tables. In some cases, the same cancers were selected in both incidence and mortality in the previous results, but in other cases, they were not. Therefore, if a cancer was selected in either incidence or mortality, it was included in the overall analysis. For males, the cancers added to incidence were esophageal, bladder, and non-Hodgkin lymphoma, and the cancers added to mortality were biliary tract and kidney. For females, the cancers added to incidence were cervical, non-Hodgkin lymphoma, and leukemia, and the cancers added to mortality were biliary tract. In total, the analysis was based on 12 cancers for males and 9 cancers for females.

Table 7-1. Cancer selection for the model projection with adjustment on ageing population in incidence by sex

Cancer Sites	ICD-10	Incidence rates dominated* in age 65+	Incidences shifting† to young generation	Selected‡ for projection modeling
Stomach	C16	Yes (Yes)	NO (NO)	Yes (Yes)
	C18-C20	Yes (Yes)	NO (NO)	Yes (Yes)
Colorectal	C22	Yes (Yes)	NO (NO)	Yes (Yes)
Liver	C23-C24	Yes (Yes)	NO (NO)	Yes (Yes)
Gallbladder	C25	Yes (Yes)	NO (NO)	Yes (Yes)
Pancreas	C33-C34	Yes (Yes)	NO (NO)	Yes (Yes)
Lung	C50	- (NO)	- (NO)	- (NO)
Breast	C61	Yes (-)	NO (-)	Yes (-)
Prostate	C64	Yes (-)	NO (-)	Yes (-)
Kidney	C67	Yes (-)	NO (-)	Yes (-)
Bladder	C73	- (NO)	- (NO)	- (NO)
Thyroid				

*The incidence rate in age 65+ is more than double of those in other age groups.

†The incidence rate in age 65+ is decreasing over the years, whereas the rate in any other age group is increasing.

‡Cancer is selected if its incidence rate dominated in age 65+, and the incidences were not shifted to young generation.

Table 7-2. Cancer selection for the model projection with adjustment on ageing population in mortality by sex

Cancer Sites	ICD-10	Mortality rates dominated* in age 65+	Mortality shifting† to young generation	Selected‡ for projection modeling
Esophagus	C15	Yes (-)	NO (-)	Yes (-)
Stomach	C16	Yes (Yes)	NO (NO)	Yes (Yes)
Colorectal & Anus	C18-C21	Yes (Yes)	NO (NO)	Yes (Yes)
Liver	C22	Yes (Yes)	NO (NO)	Yes (Yes)
Gallbladder	C23-C24	Yes (Yes)	NO (NO)	Yes (Yes)
Pancreas	C25	Yes (Yes)	NO (NO)	Yes (Yes)
Lung	C33-C34	Yes (Yes)	NO (NO)	Yes (Yes)
Breast	C50	- (NO)	- (NO)	- (NO)
Cervix uteri	C53	- (YES)	- (NO)	- (YES)
Prostate	C61	Yes (-)	NO (-)	Yes (-)
Bladder	C67	NO (-)	NO (-)	NO (-)
Non-Hodgkin lymphoma	C82-C86	Yes (Yes)	NO (NO)	Yes (Yes)
Leukemia	C91-C95	- (Yes)	- (NO)	- (Yes)

*The incidence rate in age 65+ is more than double of those in other age groups.

†The incidence rate in age 65+ is decreasing over the years, whereas the rate in any other age group is increasing.

‡Cancer is selected if its incidence rate dominated in age 65+, and the incidences were not shifted to young generation.

Table 7-3. Cancer selection for the model projection with adjustment on ageing population in incidence and mortality by sex

Cancer Sites	ICD-10	Male			Female		
		Incidence	Mortality	Selection	Incidence	Mortality	Selection
Esophagus	C15	-	Yes	Yes	-	-	-
Stomach	C16	Yes	Yes	Yes	Yes	Yes	Yes
Colorectal Colorectal & Anus	C18-C20 C18-C21	Yes	Yes	Yes	Yes	Yes	Yes
Liver	C22	Yes	Yes	Yes	Yes	Yes	Yes
Gallbladder	C23-C24	Yes	-	Yes	-	-	Yes
Pancreas	C25	Yes	Yes	Yes	Yes	Yes	Yes
Lung	C33-C34	Yes	Yes	Yes	Yes	Yes	Yes
Cervix uteri	C53	-	-	-	NO	Yes	Yes
Prostate	C61	Yes	Yes	Yes	-	-	-
Kidney	C64	Yes	NO	Yes	-	-	-
Bladder	C67	-	Yes	Yes	-	-	-
Non-Hodgkin lymphoma	C82-C86, C96 C82-C86	-	Yes	Yes	-	Yes	Yes
Leukemia	C91-C95	-	-	-	-	Yes	Yes

3.3 Validation selection of ageing-related cancer

To validate the selection of age-related cancers, we aimed to determine whether the incidence or mortality rates of these cancers were dominant among the elderly population between 2000 and 2019. We employed methods to identify patterns for cancer incidence and mortality rates to determine the extent to which aging affects cancer rates. The approach involved identifying patterns through age-period-cohort analysis using the lexis diagram. This method enabled us to identify the age groups with the highest incidence and mortality rates, as well as the average rate of increase or decrease according to age [66].

3.3.1 Age-Period-Cohort Effect in 2000 to 2019

In our age-period-cohort analysis, we focused on four main results:

- 1) Cross-Sectional Age Curve: Expected age-specific rates in reference period p_0 adjusted for cohort effects.
- 2) Longitudinal Age Curve: Expected age-specific rates in reference cohort c_0 adjusted for period effects.

The reference age and cohort period used in the analysis are as follows:

- Reference Age = (Number of Age Groups + 1)/2
- Reference Period = (Number of Periods + 1)/2
- Reference Cohort = (Reference Period – Reference Age + Number of Age Groups)

3.3.1.1 Validation male incidence and mortality rates

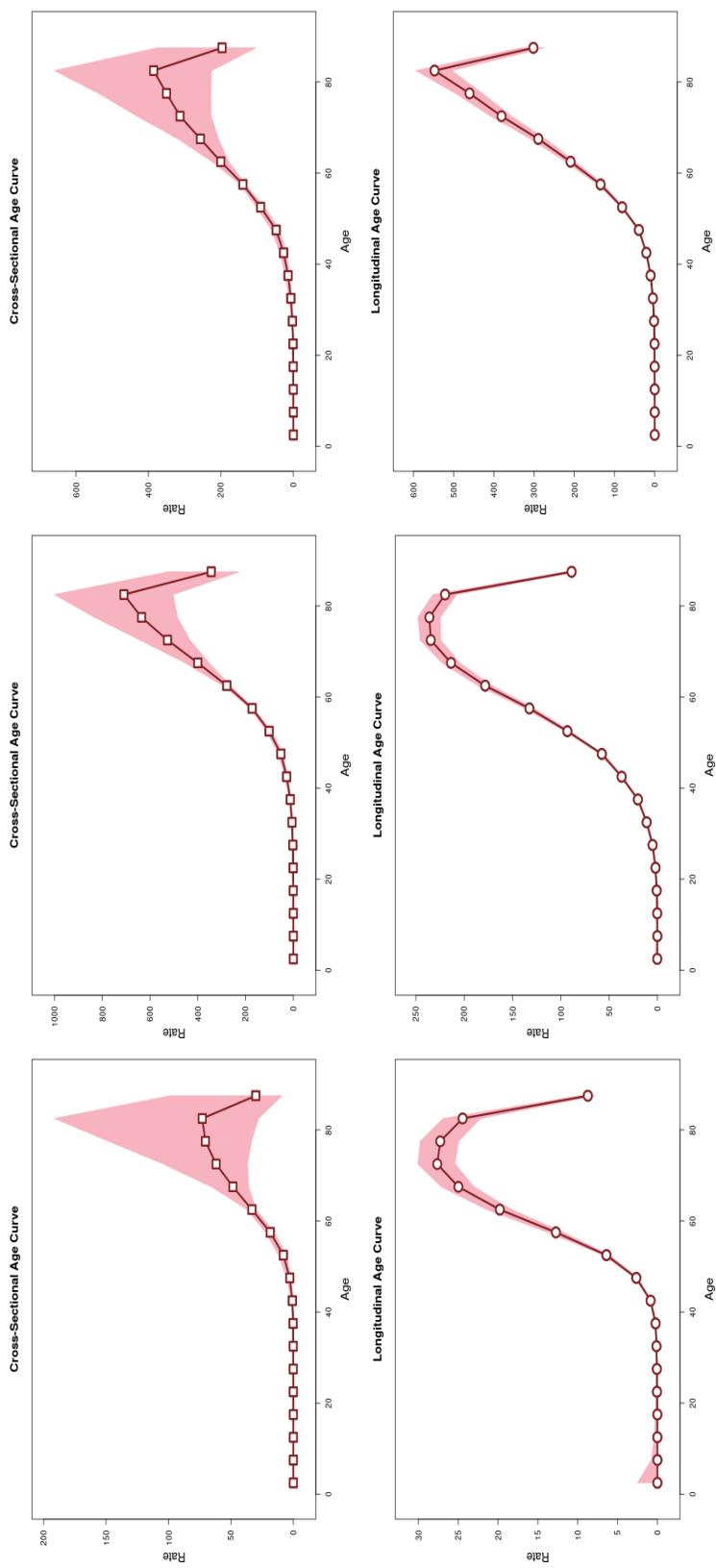
Figures 3 and 4 display the intricacies of age-specific and cohort-specific incidence and mortality rates for males. Each figure is bifurcated: the superior plot elucidates the temporal evolution of average incidence and mortality rates categorized by age, while the inferior plot focuses on the corresponding rates arranged by cohort.

Regarding incidence, an escalation in cohort-specific rates was observed as men entered older age brackets, peaking before a decline in the oldest age group, those 85 years and above. This decreases likely results from attrition attributable to deaths in this extremely old age group. Notably, this cohort-specific variation remains small across all cancer types. In a similar vein, the average incidence rate, delineated by age, also displays an upward trend in older age groups. However, the magnitude of deviation varies for each cancer type. A minor difference in deviation signifies a constant trend in incidence rates from the past to present, while a substantial difference suggests a steeper rise in the present relative to the past. High incidence rates are manifested in stomach, colorectal, liver, pancreatic, lung, and prostate cancers. Conversely, esophageal, colorectal biliary tract, and pancreatic cancers exhibit substantial deviations compared to other types. This implies that these cancers warrant heightened attention due to the current trend of a sudden upswing, diverging from past trends (Figure3).

In the realm of mortality, the effect intensifies in older age groups. Examining the average age-standardized mortality rate per cancer type reveals a notable

discrepancy from the incidence results. A modest rise is observed until reaching the oldest age group, but upon entering this category, a dramatic escalation ensues. This pattern is consistent across all types of cancer. In terms of mortality rates, discernible discrepancies exist between the average rates categorized by age and by cohort for certain cancers. For instance, in stomach and liver cancers, cohorts display a steep rise in mortality during middle age, but the age-specific mean incidence rates do not demonstrate a significant increase for this age group. This suggests that the current middle-aged cohort may influence long-term liver cancer mortality, necessitating careful monitoring. Consistent with the incidence rates, stomach, colorectal, liver, pancreatic, lung, and prostate cancers show high mortality rates, with lung cancer presenting the highest mortality rates (Figure4).

A. Esophagus Cancer (C15)



B. Stomach Cancer (C16)

C. Colorectal Cancer (C18-C20)

Figure 2-1. Age-Period-Cohort Effect for Male Incidence (Esophagus, Stomach, Colorectal Cancer)

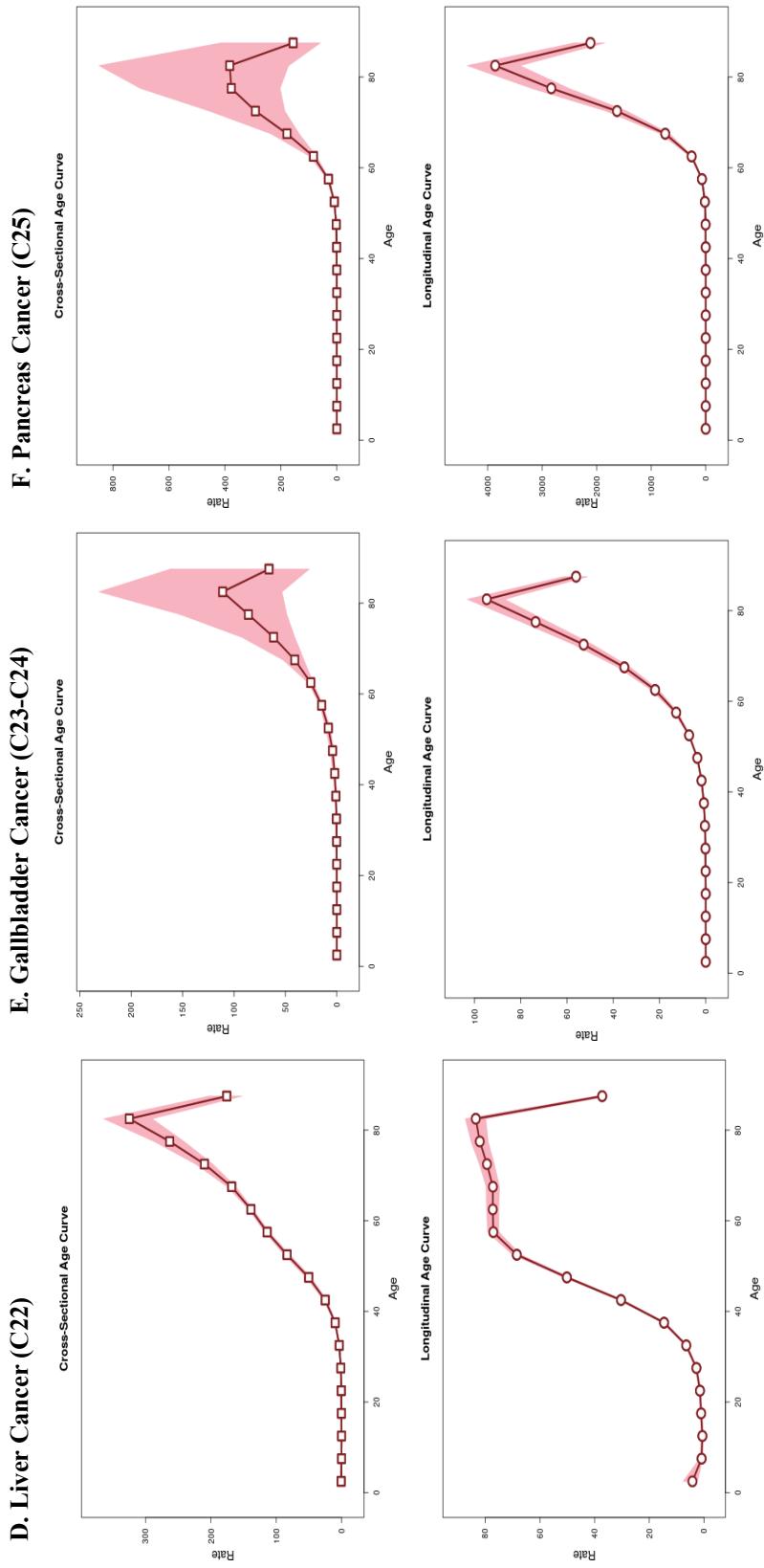


Figure 2-2. Age-Period-Cohort Effect for Male Incidence (Liver, Gallbladder, Pancreas Cancer)

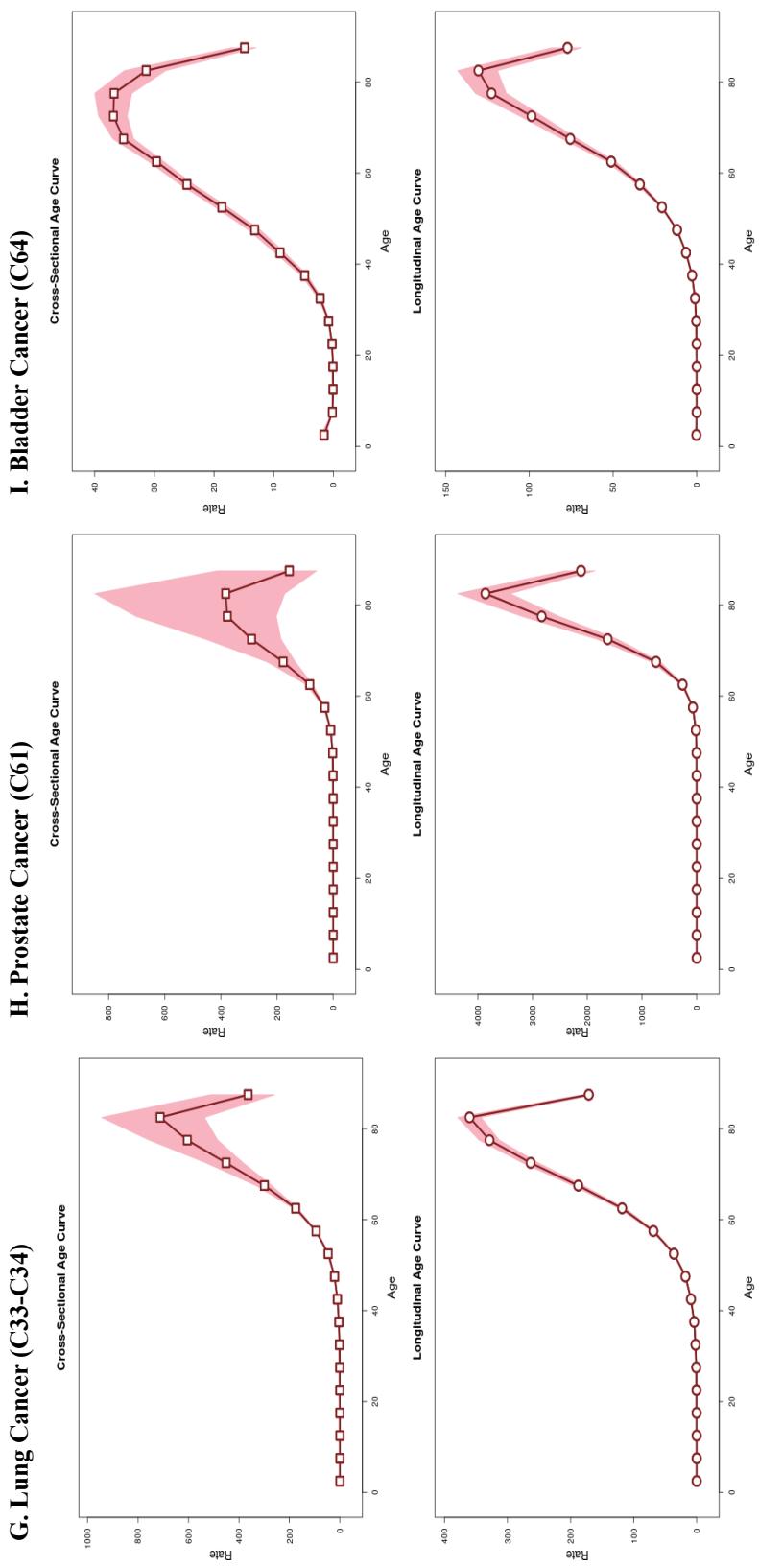
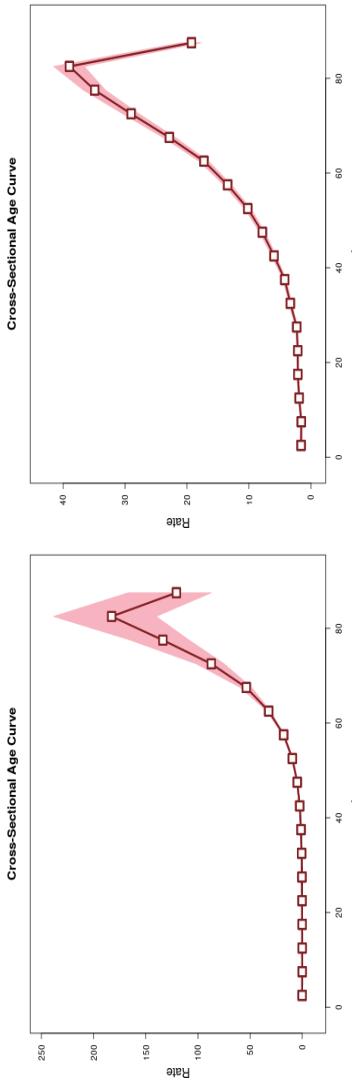


Figure 2-3. Age-Period-Cohort Effect for Male Incidence (Lung, Prostate, Bladder Cancer)

J. Kidney Cancer (C67)



K. Non-Hodgkin Lymphoma (C82-C86,C96)

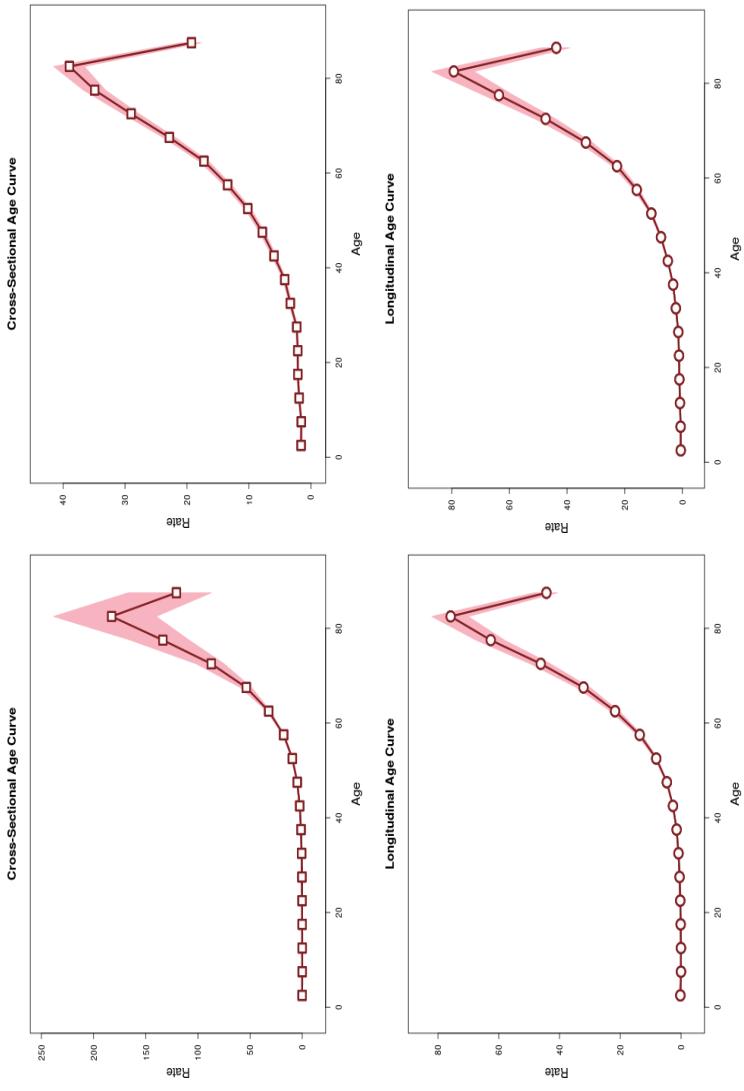
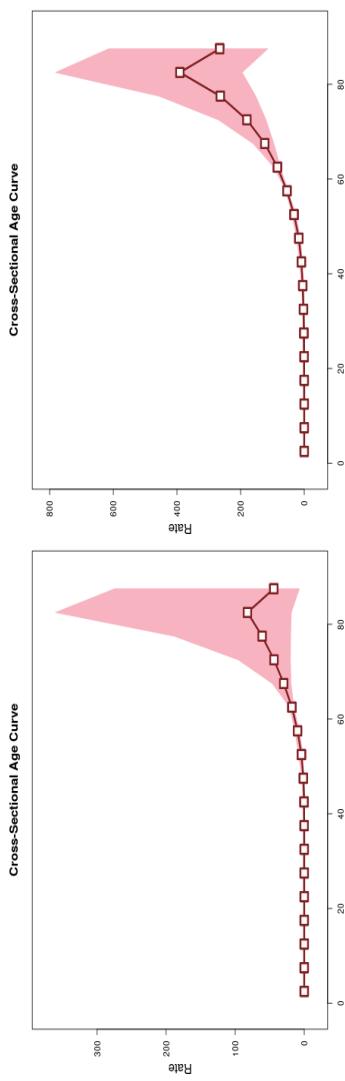
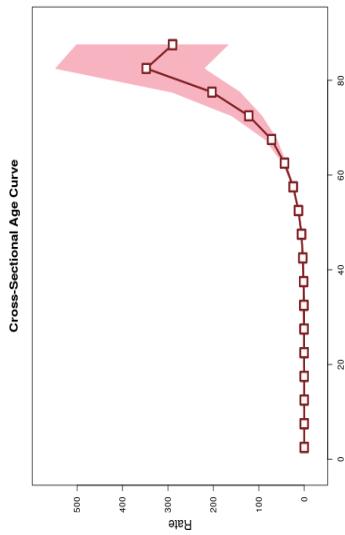


Figure 2-4. Age-Period-Cohort Effect for Male Incidence (Kidney Cancer, Non-Hodgkin Lymphoma)

A. Esophagus Cancer (C15)



B. Stomach Cancer (C16)



C. Colorectal & Anus Cancer (C18-C21)

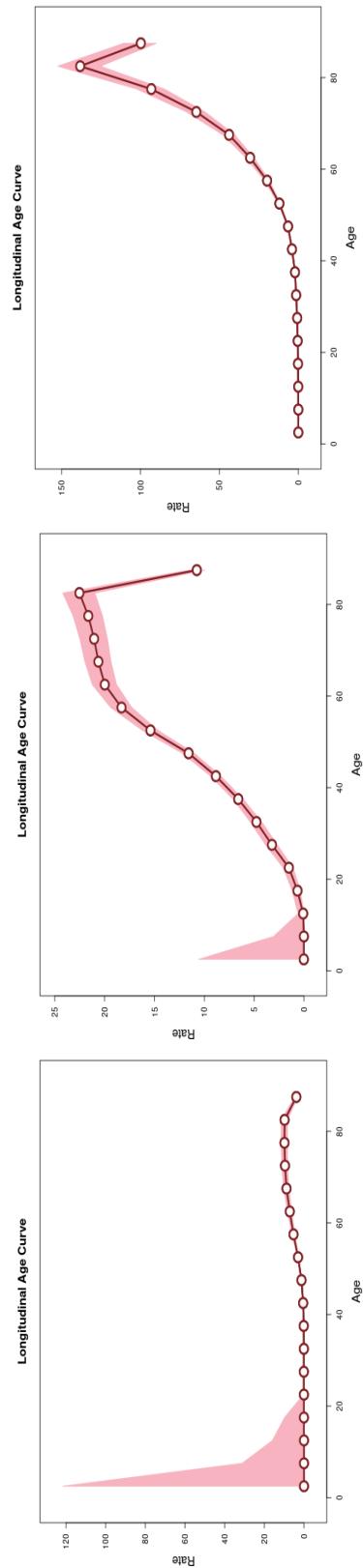


Figure 3-1. Age-Period-Cohort Effect for Male Mortality (Esophagus, Stomach, Colorectal Cancer)

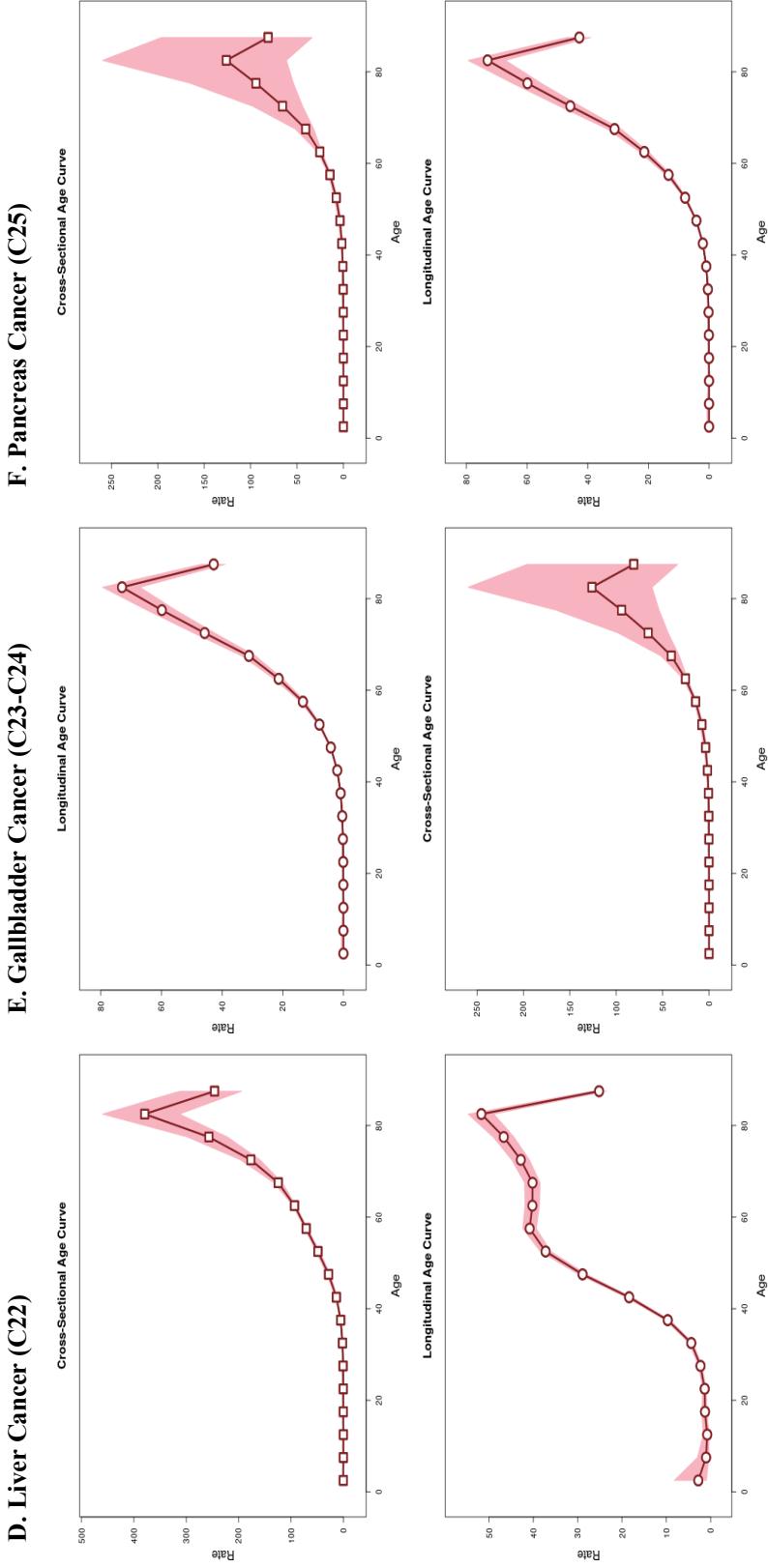


Figure 3-2. Age-Period-Cohort Effect for Male Mortality (Liver, Gallbladder, Pancreas Cancer)

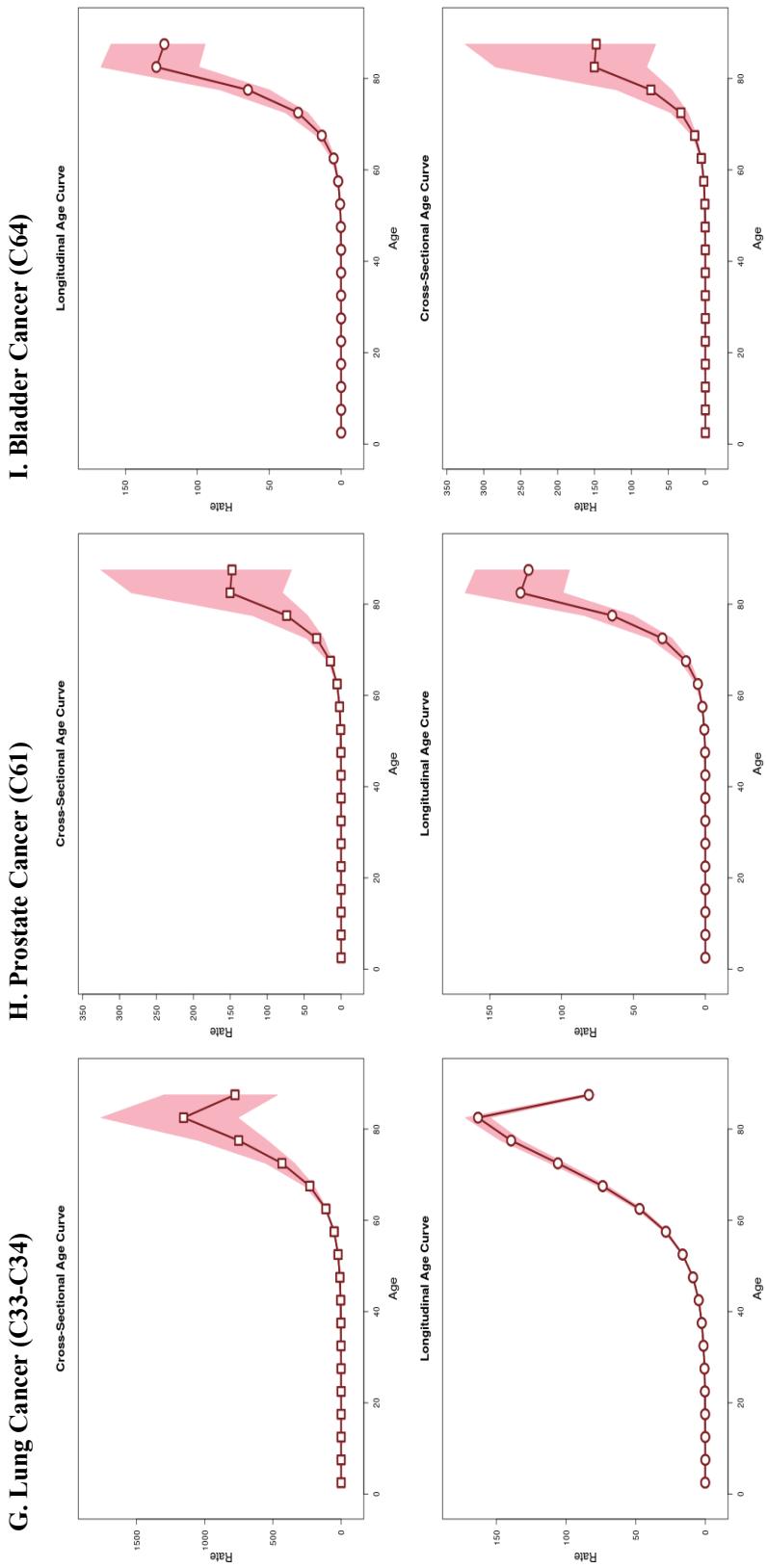
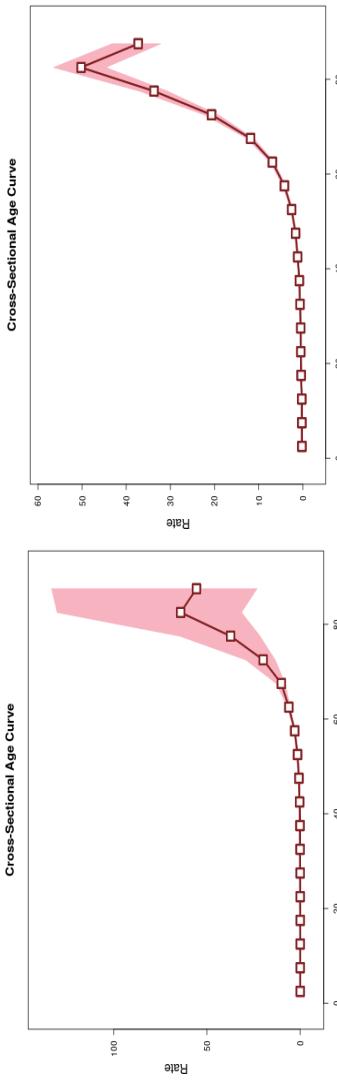


Figure 3-3. Age-Period-Cohort Effect for Male Mortality (Lung, Prostate, Bladder Cancer)

J. Kidney Cancer (C67)



K. Non-Hodgkin Lymphoma (C82-C86,C96)

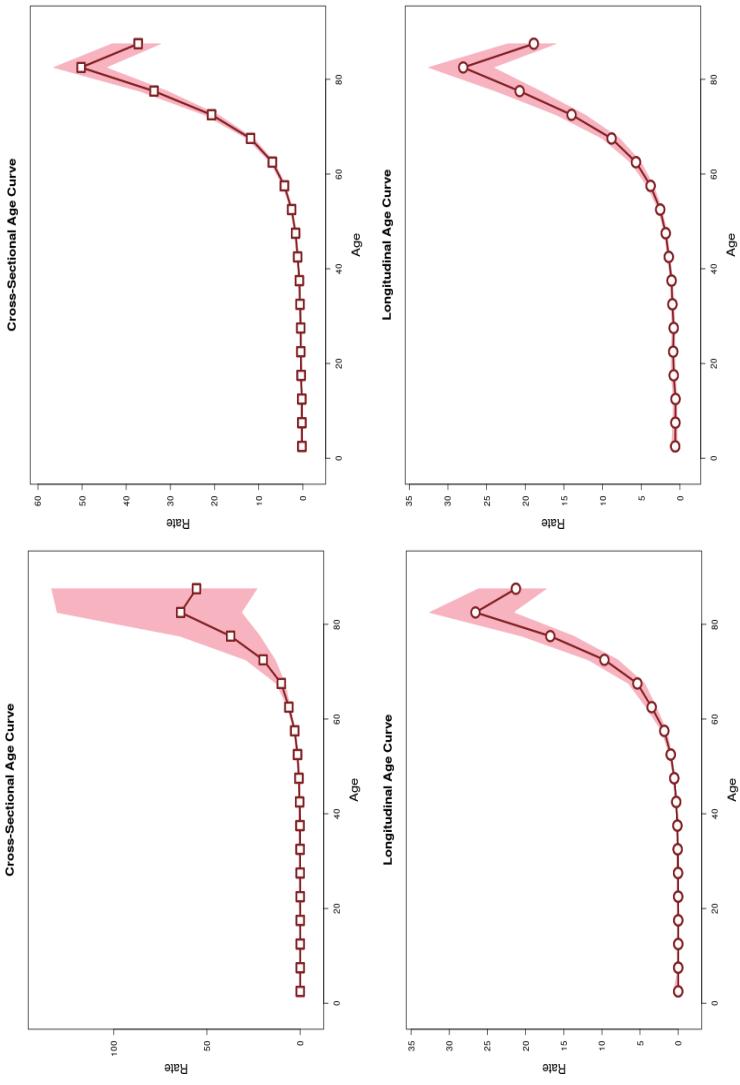


Figure 3-4. Age-Period-Cohort Effect for Male mortality (Kidney Cancer, Non-Hodgkin Lymphoma)

3.3.1.2 Female incidence and mortality rates

Figures 5 and 6 elaborate on the age- and cohort-specific incidence and mortality for women, respectively, aligning with the previous analysis performed for males. Similarly, to men, the cohort-specific incidence rates for women indicated a higher rate among older age groups and a decrease in the group aged 85 years and older. The deviations remained minimal for all cancer types within these cohorts.

Age-specific mean incidence rates also exhibited a higher incidence rate in older age groups for all cancers, barring uterine trunk cancer, but the deviation varied among different cancers. A particularly noticeable variation was identified in the incidence of uterine cervical cancer, which displayed a broad range of incidence across all age groups beyond the twenties, peaking in the middle-aged group and subsequently declining towards older age groups. This trend was also mirrored in the cohort data, indicating that the incidence rate of this cancer was consistently higher among middle-aged and especially younger women than among older age groups. In terms of incidence, high rates were observed for stomach, colorectal, liver, biliary tract, and lung cancers. Among these, colorectal cancer imposed the greatest burden with the highest incidence rate.

In relation to mortality, as previously examined, older age groups bear a significant impact. The average age-standardized mortality rate by cancer type diverges from the earlier incidence results. The rise is relatively modest until the oldest age group is reached, after which there is a considerable increase in the rate

until the very old age group is considered. This pattern holds regardless of cancer type. Furthermore, certain cancers demonstrated a discrepancy between the average mortality rate by age and by cohort. Leukemia, for example, followed a similar age-specific incidence pattern to other cancers but exhibited an unusual cohort-specific pattern, with particularly high incidence rates in early childhood and at older ages. In terms of mortality, stomach, colorectal, liver, pancreatic, and lung cancers showed high rates, with lung cancer resulting in the highest mortality.

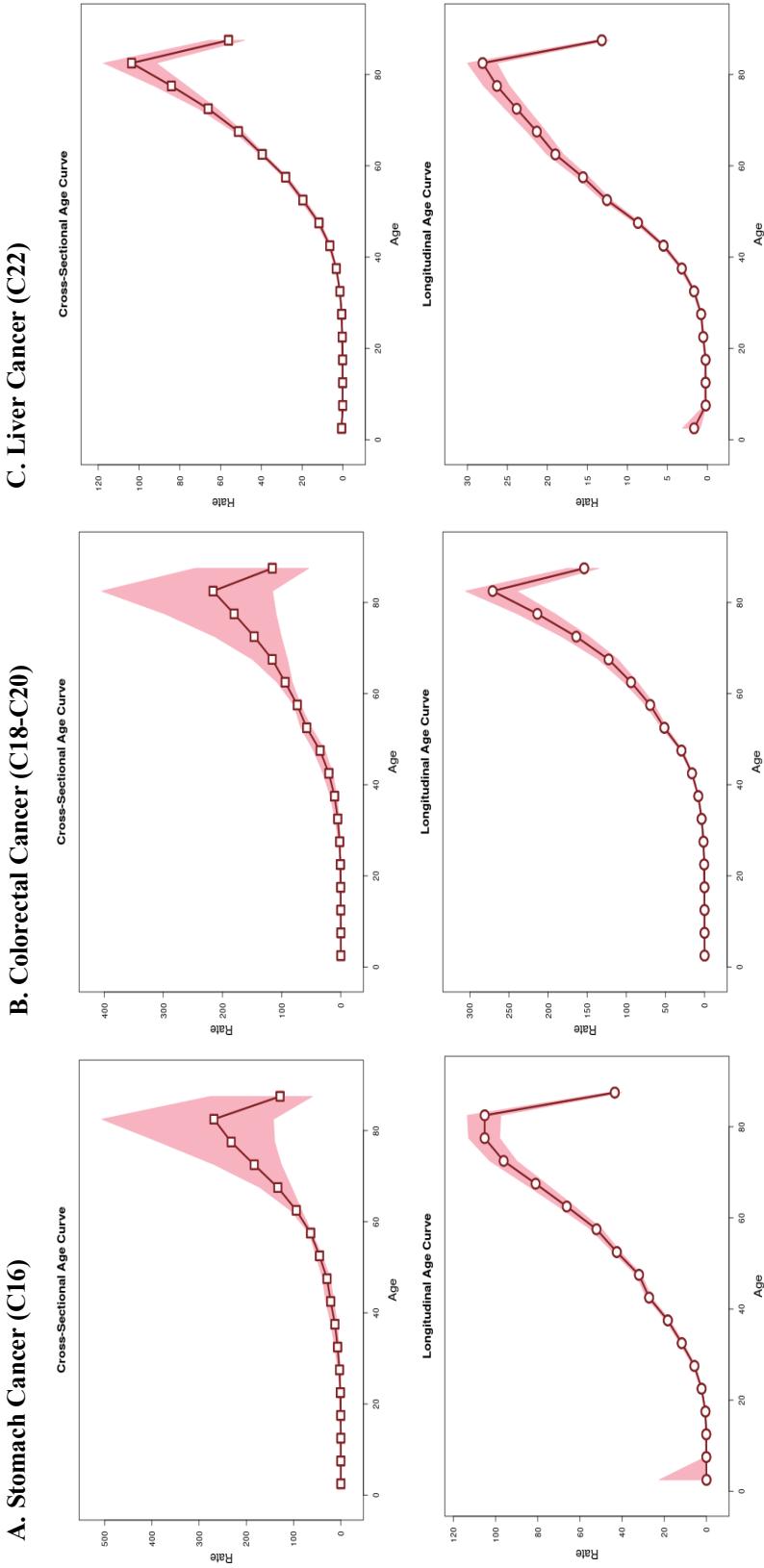
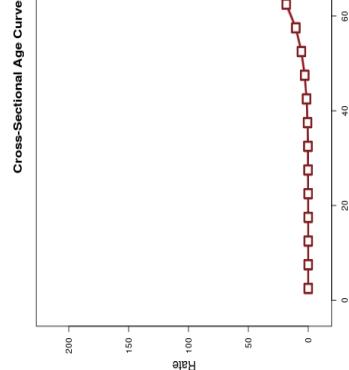
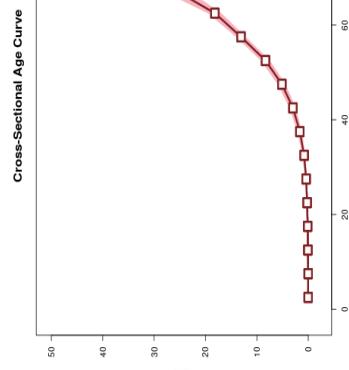


Figure 4-1. Age-Period-Cohort Effect for Female Incidence (Stomach, Colorectal, Liver Cancer)

D. Gallbladder Cancer (C23-C24)



E. Pancreas Cancer (C25)



F. Lung Cancer(C33-C34)

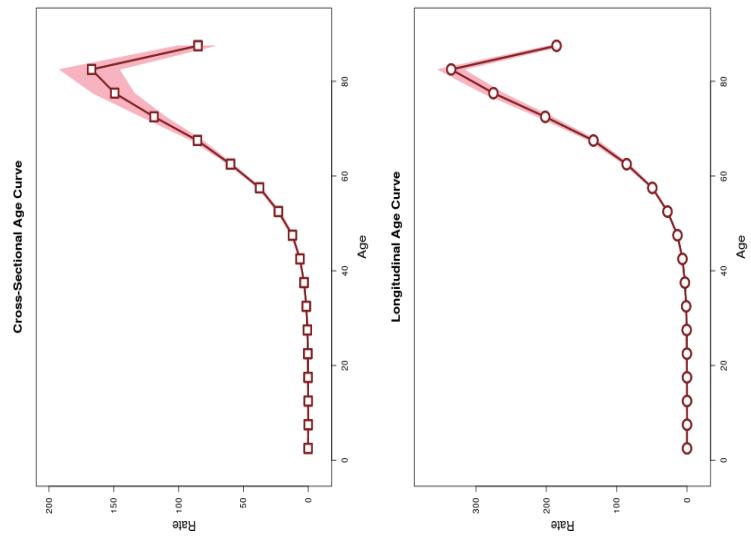


Figure 4-2. Age-Period-Cohort Effect for Female Incidence (Gallbladder, Pancreas, Lung Cancer)

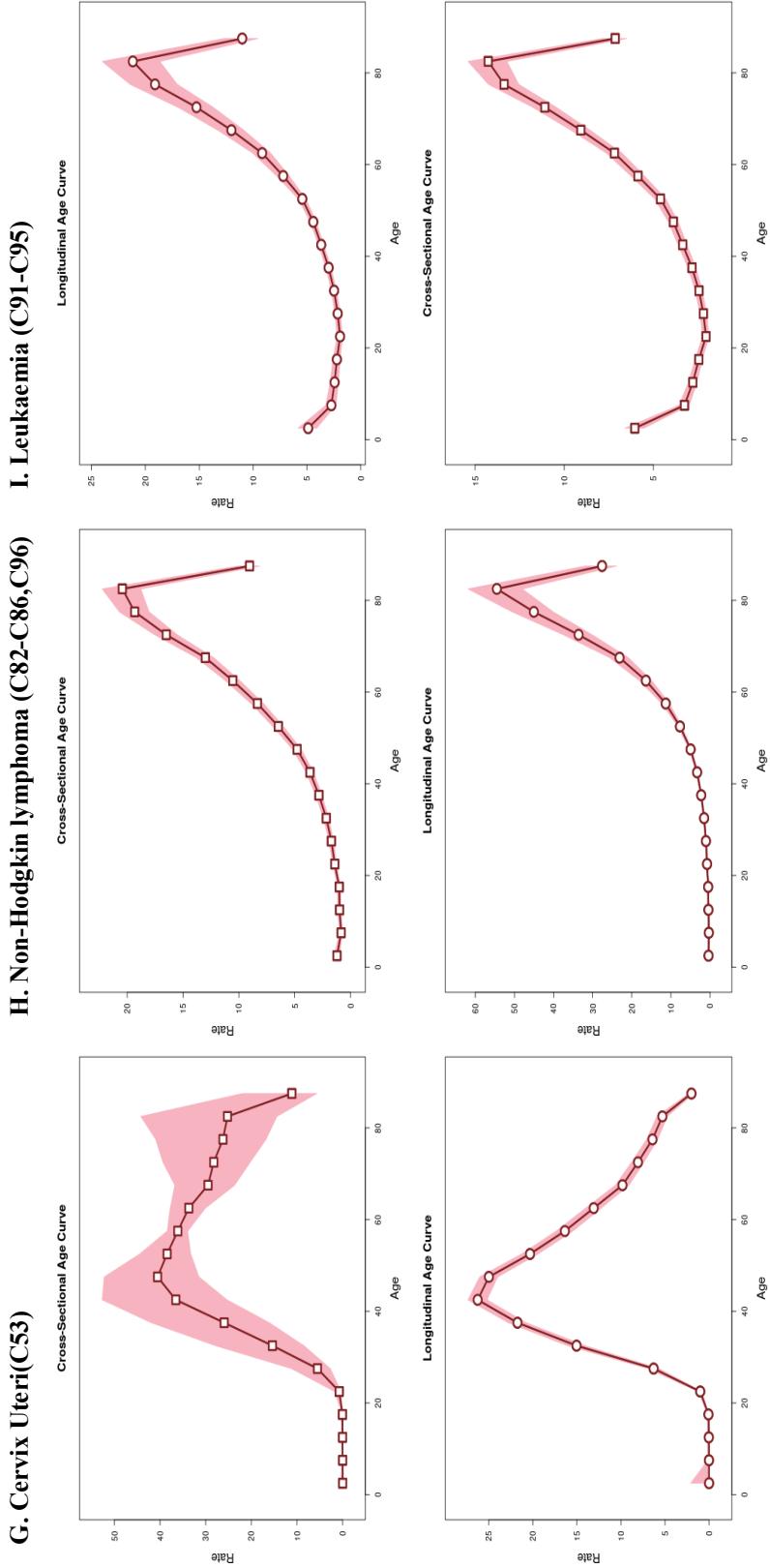


Figure 4-3. Age-Period-Cohort Effect for Female Incidence (Cervix Uteri Cancer, Non-Hodgkin Lymphoma, Leukemia)

A. Stomach Cancer (C16)

B. Colorectal & Anus Cancer (C18-C21)

C. Liver Cancer (C22)

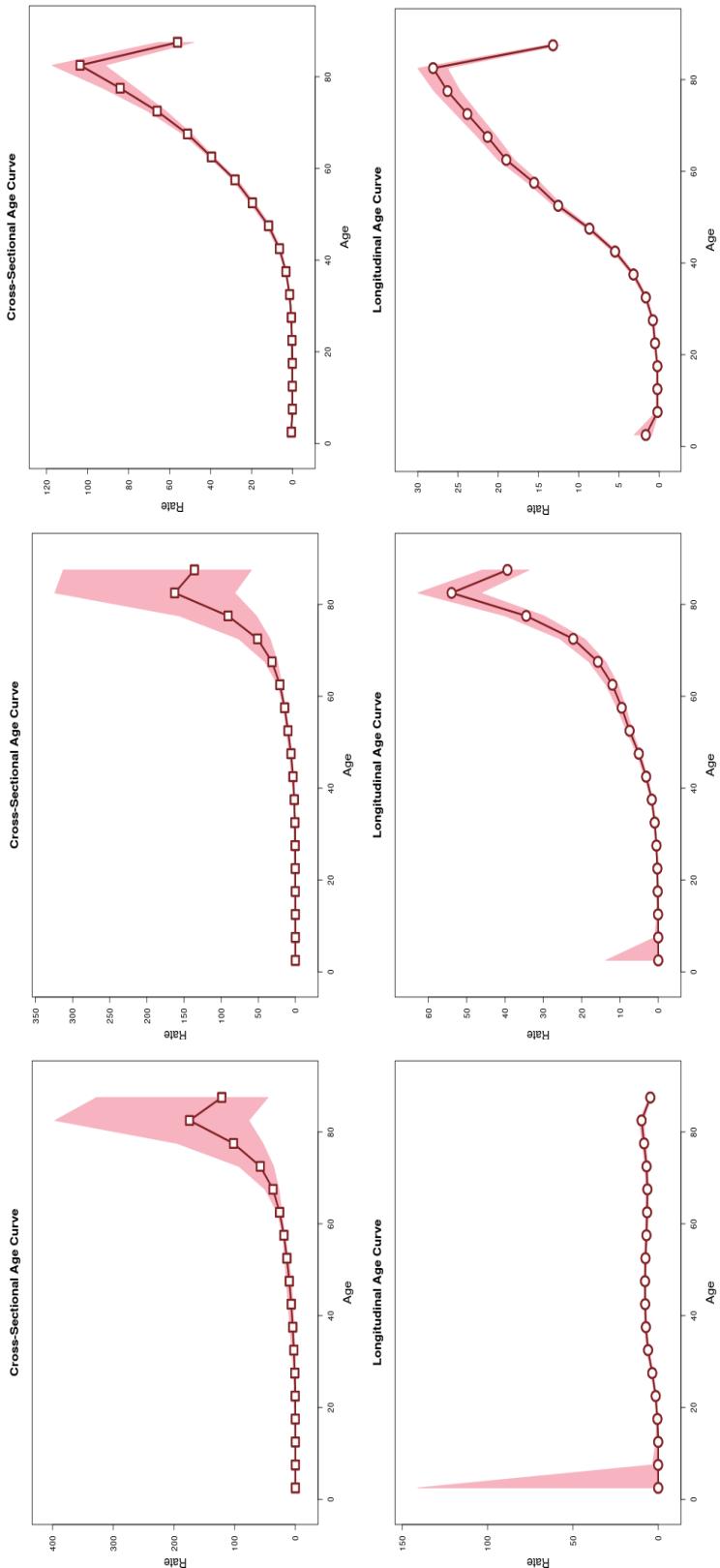


Figure 5-1. Age-Period-Cohort Effect for Female Mortality (Stomach, Colorectal, Liver Cancer)

D. Gallbladder Cancer (C23-C24)

E. Pancreas Cancer (C25)

F. Lung Cancer(C33-C34)

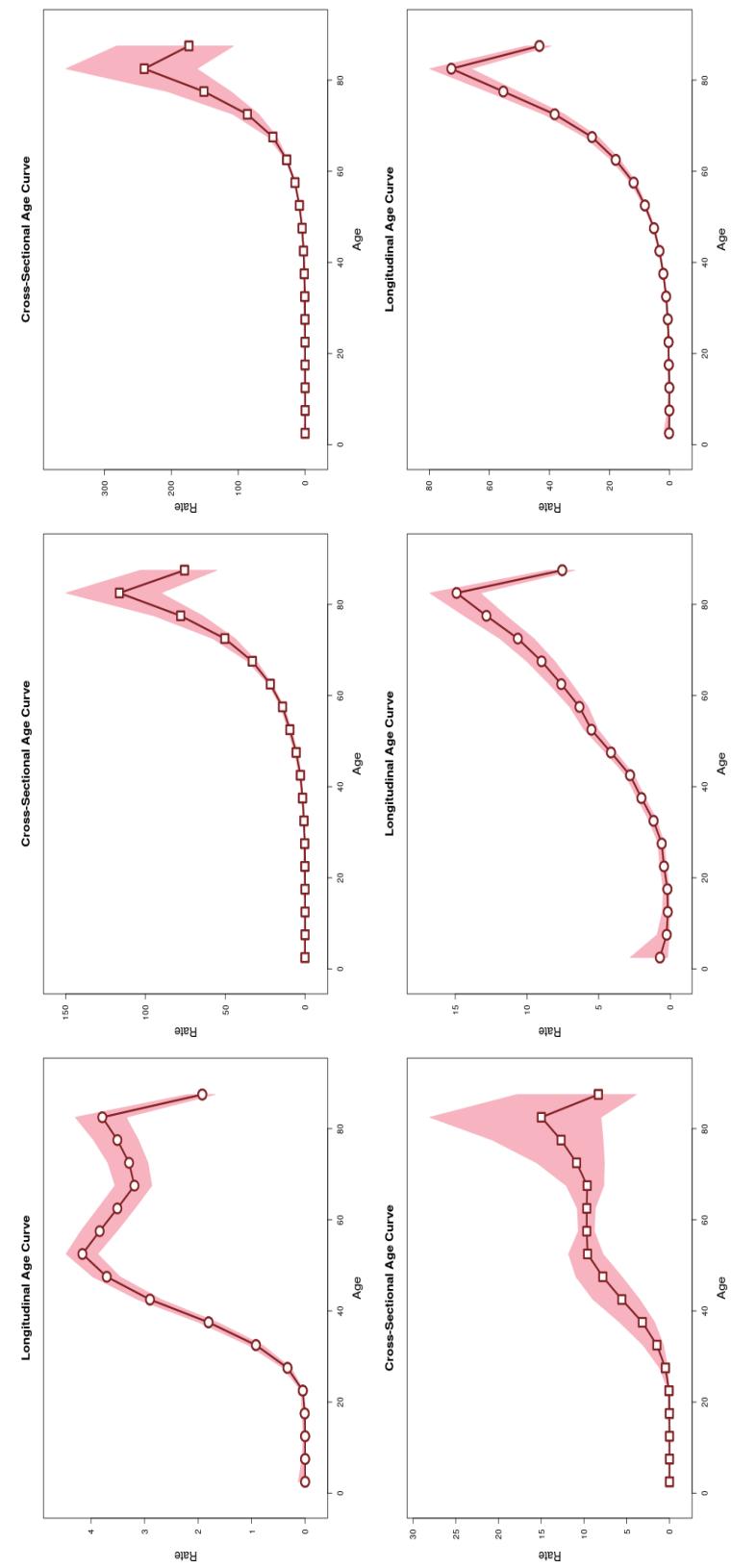
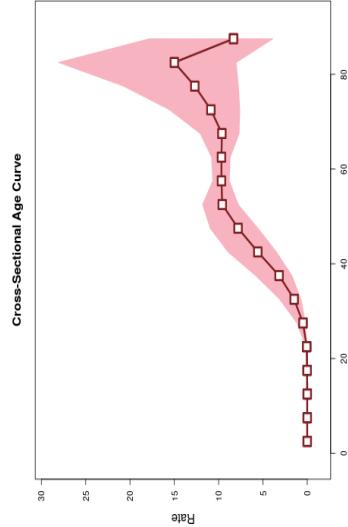
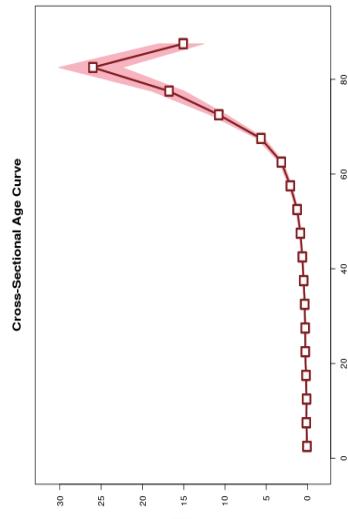


Figure 5-2. Age-Period-Cohort Effect for Female Mortality (Gallbladder, Pancreas, Lung Cancer)

G. Cervix Uteri(C53)



H. Non-Hodgkin lymphoma (C82-C86)



I. Leukaemia (C91-C95)

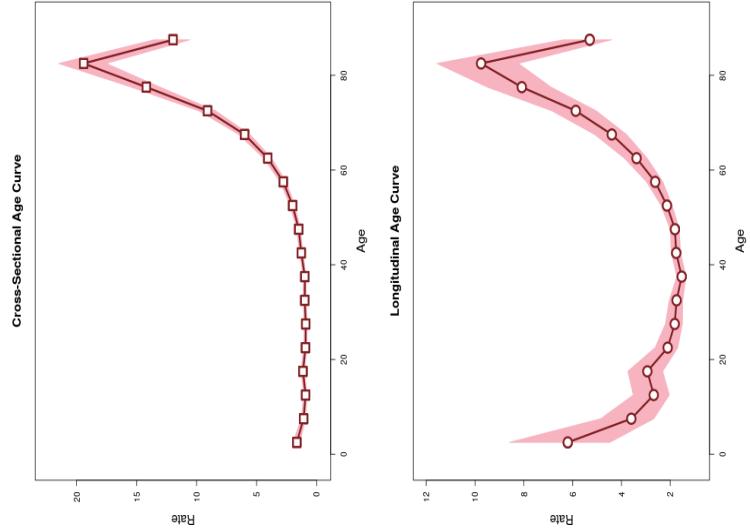


Figure 5-3. Age-Period-Cohort Effect for Female Mortality (Cervix Uteri Cancer, Non-Hodgkin Lymphoma, Leukemia)

3.4 Summary

In Chapter 3, we detailed the process of selecting age-related cancer types and verifying the validity of their aging factors. We began by selecting the top 10 cancers for both incidence and mortality in all age groups and age groups over 65 in the most recent year in 2019. We then selected cancers that were suitable for age-related cancers by identifying those with a steady increase in incidence or mortality under 65 ages and a decrease in incidence or mortality over 65 ages using the AAPC and join point for each age group over the last 10 years. In addition, for the analysis of the same cancer type in both incidence and mortality, we decided to analyze the cancer type in both analyses if it was present in either incidence or mortality, regardless of the condition of the selected cancer type. For both men and women, the primary cancers selected were stomach, liver, pancreas, and lung, which are representative cancers identified in previous Korean studies.

When conducting age-period-cohort analysis, we found that most of the selected cancers exhibited high incidence or mortality rates among the elderly (65+), and the increasing rates when considering the age-specific rate or cohort-age specific rate suggested that the high incidence or mortality among the elderly (65+) would persist in the future. It suggests that the increase in incidence or deaths among the elderly will intensify. Therefore, the selected cancers represent a reasonable selection of age-related cancers based on the age-period-cohort analysis.

CHAPTER IV. COMAPRISON BY PROJECITON MODELS WITH AND WITHOUT POPULATION STRUCTURE EFFECTS

4.1 Introduction

In this chapter, we aim to compare the results of future long-term projections using projection models based on previously selected cancers. The selected model is divided into two parts: a population structure-based prediction model and a non-population structure-based prediction model.

The population structure-based model predicts the incidence or mortality rate by age and projects it by applying the incidence or mortality rate projected on the population estimated in the future population. This model is mainly an age-period-cohort (APC)-based model, and it is known to show superior predictive power over other models. However, this model has the disadvantage of being unlearned and reducing predictive power if the number of occurrences or deaths is small. In this study, we introduce the NP model, which is mainly used as an APC-based model, and the population structure-based model based on Bayesian APC.

The non-population structure-based model is a method of predicting a given observation or rate by fitting it to a nonlinear curve regardless of future population estimates. A typical method is the JP model, which is mainly used for medium and short-term estimation using the last period's model. In addition, time series models are mainly used, and the model is constructed and used in various ways depending

on the researcher. This study uses the most representative ARIMA and selects the optimal model through the Hyndman-Khandakar algorithms [59]. The second time series model is a space-state model, which schematizes the relationship through the state of input and output in circuit or electronics. This longitudinal estimation utilized data spanning from 2000 to 2019 and was carried out by projecting observations up to the year 2040.

4.2 Population structure-based model

In the present study, we employed two distinct APC-based models, namely the NP and BAPC models. The construction of a model utilizing the NP approach entails several fundamental assumptions, which are outlined below:

- 1) Age is grouped into 5-year age groups indexed by $a = 1, 2, \dots, A$ ($A = 18$)
- 2) Period is indexed by 5-year period index $p = 1, 2, \dots, P$, where p represents a 5-year period
- 3) The 10-year overlapped cohort in the Age-Cohort and NP models is indexed by $c = 1, 2, \dots, C$, where $C = A + P - 1$ and $c = A - a + p$
- 4) The “power 5” link function, $x0.2$, is used as the default link function,
- 5) Attenuating parameters are used to degenerate any linear trend.
- 6) Recent 10-year average age-specific rates are used for age groups with small numbers.

- 7) The number of age-specific cases, in the p^{th} period, Y_{ap} , follows a Poisson distribution with offset from the age-specific population size, n_{ap} , $Y_{ap} \sim Poi(\mu_{ap}, n_{ap})$

The age-drift-period-cohort model (NORDPRED)can be written as

$$h(\mu_{ap}/n_{ap}) = \alpha_a + D \times p + \beta_p + \gamma_c, p=1,2,\dots,P$$

where μ_{ap} is the Poisson mean parameter (expected count) in age group a and period p, D is the common drift parameter, α_a , β_p and γ_c , are the non-linear components for age group a, period p and cohort c, respectively. Throughout this document, h represents the link function in the GLM.

In this BAPC methods, we focus on projections, which allows us to bypass the need for additional constraints. We need to assume an interest in mortality or incidence rates for identical age groups but t periods ahead in the future, expressed as:

$$\log(\lambda_{i(J+t)}) = \mu + \alpha_i + \beta_{J+t} + \gamma_{k+t} + z_{i(J+t)}$$

where $k = M \cdot (I - i) + J$. To accomplish this, we must extrapolate the period and cohort effects following the RW2 model's structure. Specifically, if we have data up to period $J \geq 2$, the period effect at period $J + 1$ will exhibit the conditional distribution:

$$\beta_{J+1} | \beta_1, \dots, \beta_J, \kappa\beta \sim N(2\beta_J - \beta_{J-1}, \kappa\beta^{-1}).$$

For $t > 1$, indicating an interest in t -steps ahead forecasts rather than one-step ahead we have:

$$\beta_{J+1} | \beta_1, \dots, \beta_J, \kappa\beta \sim N((1+t)\beta_J - t\beta_{J-1}, \kappa\beta_{(1,2,\dots,t)}))$$

The conditional mean is given by a linear extrapolation of the last two period effects with a cubically increasing variance. Analogously, projections for the cohort effects are obtained. Identifiability is guaranteed using a linear extrapolation scheme, such as the RW2 model.

4.3 Non-population structure-based model

In the previous analysis, the results of the decomposition analysis on the differences between the predicted values of the two models for each cancer type were compared through a demographic structure-based model. In this analysis, we aim to compare the predicted values through three models by projecting long-term trends using a non-demographic structure-based model. The models used include the JP model, ARIMA, and the Linear Space-State Model (SSML), with the following optimal model selection criteria:

For the JP model, a model with up to three join points was estimated through a permutation test, and a prediction equation corresponding to the last period was

extracted to model the long-term trend until 2040.

In the case of ARIMA, the Hyndman-Khandakar algorithm was employed, and a model with the lowest AIC among the given models was selected. For the Space-State Model, the model's fit was used to determine the degree of nonlinearity according to the order.

SSML model was using various algorithms can be used to obtain the Maximum Likelihood Estimation (MLE) when estimate values. Among the studies conducted using the model, the Kalman filter algorithm was generally utilized to build a projection model, employing a universal method in this study.

4.4 Projection results between population structure-based model and non-population structure-based model

4.4.1 Projection results for male incidence and mortality rates by models

Figure 6 presents the long-term incidence and mortality rate estimates for cancer types in men by 2040. The NP model and BAPC model were used, with the former reflecting trends from the past 10 years and the latter applying random walk2 to predict age, period, and cohort, as well as drift to reflect the average incidence trend by cancer type. The results showed a significant difference in long-term predictions between the two models, with the BAPC model appearing to be less

adequate in fitting the data when the trend of increase or decrease by section is not constant and the range of increase or decrease is large. The predictions for some cancers differed significantly between the two models, indicating the problem of robustness to predictions and increased errors with longer prediction periods. However, lung cancers showed the most consistent and robust predictions across both models.

In the case of the time series model, since the irregularity of the differences in values in the observed values is modeled, the predictions may not reflect the recent trend when the trend of increase or decrease is in the recent year. In particular, the SSML model is subject to greater limitations due to its linear equations. On the other hand, ARIMA can reflect the recent trend to some extent, depending on the model selection, as the difference between models appears through the optimal algorithm.

For the JP model, no matter the previous observation state, the slope for future prediction is set based on the recent JP. This aspect is especially evident in prostate cancer, where a large slope has been formed within the previous five years, leading to a wide prediction range. Biliary tract cancer, pancreatic cancer, and kidney cancer, which formed join points in recent years, also exhibited high slopes in the JP prediction model, rather than showing general linearity.

The NP model applied a slowdown rate every year of the forecast to make it like the actual occurrence rate or mortality trend, showing a more pronounced slowdown due to the increase in the forecast year than the incidence rate. BAPC tended to overestimate overall, particularly for cancers with irregular mortality rates, such as esophageal cancer. The predicted values for some cancers, such as stomach and liver cancer, showed a decreasing trend over time, but with a slowdown in the decreasing trend, the pattern shifted to an increasing trend. The BAPC model predicted a smaller value for leukemia than the NP model, as it reflects future trends based on the average rate of change in the entire observed year.

In this figure, the time series models show JP prediction values regardless of cancer type classification, with the JOINPOINT model and the other two time series models overlapping on the coordinates. Unlike the incidence rate, the increase or decrease in mortality rate is not significant, so the difference value for the observed value is not large, and the global minimum in the process of minimizing the error appears to be small. In most cancer types, except esophageal cancer, the predicted values of the time series models were linear according to the average trend of the entire observation. However, for liver cancer, the change in value gradually decreased as the long-term trend entered, and the value was observed to be stationary. This was considered to have an insufficient effect in estimating the predicted value, even if it reflected the previously observed trend in 2018-2019 when estimated in the model.

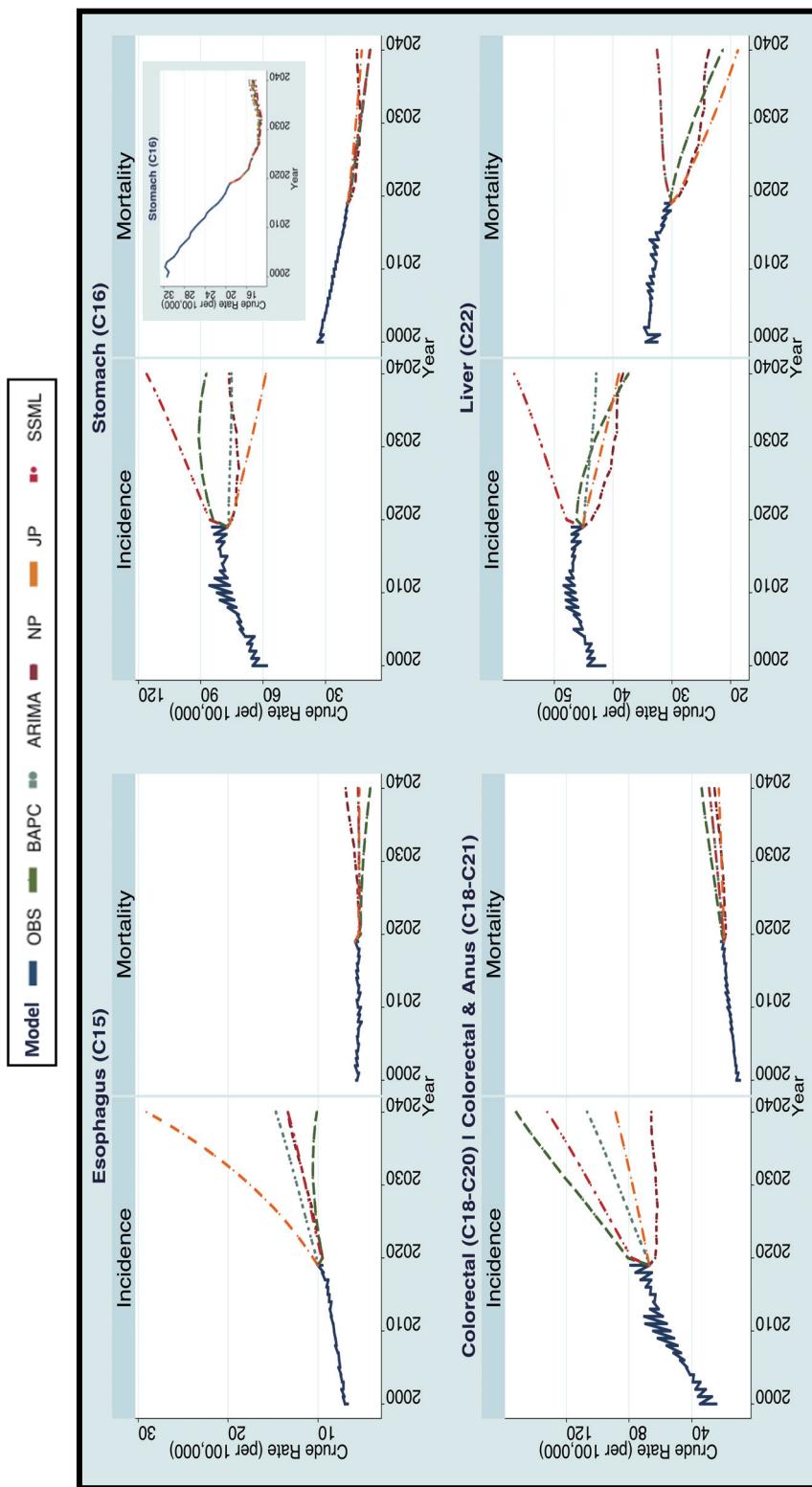


Figure 6-1. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Esophagus, Stomach, Colorectal, Colorectal & Anus, Liver Cancer)

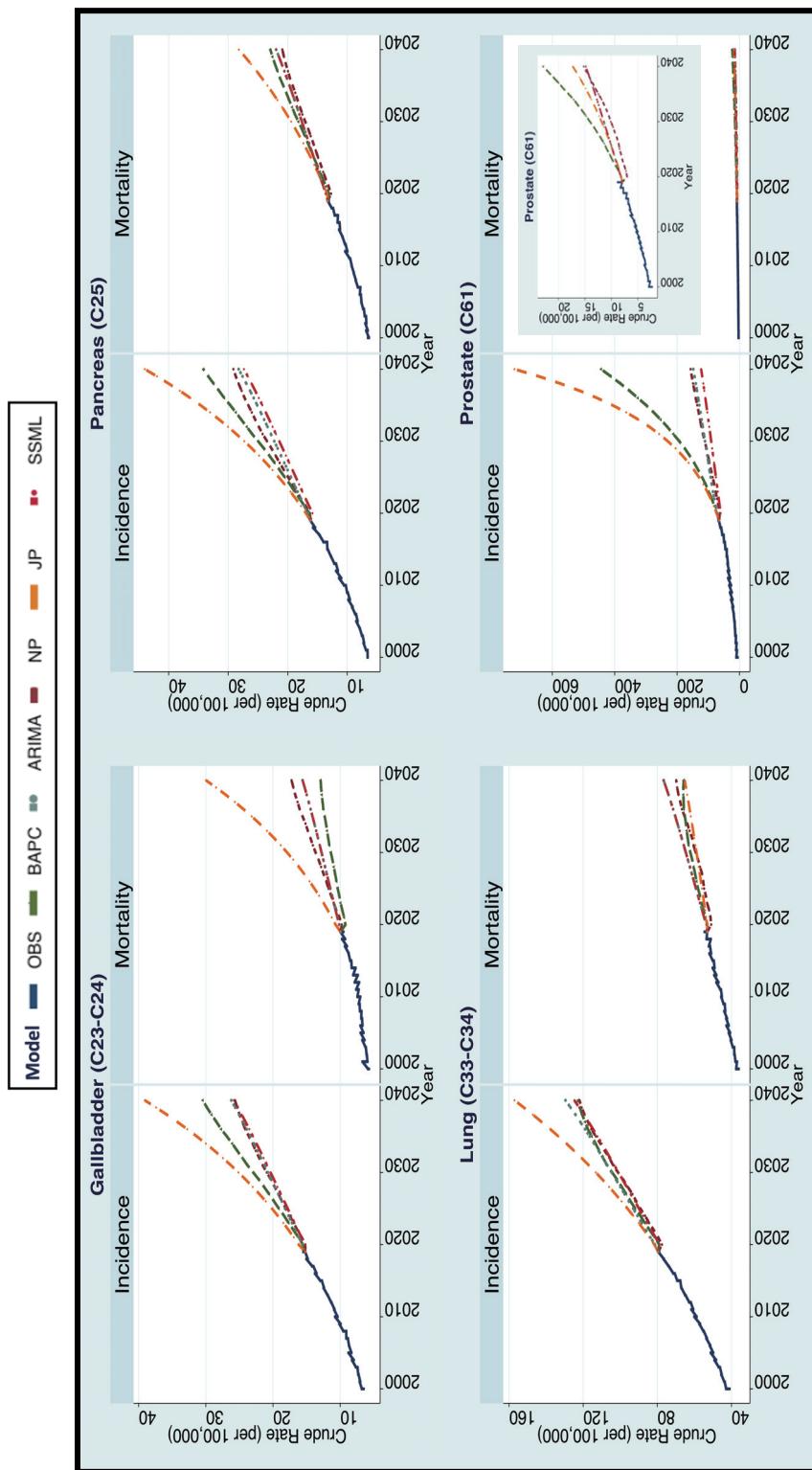


Figure 6-2. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Gallbladder, Pancreas, Lung, Prostate Cancer)

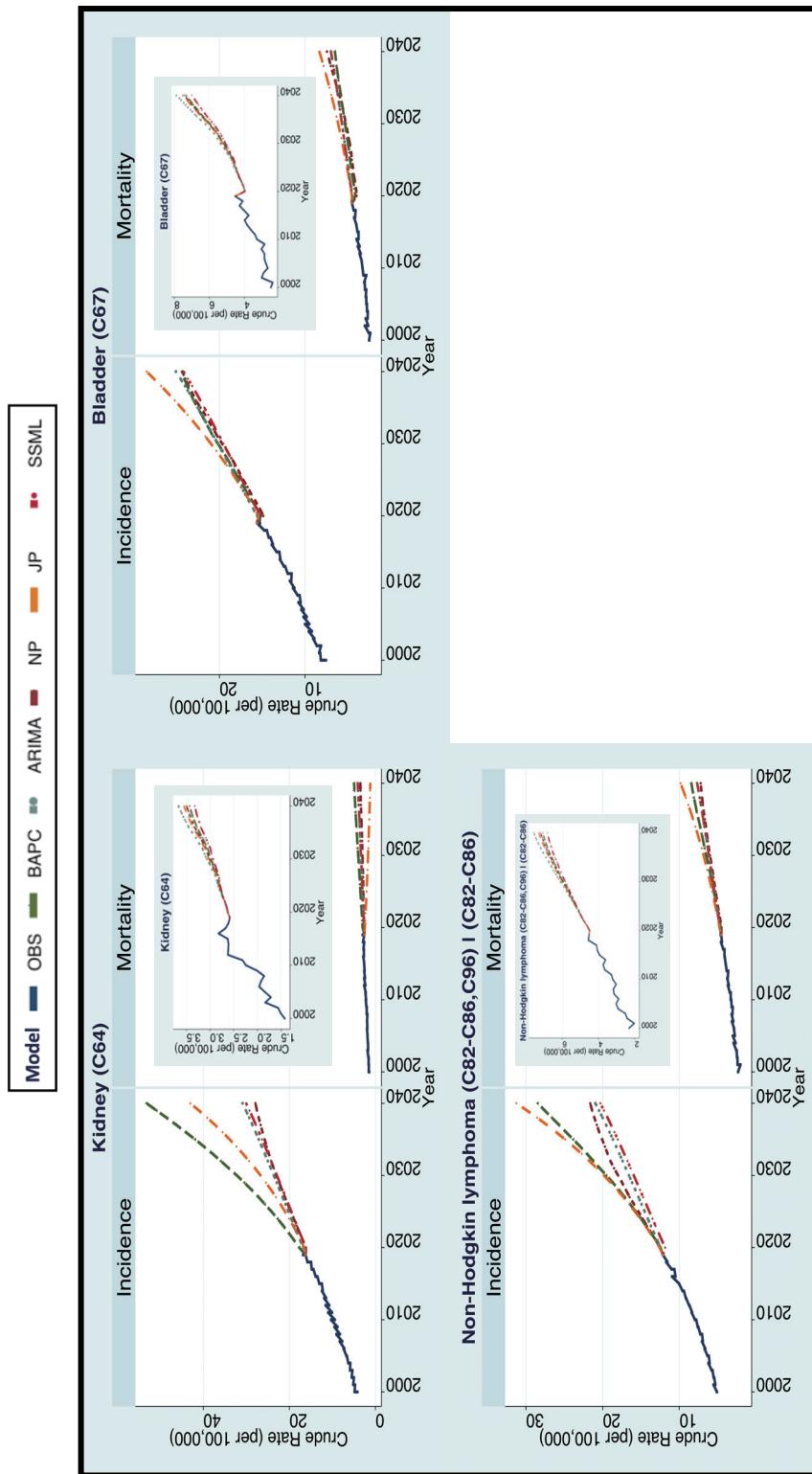


Figure 6-3. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Kidney, Bladder Cancer, Non-Hodgkin Lymphoma)

4.4.2 Projection results for female incidence and mortality rates by models

Figure 7 presents the long-term incidence and mortality rate estimates for cancer types in women by 2040. The trends exhibited by these models for cancer incidence in women show notable differences based on their respective characteristics. For lung cancer, both models display similar prediction trends during the observed period. However, due to the randomwalk2 characteristics, the disparity between the two models' predictions increases over time. In the cases of stomach and liver cancer, the predictive directions of the two models diverge as the patterns of increase and decrease fluctuate over time. Colorectal cancer has experienced a deceleration in the past decade, and the NP model predicts a specific percentage for long-term predictions, taking this trend into account. Conversely, the BAPC model continues to forecast a steady increase, based on the increasing trend observed in previous years, leading to a widening gap in predictions between the two models. Pancreas cancer displays a similar pattern, with the discrepancy in predictions gradually expanding over time.

In the previous analysis, it was confirmed that, like men, the predicted values in the time series model exhibited similar trends when linearity or specific patterns were present. In the case of the JP model, the slope of the prediction model may vary depending on the location of the JP, allowing for the observation of either linearity or exponential changes for each model. Particularly, for gastric and liver

cancer, the location of the inflection point is distant from recent years, resulting in a linear tendency to adjust the recent year observations in the model.

In contrast to the model fitting for incidence, the variance in mortality across different models was not particularly pronounced, with liver, gallbladder, and pancreas cancers. being an exception. Generally, the total number of deaths is significantly lower than the number of incidences, which likely leads to an insignificant difference in mortality in the fitted models. While the figures illustrate minimal variance in actual model predictions, given that the scales are adjusted to align with incidence criteria, amplification of these scales brings the model differences into clearer view.

However, caution should be exercised in applying these models for liver, gallbladder, and pancreas cancers. The differences in the predicted values for these cancers are relatively minor yet sufficient to induce disparities in trends across the various models. Therefore, it is critical to consider these differences when utilizing these models to interpret trends and make predictions.

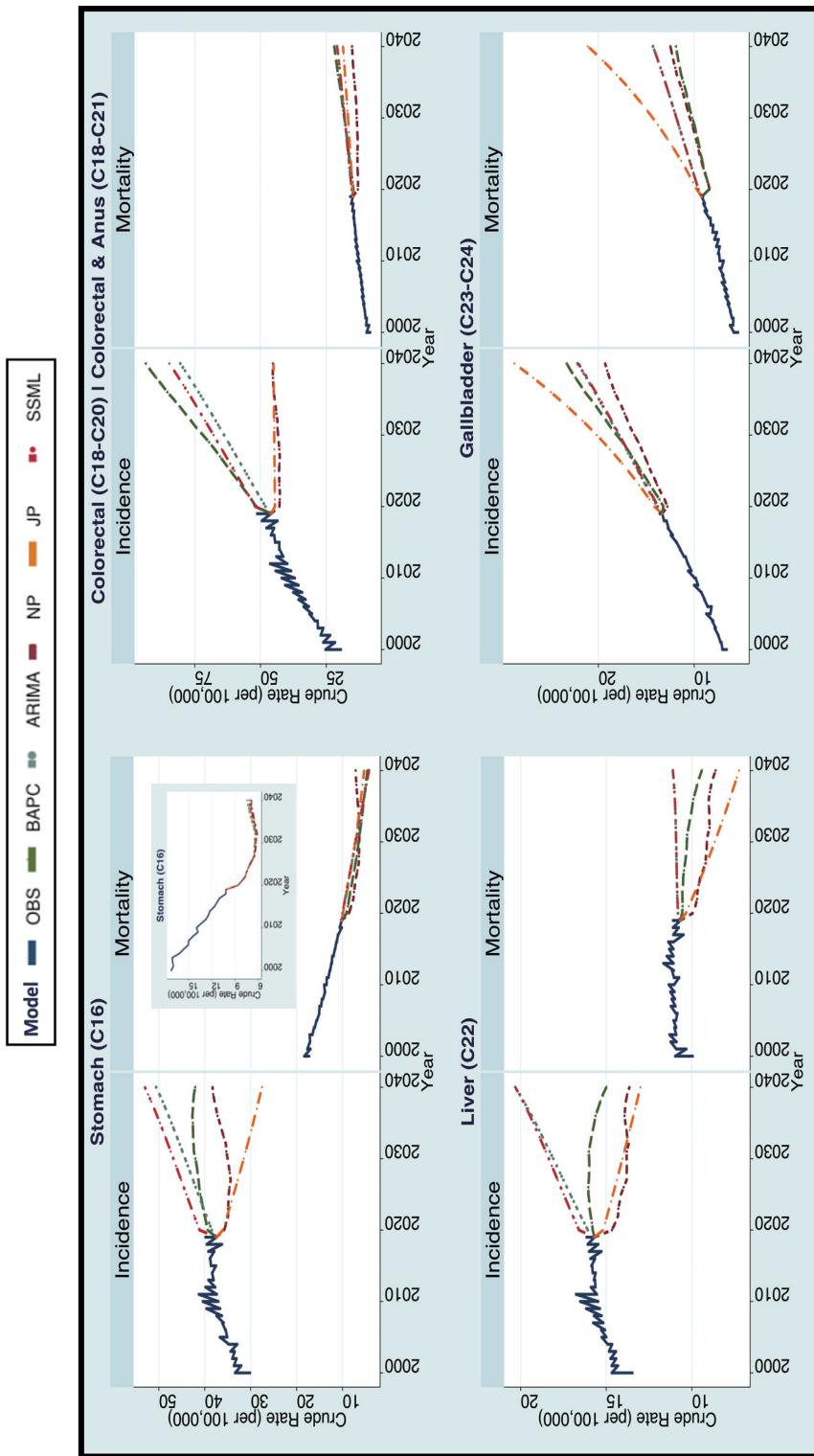


Figure 7-1. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder

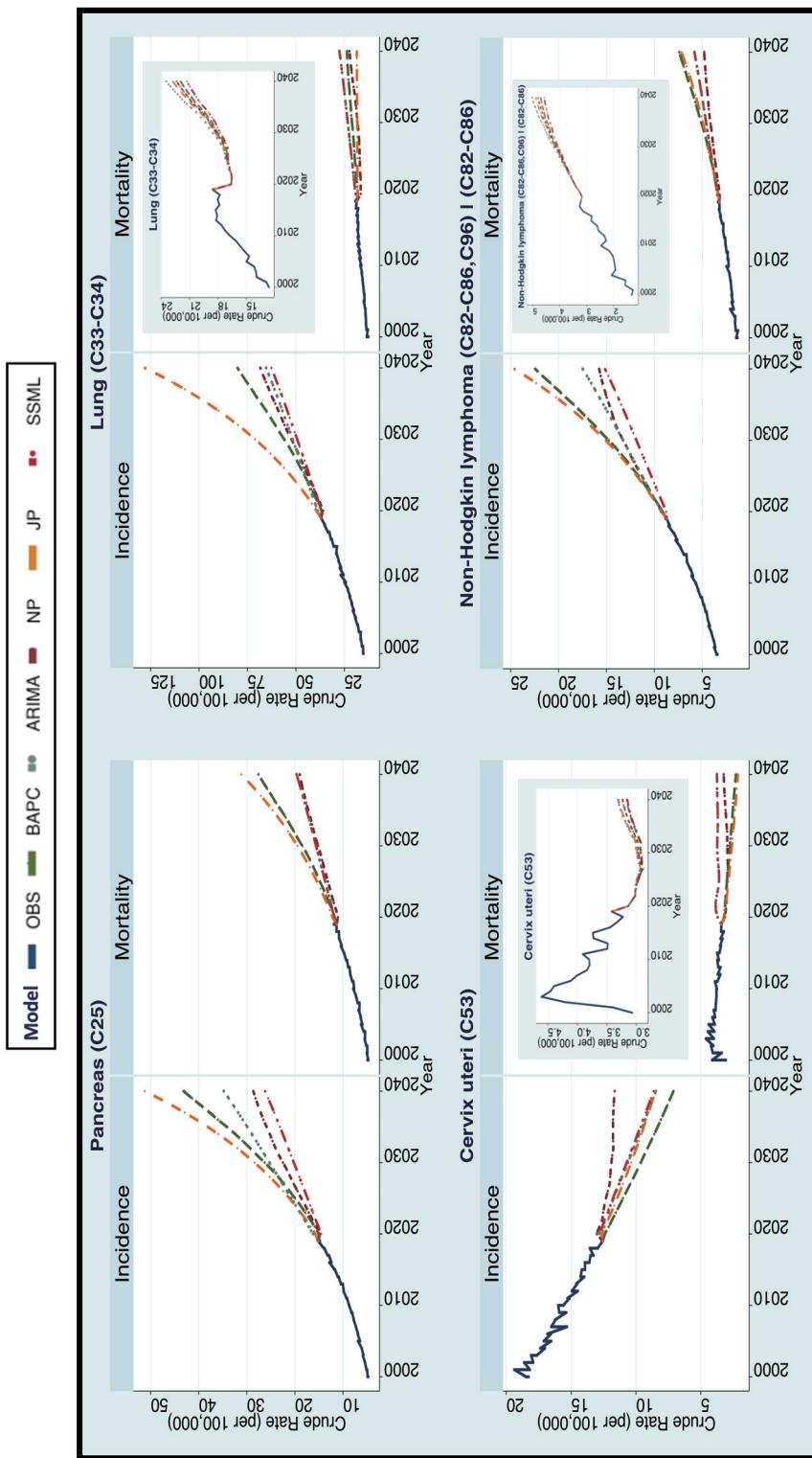


Figure 7-2. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Pancreas, Lung, Cervix uteri Cancer, Non-Hodgkin

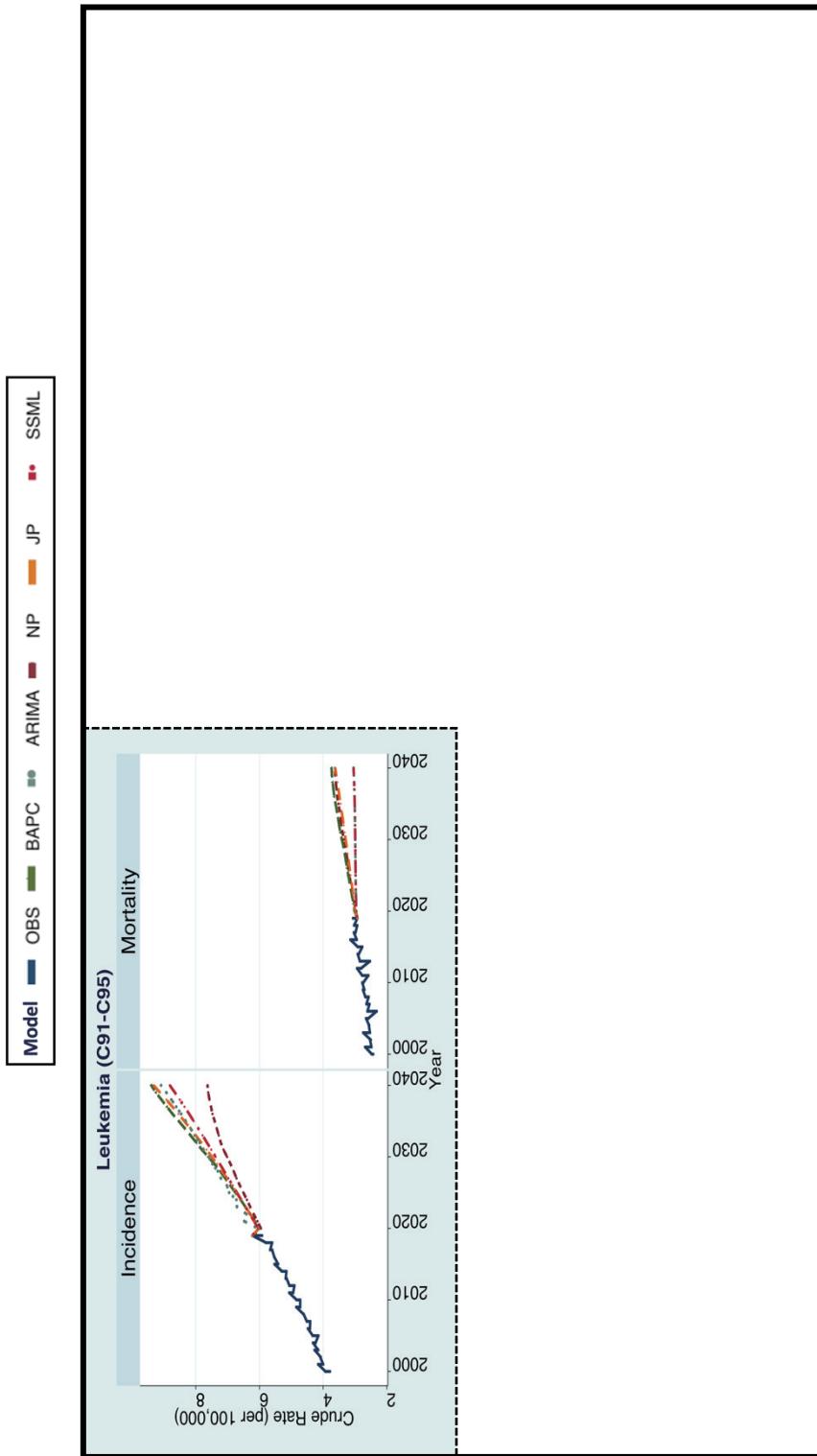


Figure 7-3. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using population structure-based models and non-population structure-based models (Leukemia)

4.5 Compare model characteristics based on outcomes by cancer type

Projection models, such as projection models, are typically evaluated using an assortment of methods. A conventional approach entails segmenting the held data into two subsets, namely observation and test years. The model's performance is subsequently gauged by quantifying the error ratio, i.e., the degree of correspondence between the model's predictions and the actual data from the test year, when the model is fit exclusively using observed data.

Despite its common usage, this method, which fundamentally uses historical data to assess the present, does present certain limitations when extended to forecast future data. Essentially, the uncertainty inherent in predicting future events based on past and present data calls into question the validity and reliability of such projections. Consequently, the absence of established guidelines or metrics to evaluate the future predictive value of these models compounds these challenges.

When it comes to NP models, they typically predict the most recent 10 years of data in the observations, incorporating an adjustment known as "cut-trend" to account for a 5-year update. This adjustment considers the rate of increase or decrease by assigning weights to each update, aiming to capture a trend like the actual trend [6,14,41,42]. However, one drawback of this model is that it relies solely on the most recent 10 years of data, which may result in a poorer fit when

the data pattern is irregular compared to models utilizing the complete dataset. The BAPC model, which is derived from the age-period-cohort model, addresses Maximum Likelihood Estimation (MLE) by constraining one variable based on the Bayesian method, specifically the Markov Chain Monte Carlo (MCMC) method, to approximate the actual value [43,45,48]. This model has been reported to offer a realistic representation of age-specific patterns of change when long-term trend data are available.

However, if there are numerous periods with missing values or if there is a significant increase or decrease in value over the observed period, the BAPC model may overestimate future trends. It also tends to adjust age-specific rates to match the trend of age-standardized rates, potentially resulting in overestimation or underestimation in specific age groups.

On the other hand, both the ARIMA and SSML models are derived from time series models and primarily focus on predicting the trend of age-standardized rates, considering only the total number of cases or deaths. These models utilize the entire observation period as data, which may lead to challenges in capturing irregular patterns of increase or decrease or sudden changes in observations [59,60,61].

The JP model, often used for short-term predictions, adopts an approach that identifies the joinpoint of the last period and selects the corresponding model to

predict future trends [15,58]. However, due to its reliance on an exponential function, this model tends to overestimate compared to other models when making long-term predictions.

Moreover, as the model incorporates the year variable in the data, it does not account for the age structure, limiting its use in conducting decomposition analysis to examine age effects as seen in time series models. These features suggest that the models have common characteristics in predicting cancer. In the case of the NP model, regardless of whether there is an increase or decrease in the year, if the incidence or mortality trend in the last 10 years decreases, the long-term trend also decreases, and if it increases, the trend increases.

The volatility of the trend was observed to be either increasing or decreasing in a continuous direction. BAPC is sensitive to sudden increases or decreases in observations. Therefore, for cancers with originally low incidence and mortality rates, BAPC estimates similar results to the NP model, but when there is a sudden change in the trend, the change in the estimate is very large (e.g., prostate cancer).

This tendency leads to underestimation of low observations and overestimation of high observations, so caution should be exercised for cancers with sudden increases or decreases. For ARIMA models, trends are calculated based on the entire observation period. In this model, the results are like NP for cancers with a steady increase or decrease. However, when adjusting for noise, the long-term

trend shows an irregular pattern. In addition, there are some cancer types that tend to be difficult to provide realistic observations because the phenomenon that the predicted value becomes constant in the future if the pattern of observations is irregular is also observed (leukemia).

The SSML model produces predictions like ARIMA, but it may differ from the realistic part because even if the incidence or mortality rate shows a decrease in cancer types with low incidence or mortality rate, it tends to increase in the long-term trend (Cervix Uteri). In the case of the JP model, as mentioned earlier, the trend is based on the last observed period, so if the period for the last inflection point is 2018-2019, there is a tendency to overestimate because the model is constructed based on an exponential function to estimate the model for that period. This is especially true for biliary tract and pancreas cancers, where there is a sharp increase, but also for prostate and lung cancers. The largest difference was for esophageal cancer in males, which is likely due to the large slope of the model for the last joinpoint period.

While there are no definitive criteria for selecting the most appropriate projection model for each cancer type, empirical experience suggests that APC models, particularly the NP model, have been widely used in recent studies. Although there is a lack of mathematical or objective evidence to establish the perfection or validity of the NP model for every cancer type, researchers have generally been

satisfied with the results obtained when applying the model according to their specific studies. It is important to note that the NP model is commonly employed in the context of comparing multiple cancer types simultaneously. While there may be variations in its performance when applied to different cancer types, empirical experience suggests that it provides a reasonable basis for such comparisons to evaluate the future predictive value of these models compounds these challenges [50,52].

4.6 Summary

This chapter compares different models used for the prediction of long-term incidence and mortality rates for various cancer types by 2040.

Lung cancer prediction trends exhibited similar patterns across models. However, randomwalk2 characteristics caused the predictions between the two models to diverge over time. The models' predictions for stomach and liver cancer also differed, as trends of increase and decrease fluctuated over time. For colorectal cancer, the NP model forecasted taking recent deceleration trends into account, while the BAPC model predicted a steady increase based on past trends, resulting in a widening gap in predictions. A similar pattern was observed with pancreas cancer, with discrepancies in predictions gradually expanding over time.

The JP model demonstrated variation in prediction model slopes, allowing for the observation of linearity or exponential changes. For gastric and liver cancer, a linear trend was seen as the inflection point was distant from recent years. The variance in mortality across models was generally less pronounced, with exceptions for liver, gallbladder, and pancreas cancers. Comparison of male incidence and mortality rates revealed a significant difference in long-term predictions between the NP model and the BAPC model, particularly for cancers with large increase or decrease ranges. Lung cancers showed the most consistent predictions across both models. The SSML model, due to its linear equations, faced

limitations in representing trends of increase or decrease in recent years. ARIMA, however, could reflect the recent trend to some extent, depending on the model selection. The JP model's slope for future prediction is set based on the recent JP, leading to a wide prediction range in certain cancers. The BAPC model was found to overestimate, particularly for cancers with irregular mortality rates, while the NP model showed a more pronounced slowdown with increase in the forecast year.

While most models predict cancer trends with some degree of success, their effectiveness varies depending on the specifics of the data and the cancer type being examined. The NP model, the BAPC model, the ARIMA and SSML models, and the JP model each have their strengths and limitations. While no model is universally superior, the most appropriate model for a specific cancer type needs to be chosen based on empirical evidence and the specifics of the data available. Based on many excerpted papers, the NP model is a good choice for universal use.

CHAPTER V. FUTURE PREDICTIONS ACCORDING TO POPULATION AGING SCENARIOS

5.1 Introduction

In the previous chapter, we employed the median population as a basis for making predictions when comparing different models [69,70,71]. Numerous studies have prioritized median values when evaluating the performance of models and comparing their predictive power. However, given the inherent uncertainty of future projections, it is essential to consider a wider range of factors. Consequently, this study will not only compare models but also incorporate additional scenarios to account for future uncertainty and provide more robust information on cancer incidence and mortality through scenario-specific predictive values. In contrast to models used in earlier analyses, the NP model utilized in this study demonstrated a superior ability to consistently estimate the contribution of incidence or mortality effects due to irregular periodicity of observations and demographic changes. Among the various models examined, the NP model was that appeared to have fewer overestimations or underestimations while maintaining a consistent trend across all cancer types. Consequently, this study has adopted a selection of scenarios provided by the KOSIS and compared model-specific predictions through the NP model.

The primary objective of this study was to furnish more reliable information on cancer incidence and mortality rates by incorporating diverse scenarios to offset

future uncertainties. This was achieved by going beyond the comparison of model predictions using median population values and offering scenario-specific predictions. By doing so, the study contributes to a more comprehensive understanding of the potential future landscape of cancer incidence and mortality, thereby supporting well-informed decision-making processes in cancer prevention and control.

5.2 Scenarios setting

In constructing the scenarios for this study, a suitable scenario was chosen based on the criteria provided by KOSIS. Generally, to establish a confidence interval for future population estimates, the range of errors is recognized using median, high, and low scenarios [72]. Consequently, both basic high and low scenarios were incorporated in our analysis. Additionally, given the importance of examining the impact of aging on cancer incidence and mortality rates, we assessed how the actual population structure change would differ depending on the rate of aging. Three key assumptions are required when constructing these scenarios, representing fertility, life expectancy, and international migration. The population inflow and outflow are calculated annually through fertility and life expectancy, while the inflow and outflow of international movement are additionally considered, altering the scenario's type according to the household level corresponding to each factor. In this study, fertility rate and life expectancy were adjusted to reflect the scenario according to the rate of aging, and the assumption

of international movement was also modified to significantly represent the difference in value (Table8).

Table 8. Key scenarios for prospective population projections

Category	Scenario Name	Hypothesis Setting		
		Birth Rate	Life Expectancy	International Migration
Base	Median	Median	Median	Median
	High	High	High	High
	Low	Low	Low	Low
Combination	Fast Ageing	Low	High	Low
	Slow Ageing	High	Low	High

5.3 Applying decomposition analysis in future projection

The decomposition analysis reflects changes according to the population structure, specifically the distribution by age, and indicates the contribution to the occurrence or death according to the distribution by age. In general, after calculating the difference between the occurrence or death based on the base year, it indicates how much the difference contributes to forming the difference by the factor depending on the factor being decomposed.

This allows for the specification of a base year, and the visualization of how much it contributes to the difference between the base year and the current year in performing the analysis. Four indicators can be found in this decomposition analysis.

The first is net change, which simply refers to the value of the current year minus the base year's occurrence or death. The second is Population Aging Risk, which primarily addresses the aging impact in this study. This factor refers to the effect that the population growth in the elderly contributes to cancer incidence or death. The third is Population Growth Risk, which shows the effect of population increase or decrease in all age groups contributing to occurrence or death. The last represents an incidence rate risk or a mortality rate risk, commonly referred to as a rate risk. This factor represents the incidence or mortality rate by age and refers to the effect of contributing to the occurrence or death when the incidence or mortality rate by age changes.

The decomposition analysis method of Das Gupta is primarily used in demographic studies. However, this method has the disadvantage of not being able to show robust results as the demographic changes, as it does not consider interactions for three factors. In the case of the previous Bashir method, the method was simplified, showing an advantage in terms of computational cost.

However, bias can occur because the interaction is not considered. Cheng's method is a decomposition analysis algorithm that combines the advantages of the das Gupta and Bashir methods to reduce computational costs while minimizing bias. By incorporating interaction effects among the three factors, Cheng's method provides robust and accurate results that can be efficiently applied to large datasets.

Therefore, as mentioned in Chapter 2, we will conduct the decomposition analysis using an algorithm that corrects the bias of the model while reducing the computational cost using the Cheng's method.

5.4 Comparison of projection according to prospective population projection scenarios

5.4.1 Comparison male incidence and mortality rates by scenarios

Figure 8 presents the comparative results of estimated population by scenario for male cancer incidence. In the case of men, across all carcinoma types, the incidence was higher in the scenario with the fastest aging rate and lower in the scenario with the slowest aging rate. Specifically, for gastric cancer, colorectal cancer, and liver cancer, the incidence rate varied considerably according to the scenario. In the case of gastric cancer, it was observed that the incidence rate was approximately five individuals higher in the scenario with the fastest aging rate based on the median population by 2040. In contrast, compared to the scenario with the slowest aging rate, the difference of up to ten individuals suggests a potential reduction in the absolute incidence of gastric cancer by decelerating the rate of aging entry. For other cancer types, the increase or decrease in incidence by scenario revealed an increase of up to 5% in cases of fast aging based on the median population and a decrease of up to 5-7% in cases of slow aging (Table9-1).

Analogous to the incidence rates, the mortality rates were higher in the scenario with the fastest aging rate and lower in the scenario with the slowest aging rate. The differences in the scenarios were relatively smaller compared to the incidence rates, with an average increase or decrease rate of 5% (Table9-2).

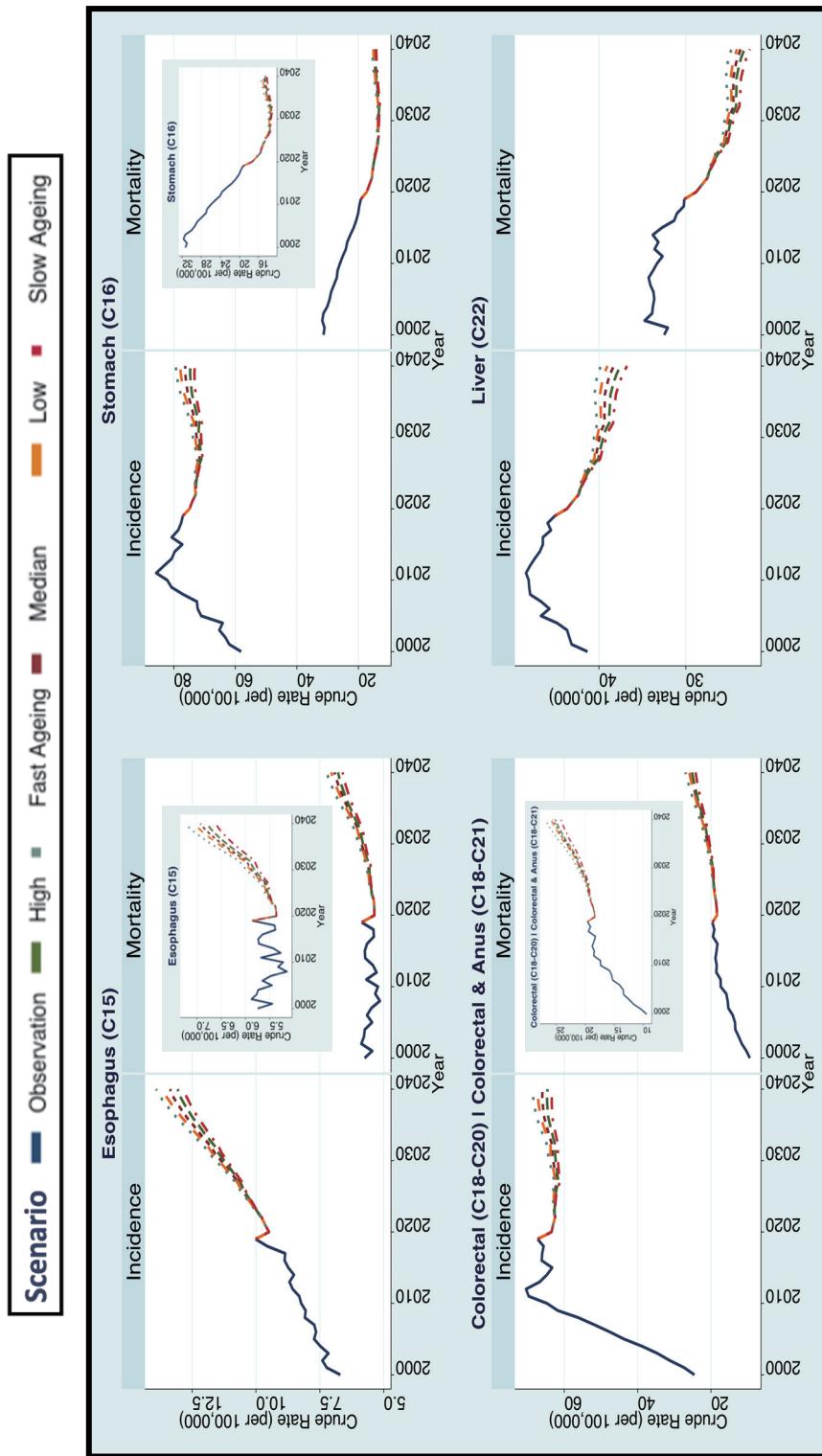


Figure 8-1. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Esophagus, Stomach, Colorectal & Anus, Liver Cancer)

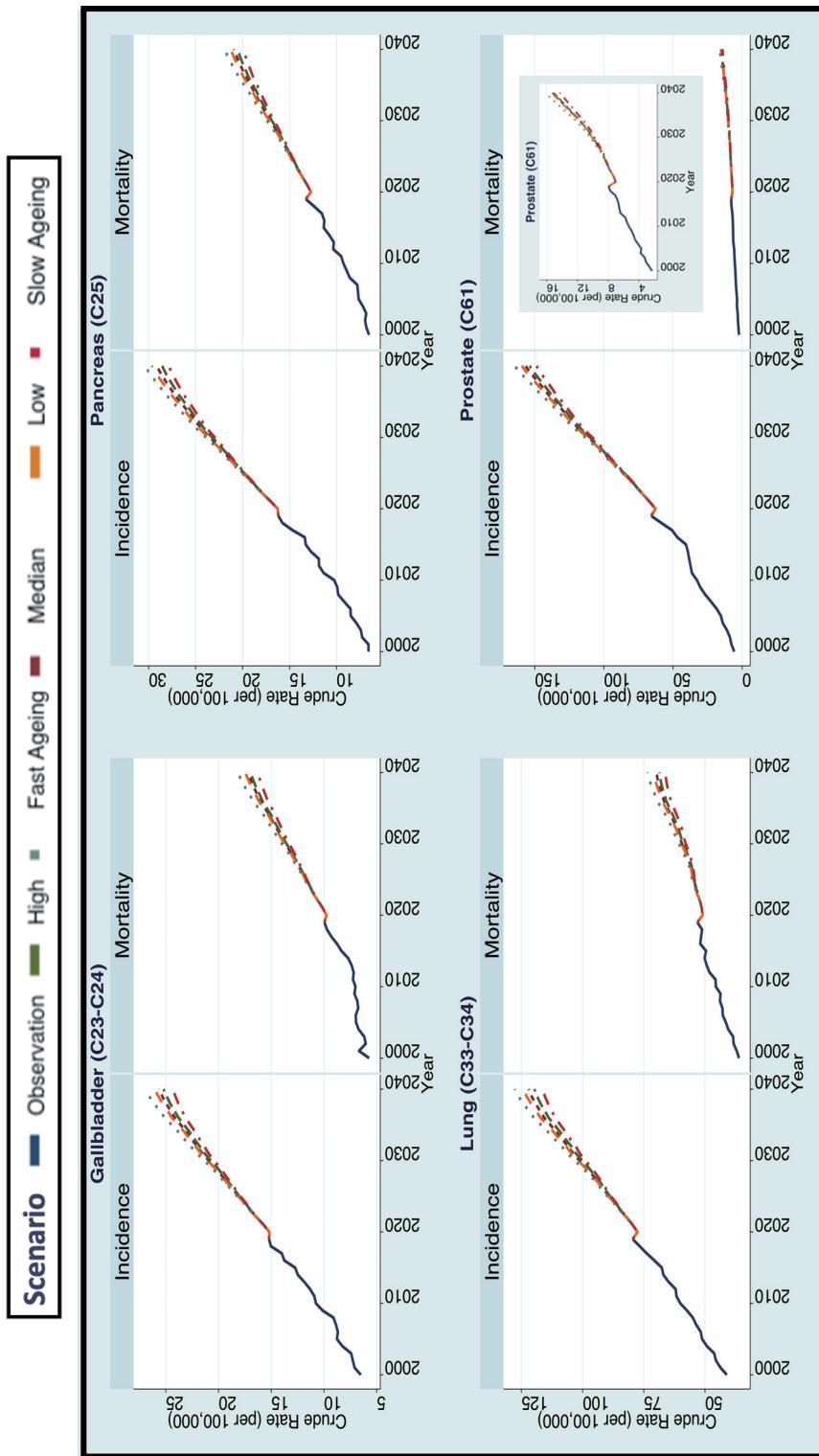


Figure 8-2. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer)

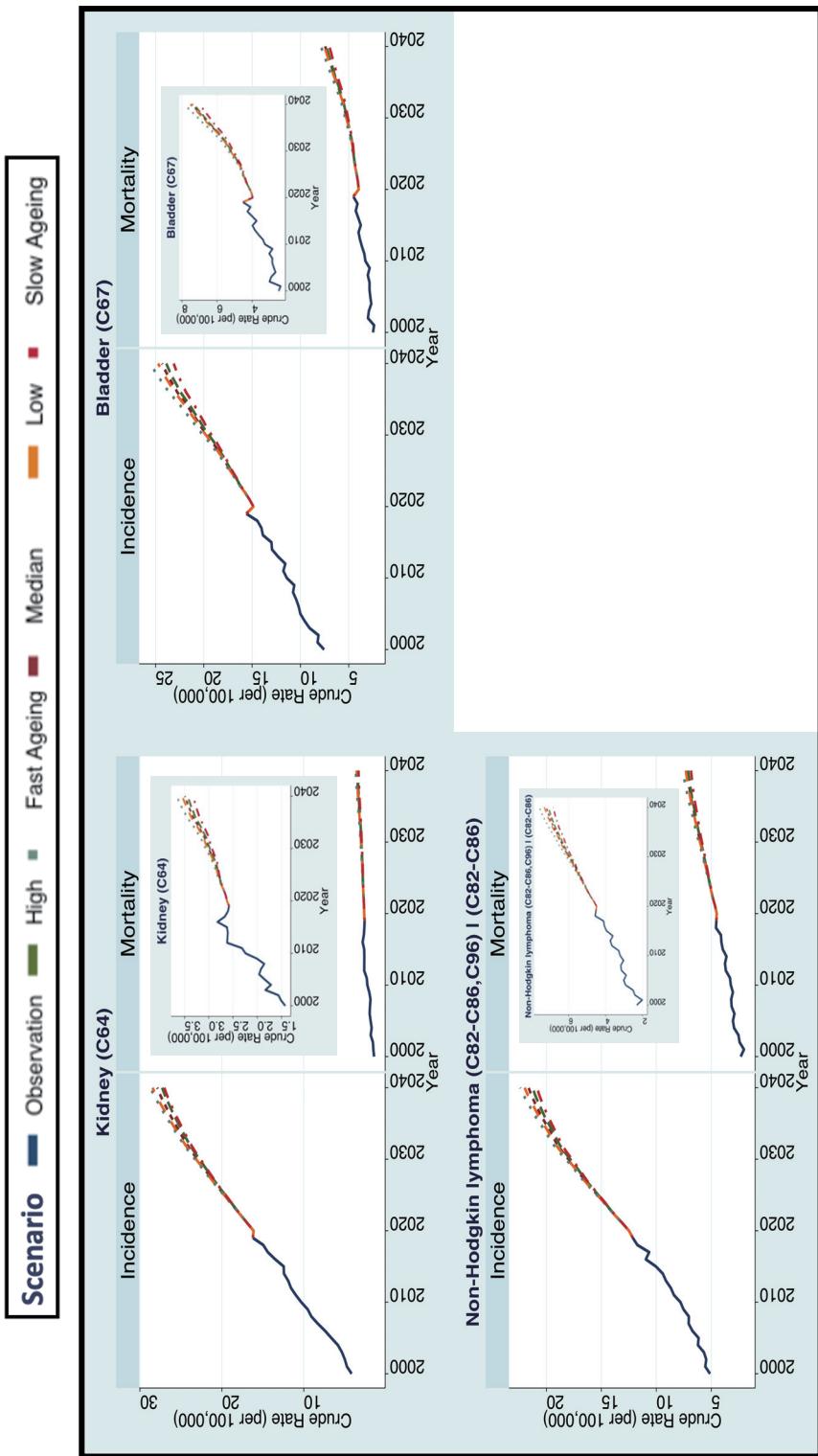


Figure 8-3. Projection for male age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Kidney, Bladder Cancer, Non-Hodgkin Lymphoma)

Table 9-1. Male incidence rate of each cancer type by year by scenarios

Cancer	Scenarios				
	Median	High	Low	Fast Ageing	Slow Ageing
Esophagus (C15)					
2019	10.0	10.0	10.0	10.0	10.0
2030	11.4	11.2	11.5	11.6	11.1
2040	13.4	13.1	13.6	13.9	12.8
Stomach (C16)					
2019	77.1	77.1	77.1	77.1	77.1
2030	72.6	71.8	73.1	73.8	71.1
2040	76.3	74.6	77.9	79.4	73.2
Colorectal (C18-C20)					
2019	67.3	67.3	67.3	67.3	67.3
2030	63.0	62.3	63.4	64.1	61.7
2040	65.9	64.5	67.2	68.6	63.2
Liver (C22)					
2019	45.1	45.1	45.1	45.1	45.1
2030	39.9	39.5	40.2	40.6	39.2
2040	38.3	37.5	39.0	39.8	36.8
Gallbladder (C23-C24)					
2019	15.2	15.2	15.2	15.2	15.2
2030	20.9	20.7	21.0	21.4	20.4
2040	25.7	25.2	26.1	27.0	24.3
Pancreas (C25)					
2019	16.2	16.2	16.2	16.2	16.2
2030	23.5	23.2	23.6	23.9	23.0
2040	29.1	28.6	29.7	30.4	27.9
Lung (C33-C34)					
2019	79.5	79.5	79.5	79.5	79.5
2030	101.2	100.2	101.8	103.2	98.8
2040	122.0	119.7	124.3	127.8	116.3
Prostate (C61)					
2019	65.8	65.8	65.8	65.8	65.8
2030	111.2	110.1	111.9	113.5	108.5
2040	156.7	153.7	159.4	164.6	148.7
Kidney (C64)					
2019	16.2	16.2	16.2	16.2	16.2
2030	23.0	22.7	23.1	23.3	22.6
2040	27.9	27.3	28.5	28.7	27.0
Bladder (C67)					
2019	15.6	15.6	15.6	15.6	15.6
2030	19.7	19.5	19.8	20.1	19.2
2040	24.3	23.9	24.7	25.5	23.1
Non-Hodgkin lymphoma (C82-C86,C96)					
2019	12.1	12.1	12.1	12.1	12.1
2030	18.0	17.8	18.1	17.7	18.2
2040	21.6	21.2	22.0	20.8	22.4

Table 9-2. Male mortality rate of each cancer type by year by scenarios

Cancer	Scenarios				
	Median	High	Low	Fast Ageing	Slow Ageing
Esophagus (C15)					
2019	5.9	5.9	5.9	5.9	5.9
2030	5.9	5.8	5.9	6.0	5.7
2040	6.9	6.8	7.1	7.2	6.6
Stomach (C16)					
2019	19.2	19.2	19.2	19.2	19.2
2030	13.6	13.5	13.7	13.9	13.3
2040	14.8	14.5	15.1	15.5	14.2
Colorectal & Anus (C18-C21)					
2019	19.6	19.6	19.6	19.6	19.6
2030	20.9	20.7	21.0	21.4	20.4
2040	25.6	25.2	25.9	27.0	24.2
Liver (C22)					
2019	30.2	30.2	30.2	30.2	30.2
2030	25.1	24.8	25.2	25.5	24.5
2040	23.7	23.2	24.1	24.7	22.6
Gallbladder (C23-C24)					
2019	10.0	10.0	10.0	10.0	10.0
2030	13.7	13.6	13.7	14.0	13.3
2040	17.2	17.0	17.5	18.2	16.3
Pancreas (C25)					
2019	13.3	13.3	13.3	13.3	13.3
2030	17.1	16.9	17.2	17.4	16.7
2040	20.9	20.5	21.3	21.9	19.9
Lung (C33-C34)					
2019	53.1	53.1	53.1	53.1	53.1
2030	59.3	58.8	59.6	60.6	57.8
2040	69.8	68.6	70.9	73.4	66.2
Prostate (C61)					
2019	7.9	7.9	7.9	7.9	7.9
2030	10.1	10.0	10.1	10.4	9.8
2040	15.2	15.1	15.4	16.2	14.3
Kidney (C64)					
2019	2.6	2.6	2.6	2.6	2.6
2030	3.0	3.0	3.0	3.1	2.9
2040	3.5	3.4	3.5	3.7	3.3
Bladder (C67)					
2019	4.5	4.5	4.5	4.5	4.5
2030	5.2	5.2	5.2	5.4	5.1
2040	7.4	7.3	7.5	7.9	7.0
Non-Hodgkin lymphoma (C82-C86)					
2019	4.5	4.5	4.5	4.5	4.5
2030	6.0	5.9	6.0	6.1	5.8
2040	7.2	7.1	7.3	7.6	6.8

5.4.2 Comparison female incidence and mortality rates by scenarios

Figure 9 presents the long-term estimation results by scenario for female cancer incidence. In accordance with male cancer incidence, substantial variation across scenarios is observed for stomach and liver cancer. However, the absolute differences in incidence are relatively small, with variations within 5%. The most considerable differences are found in cancers other than stomach and liver, which display variations within 3-5% (the difference between fast and slow aging based on the median population) (Table10-1).

For stomach cancer, minimal variation across scenarios is observed, and as in the previous analysis, the variation in predictions across scenarios for liver cancer is notably large at 7%. Nonetheless, the low age-adjusted mortality rates and the ongoing decline in these cancers suggest that surveillance may require less emphasis than for colorectal and pancreatic cancers, which continue to increase, even if the variation is small. For other cancers, the age-adjusted mortality rate has been steadily increasing, and due to the small age-adjusted mortality rate, it is improbable that a significant increase in deaths will be observed even if a long-term trend is entered, unless the increase or decrease becomes extreme. In the case of colorectal cancer, pancreatic cancer, and lung cancer, sufficiently high mortality rates are observed to warrant caution (Table10-2).

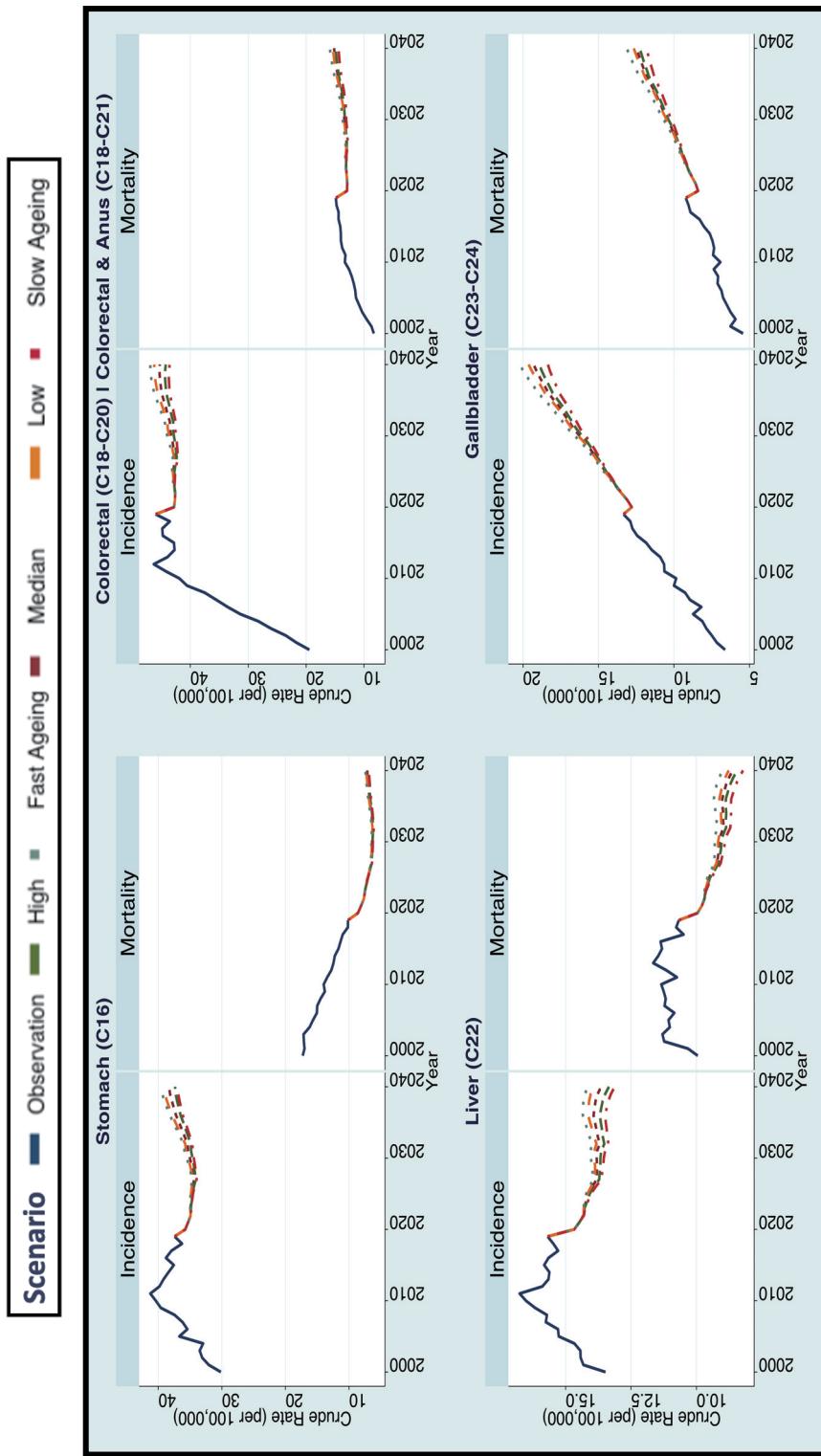


Figure 9-1. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer)

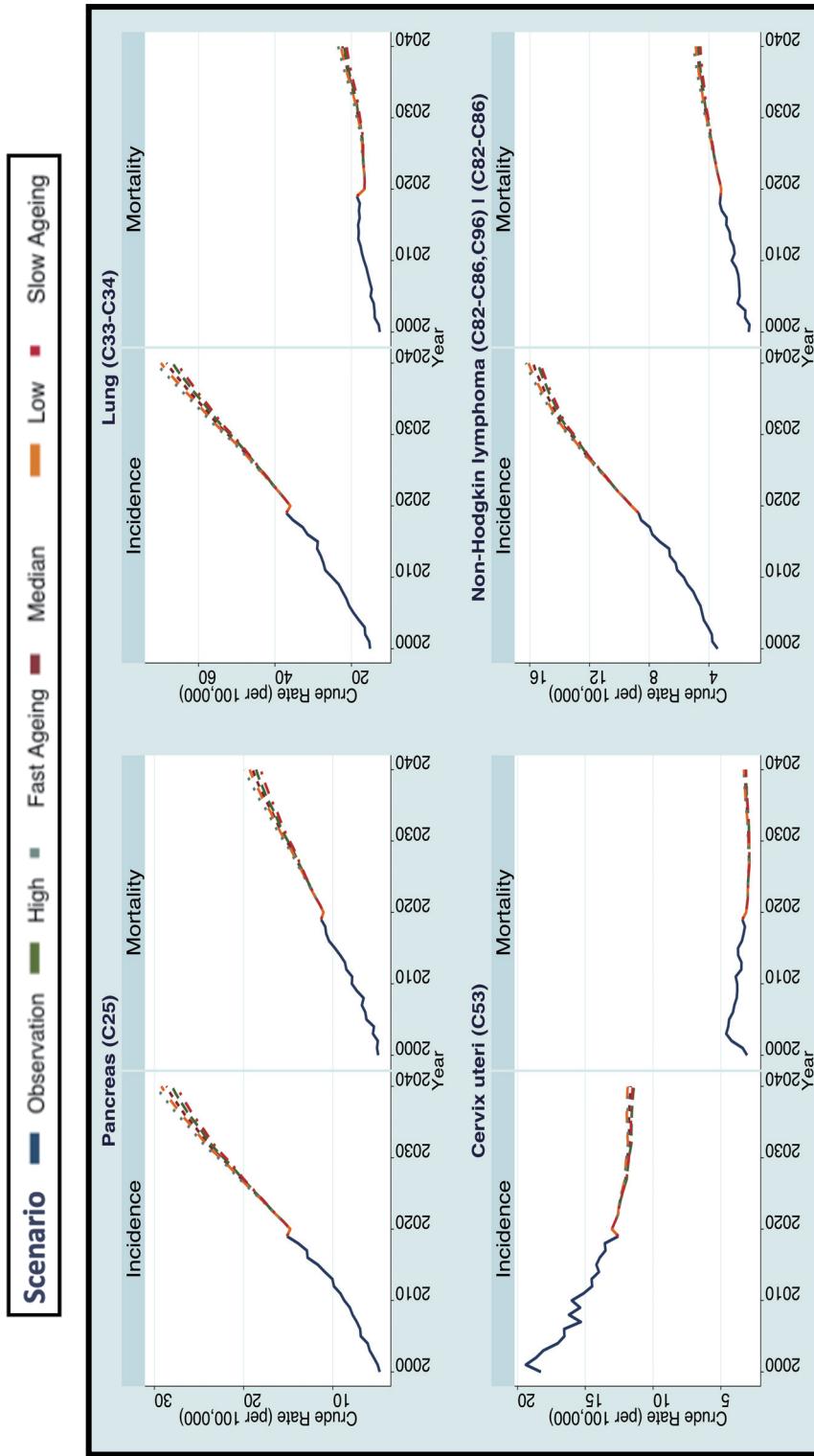


Figure 9-2. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Pancreas, Lung, Cervix Uteri, Cancer, Non-Hodgkin Lymphoma)

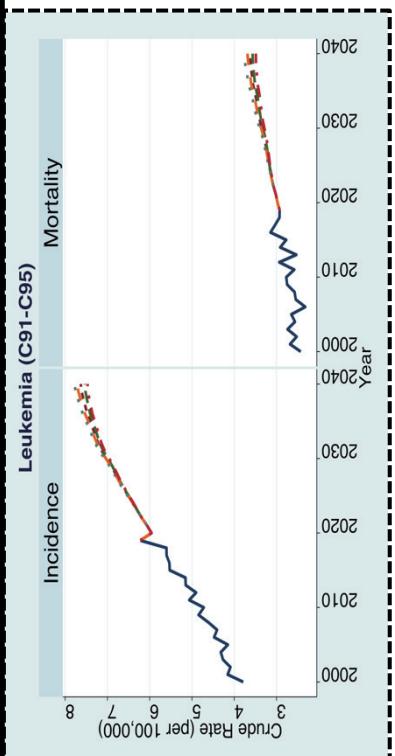


Figure 9-3. Projection for female age-adjusted incidence & mortality rates in 2000 to 2040 using NP model by Scenarios (Leukemia)

Table 10-1. Female incidence rate of each cancer type by year by scenarios

Cancer	Scenarios				
	Median	High	Low	Fast Ageing	Slow Ageing
Stomach (C16)					
2019	37.4	37.4	37.4	37.4	37.4
2030	35.0	34.6	35.3	35.5	34.4
2040	38.3	37.4	39.2	39.6	36.9
Colorectal (C18-C20)					
2019	46.0	46.0	46.0	46.0	46.0
2030	43.3	42.8	43.6	44.0	42.5
2040	45.3	44.2	46.2	46.9	43.6
Liver (C22)					
2019	15.7	15.7	15.7	15.7	15.7
2030	13.8	13.7	13.9	14.0	13.5
2040	13.6	13.3	13.9	14.2	13.1
Gallbladder (C23-C24)					
2019	13.3	13.3	13.3	13.3	13.3
2030	16.2	16.1	16.3	16.6	15.9
2040	19.3	18.9	19.6	20.2	18.4
Pancreas (C25)					
2019	15.1	15.1	15.1	15.1	15.1
2030	22.5	22.2	22.6	22.9	22.0
2040	28.6	28.0	29.2	29.8	27.5
Lung (C33-C34)					
2019	37.0	37.0	37.0	37.0	37.0
2030	52.3	51.7	52.7	53.1	51.3
2040	68.4	66.8	69.8	71.1	65.6
Cervix Uteri (C53)					
2019	12.6	12.6	12.6	12.6	12.6
2030	11.9	11.8	12.0	12.0	11.8
2040	11.6	11.4	11.9	11.8	11.5
Non-Hodgkin lymphoma (C82-C86,C96)					
2019	8.7	8.7	8.7	8.7	8.7
2030	13.3	13.1	13.3	13.4	13.1
2040	15.8	15.4	16.1	16.3	15.3
Leukemia (C91-C95)					
2019	6.2	6.2	6.2	6.2	6.2
2030	7.0	7.0	7.0	7.1	7.0
2040	7.6	7.5	7.7	7.8	7.5

Table 10-2. Female mortality rate of each cancer type by year by scenarios

Cancer	Scenarios				
	Median	High	Low	Fast Ageing	Slow Ageing
Stomach (C16)					
2019	10.2	10.2	10.2	10.2	10.2
2030	6.4	6.3	6.4	6.5	6.2
2040	7.1	7.0	7.3	7.4	6.9
Colorectal (C18-C20)					
2019	14.9	14.9	14.9	14.9	14.9
2030	13.4	13.3	13.5	13.7	13.1
2040	15.2	14.9	15.4	15.9	14.4
Liver (C22)					
2019	10.7	10.7	10.7	10.7	10.7
2030	9.1	9.0	9.2	9.3	8.9
2040	8.6	8.4	8.8	9.0	8.2
Gallbladder (C23-C24)					
2019	9.2	9.2	9.2	9.2	9.2
2030	10.5	10.4	10.5	10.7	10.2
2040	12.5	12.3	12.7	13.1	11.8
Pancreas (C25)					
2019	11.3	11.3	11.3	11.3	11.3
2030	15.2	15.0	15.3	15.5	14.8
2040	19.0	18.6	19.3	19.9	18.1
Lung (C33-C34)					
2019	18.6	18.6	18.6	18.6	18.6
2030	18.2	18.0	18.3	18.6	17.8
2040	22.4	21.9	22.8	23.5	21.3
Cervix Uteri (C53)					
2019	3.4	3.4	3.4	3.4	3.4
2030	3.0	2.9	3.0	3.0	2.9
2040	3.2	3.2	3.3	3.3	3.1
Non-Hodgkin lymphoma (C82-C86)					
2019	3.2	3.2	3.2	3.2	3.2
2030	4.1	4.1	4.2	4.2	4.1
2040	4.8	4.7	4.9	5.0	4.6
Leukemia (C91-C95)					
2019	2.9	2.9	2.9	2.9	2.9
2030	3.3	3.3	3.4	3.4	3.3
2040	3.6	3.6	3.7	3.8	3.5

5.5 Comparison of decomposition about projection according to prospective population projection scenarios

5.5.1 Comparison decompose male incidence and mortality by scenarios

Figures 10 and 11 display the decomposition of scenario-specific predicted values for cancer incidence and mortality in men, respectively. Since the differences in predicted values by scenario have already been acknowledged, no significant differences in the effects of the factors are observed, as no cancer type substantially increases or decreases by scenario. In terms of the pattern of effects by scenario for each cancer type, the effect of the factor diminishes after the projection time when the population growth rate increases in the median and low and fast aging scenarios. Conversely, in the high and slow aging scenarios, the contribution of population growth to incidence or mortality gradually increases.

Additionally, the contribution of the aging factor in the fast-aging scenario is higher than in the other scenarios. For male mortality patterns, the contribution of advanced aging factors varied not only across scenarios but also across specific cancer types. For esophagus, stomach, liver, and kidney cancers, the contribution of aging factors was largest in the moderate population scenario, while for the remaining cancers—colorectal, pancreas, lung, prostate, non-Hodgkin lymphoma—the contribution of aging factors was higher in the fast-aging scenario.

Like the percentage difference in projections in the previous scenario-specific projection model, the percentage difference in the corresponding aging factors varied by up to $\pm 10\%$ (Table11-1, Table11-2).

However, there was little difference in the proportion of net change in the predictions, and adjusting the scenarios did not seem to change the differences in the contributions of the internal factors (Table12-1, Table12-2). The rate is calculated as the difference between the increase attributable to each factor relative to the total number of cases or mortality case in the base year, which is 2000 for incidence cases and 2005 for mortality case.

In the case of men, the incidence and deaths showed an increasing trend in the long term, except for liver cancer, when the aging factor and net change were considered. In the case of liver cancer, the number of incidents or deaths tended to decrease more than the baseline year as the long-term trend entered the long term.

However, the aging factor showed a steadily increasing trend, which is thought to be the result of a steady decrease in the incidence or deaths in the middle-aged group and a gradual increase in the contribution of the elderly group. Major cancers continue to show a high level of increase. In the case of prostate cancer, both incidence and deaths are showing very large increases, and the contribution of aging factors is almost 15 times higher in terms of incidence compared to the baseline year, and more than 4 times higher in terms of deaths. The net change is

similarly high, suggesting that preparing for prostate cancer in the elderly is an essential requirement for the long-term projection.

There were some cancers where the mortality rate was consistently decreasing compared to the baseline year. Stomach and liver cancers showed a decrease in the number of deaths as they long term projection period, with stomach cancer showing the largest decrease in deaths. However, the impact of aging continues to increase, which can be understood as a shift in the age group of the main deaths from the middle-aged to the elderly.

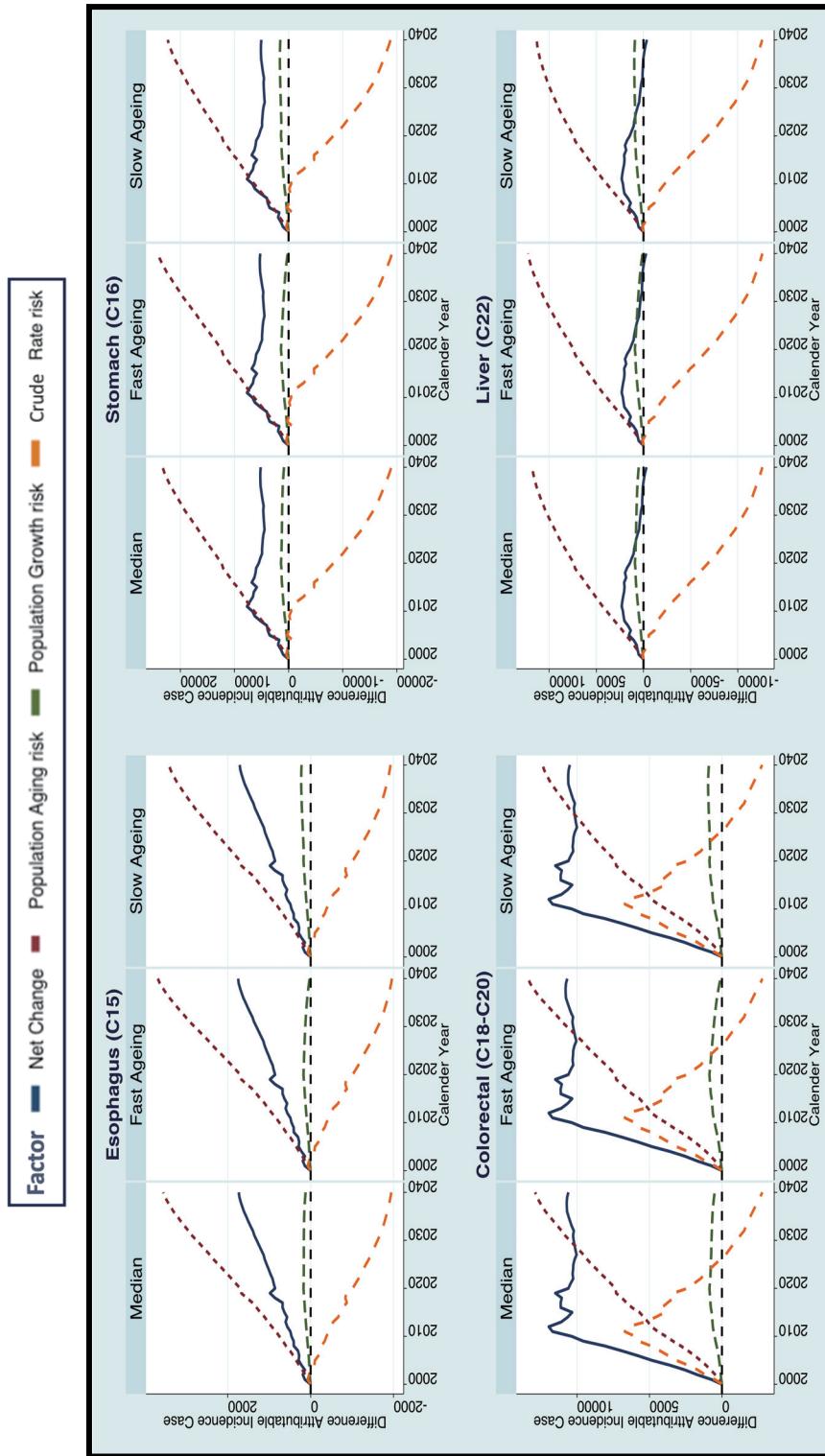


Figure 10-1. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Esophagus, Stomach, Stomach, Colorectal, Colorectal & Anus, Liver Cancer)

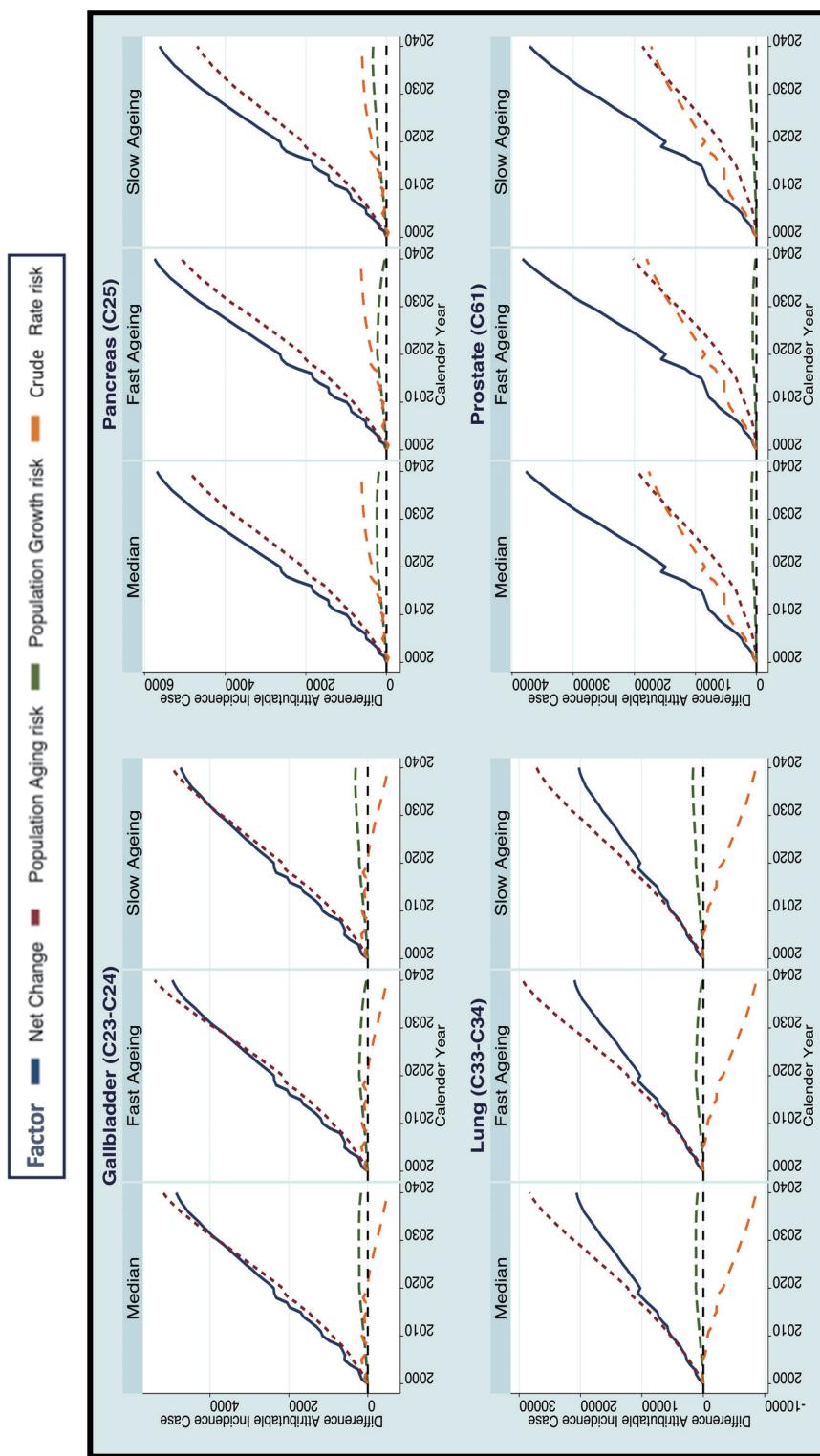


Figure 10-2. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing, Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer)

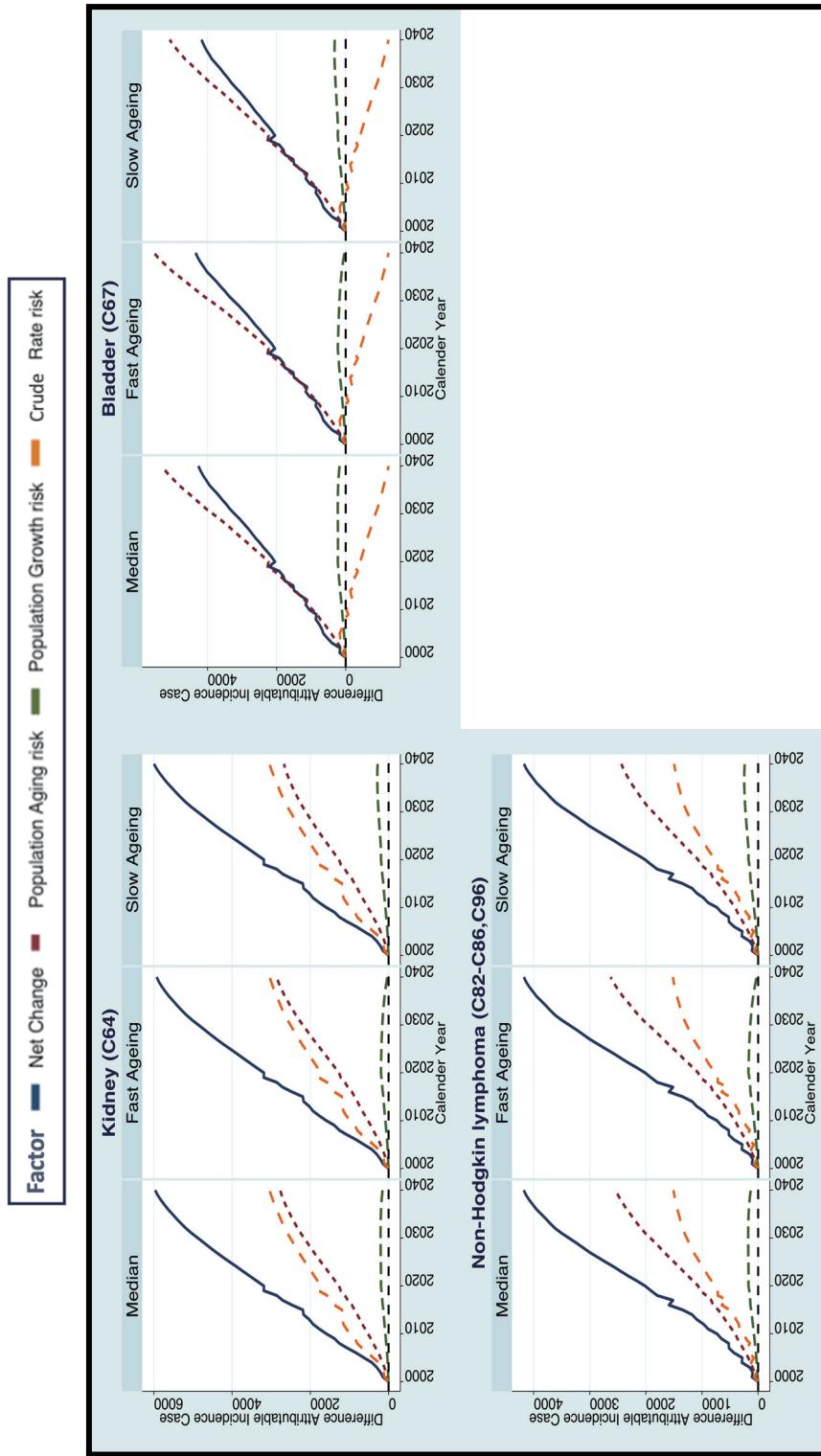


Figure 10-3. Comparison decomposition for male incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Kidney, Bladder Cancer, Non-Hodgkin Lymphoma)

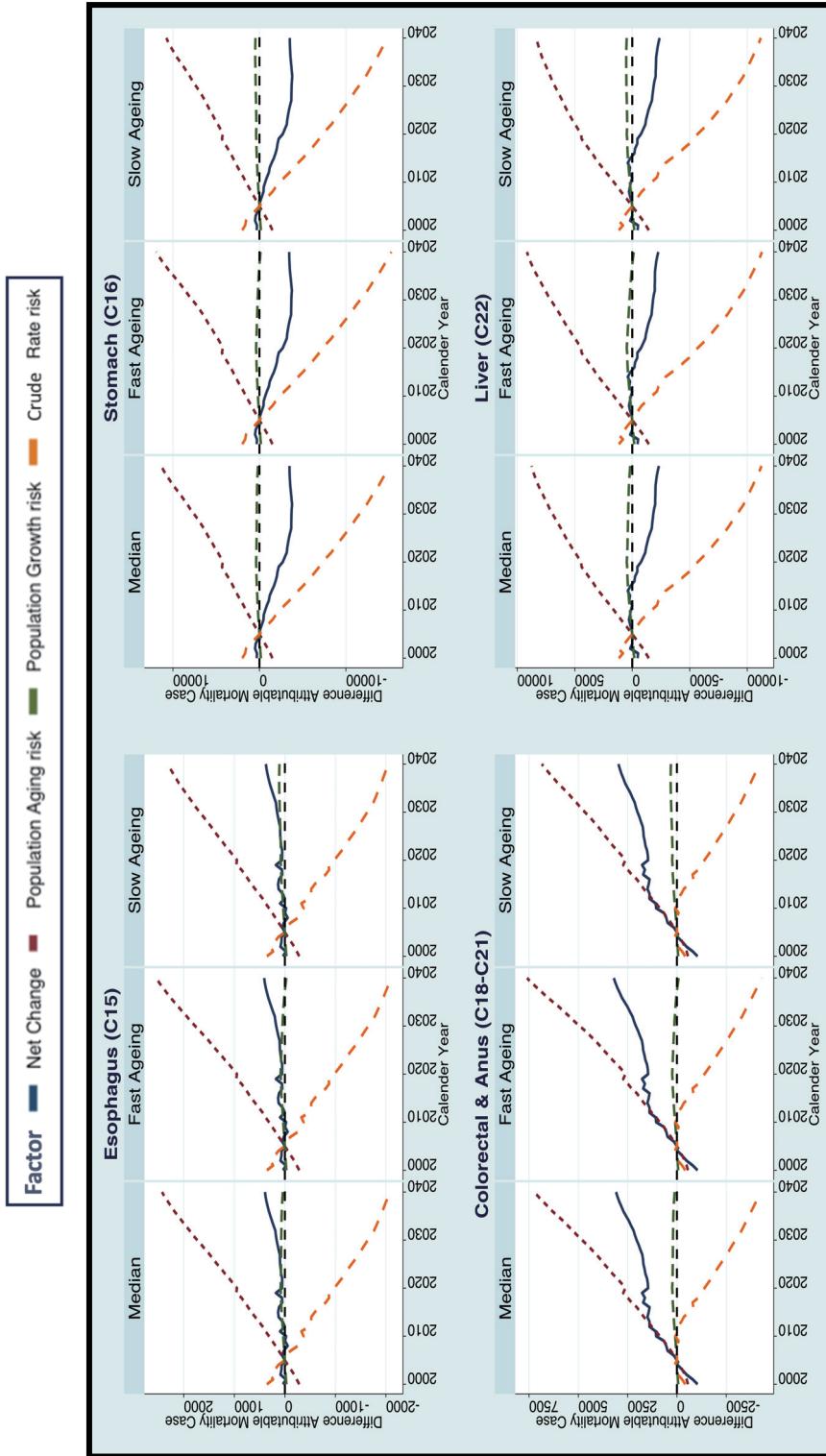


Figure 11-1. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Esophagus, Stomach, Stomach, Colorectal & Anus, Liver Cancer)

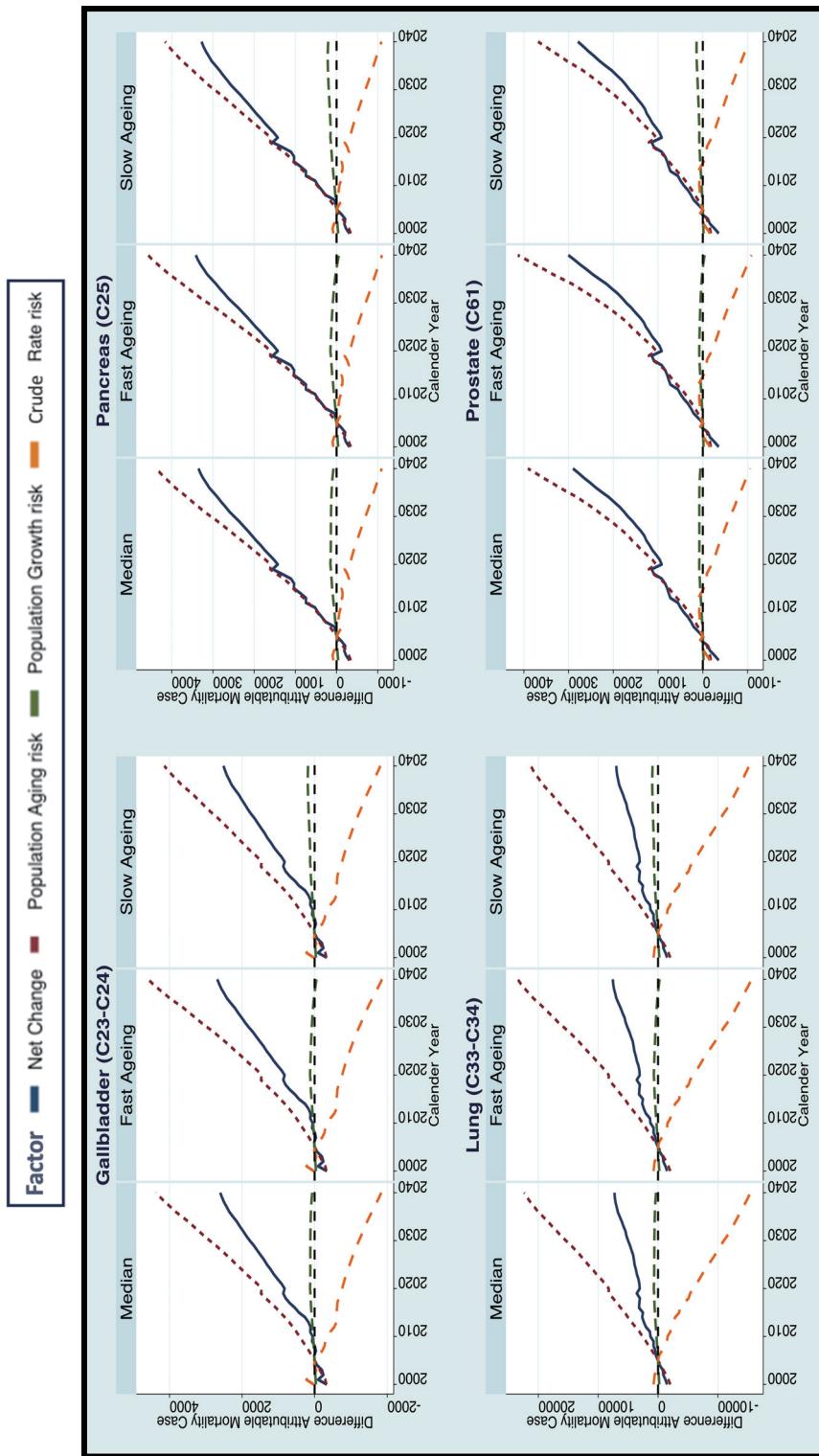


Figure 11-2. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer)

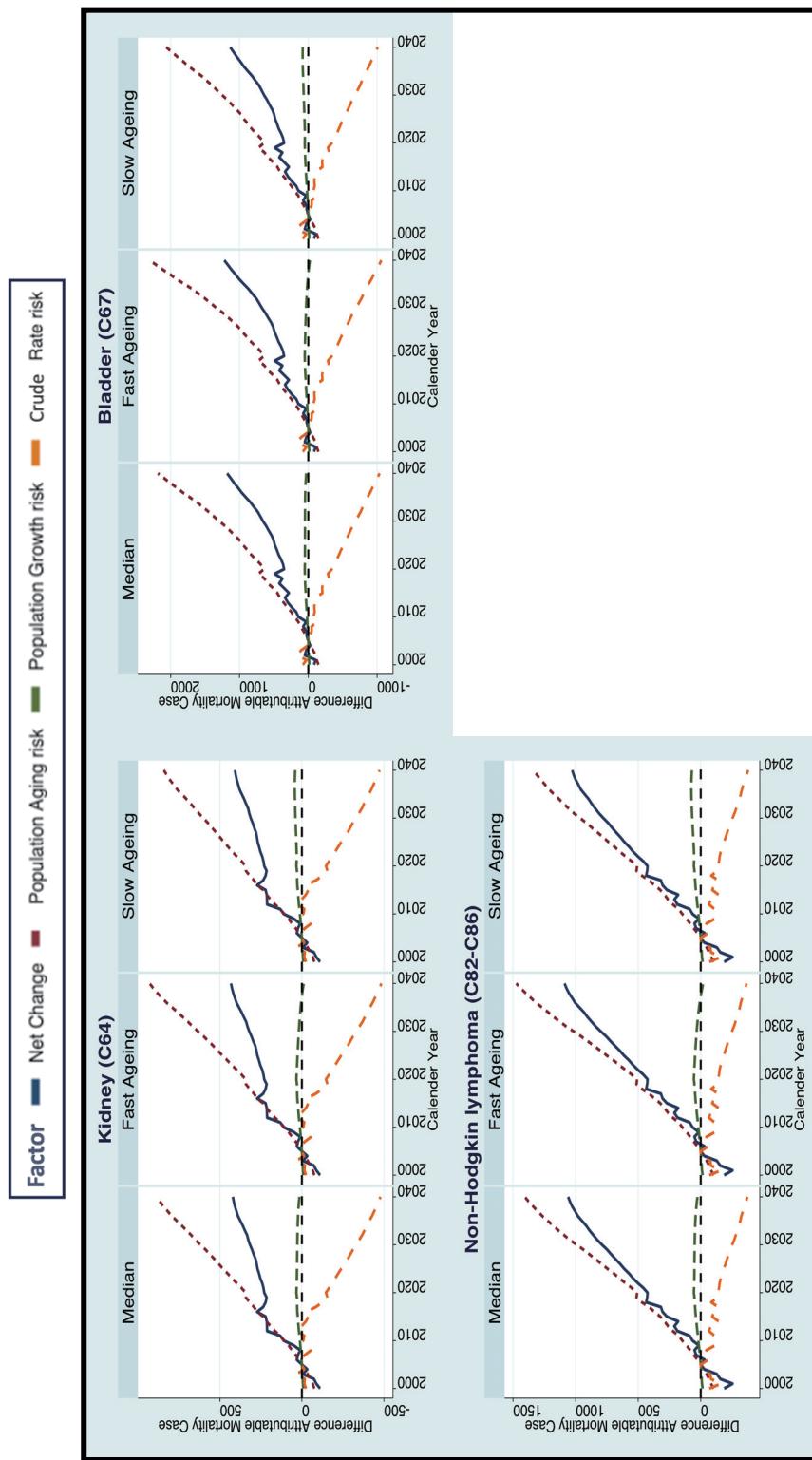


Figure 11-3. Comparison decomposition for male mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Gallbladder, Pancreas, Lung, Prostate Cancer)

Table 11-1. Male incidence case difference based on population ageing risk

Cancer	Scenarios (%) (Ref 2000)		
	Median	Fast Ageing	Slow Ageing
Esophagus (C15)			
2019	105	105	105
2030	172	175	168
2040	222	231	214
Stomach (C16)			
2019	87	87	87
2030	134	136	131
2040	168	174	161
Colorectal (C18-C20)			
2019	125	125	125
2030	179	182	175
2040	220	228	211
Liver (C22)			
2019	74	74	74
2030	103	105	101
2040	119	124	115
Gallbladder (C23-C24)			
2019	136	136	136
2030	244	249	238
2040	334	347	319
Pancreas (C25)			
2019	127	127	127
2030	230	234	225
2040	310	322	298
Lung (C33-C34)			
2019	123	123	123
2030	214	218	209
2040	289	300	277
Prostate (C61)			
2019	438	438	438
2030	924	941	904
2040	1421	1475	1364
Kidney (C64)			
2019	123	123	123
2030	217	221	213
2040	279	288	269
Bladder (C67)			
2019	125	125	125
2030	213	217	208
2040	295	308	283
Non-Hodgkin lymphoma (C82-C86,C96)			
2019	84	84	84
2030	157	160	154
2040	207	215	199

Table 11-2. Male mortality case difference based on population ageing risk

Cancer	Scenarios (%) (Ref 2005)		
	Median	Fast Ageing	Slow Ageing
Esophagus (C15)	Median	Fast Ageing	Slow Ageing
	2019	72	72
	2030	130	133
	2040	182	191
Stomach (C16)	Median	Fast Ageing	Slow Ageing
	2019	60	60
	2030	105	108
	2040	158	166
Colorectal & Anus (C18-C21)	Median	Fast Ageing	Slow Ageing
	2019	83	83
	2030	146	150
	2040	219	231
Liver (C22)	Median	Fast Ageing	Slow Ageing
	2019	54	54
	2030	84	86
	2040	106	111
Gallbladder (C23-C24)	Median	Fast Ageing	Slow Ageing
	2019	87	87
	2030	170	174
	2040	257	270
Pancreas (C25)	Median	Fast Ageing	Slow Ageing
	2019	86	86
	2030	163	167
	2040	234	244
Lung (C33-C34)	Median	Fast Ageing	Slow Ageing
	2019	81	81
	2030	152	155
	2040	221	231
Prostate (C61)	Median	Fast Ageing	Slow Ageing
	2019	133	133
	2030	244	236
	2040	433	407
Kidney (C64)	Median	Fast Ageing	Slow Ageing
	2019	76	76
	2030	137	140
	2040	196	206
Bladder (C67)	Median	Fast Ageing	Slow Ageing
	2019	107	107
	2030	191	196
	2040	322	340
Non-Hodgkin lymphoma (C82-C86)	Median	Fast Ageing	Slow Ageing
	2019	69	69
	2030	134	137
	2040	190	199

Table 12-1. Net change by male incidence case

Cancer	Scenarios (%) (Ref 2000)		
Esophagus (C15)	Median	Fast Ageing	Slow Ageing
2019	61	61	61
2030	81	82	81
2040	108	109	107
Stomach (C16)			
2019	43	43	43
2030	33	34	33
2040	37	38	36
Colorectal (C18-C20)			
2019	196	196	196
2030	175	175	174
2040	181	182	179
Liver (C22)			
2019	18	18	18
2030	3	3	3
2040	-3	-3	-4
Gallbladder (C23-C24)			
2019	152	152	152
2030	244	246	241
2040	312	318	305
Pancreas (C25)			
2019	165	165	165
2030	280	282	278
2040	361	365	357
Lung (C33-C34)			
2019	109	109	109
2030	163	165	162
2040	210	214	207
Prostate (C61)			
2019	1140	1140	1140
2030	1975	1988	1961
2040	2759	2800	2713
Kidney (C64)			
2019	321	321	321
2030	490	489	491
2040	600	597	603
Bladder (C67)			
2019	123	123	123
2030	178	180	176
2040	236	241	231
Non-Hodgkin lymphoma (C82-C86,C96)			
2019	156	156	156
2030	276	276	276
2040	342	342	341

Table 12-2. Net change by male mortality case

Cancer	Scenarios (%) (Ref 2005)		
	Median	Fast Ageing	Slow Ageing
Esophagus (C15)			
2019	13	13	13
2030	12	13	12
2040	30	31	28
Stomach (C16)			
2019	-31	-31	-31
2030	-52	-51	-52
2040	-48	-48	-49
Colorectal & Anus (C18-C21)			
2019	54	54	54
2030	62	63	61
2040	94	97	90
Liver (C22)			
2019	-5	-5	-5
2030	-22	-22	-23
2040	-28	-28	-29
Gallbladder (C23-C24)			
2019	51	51	51
2030	105	107	103
2040	153	158	148
Pancreas (C25)			
2019	83	83	83
2030	133	134	131
2040	179	182	175
Lung (C33-C34)			
2019	35	35	35
2030	49	51	48
2040	72	75	69
Prostate (C61)			
2019	126	126	126
2030	185	189	181
2040	321	333	309
Kidney (C64)			
2019	48	48	48
2030	70	71	69
2040	93	96	91
Bladder (C67)			
2019	72	72	72
2030	97	99	95
2040	174	180	167
Non-Hodgkin lymphoma (C82-C86)			
2019	58	58	58
2030	107	108	105
2040	143	147	139

5.5.2 Comparison decompose female incidence and mortality by scenarios

Figures 12 and 13 presents the decomposition of the predicted values by cancer incidence and mortality scenarios for women, respectively. For incidence, the same pattern as for men is observed, with a higher contribution from aging factors in the fast-aging scenario, and the contribution from population growth is the same as for men. For mortality, a higher contribution of the aging factor was observed in the fast-aging scenario for most cancers, but for pancreatic cancer, a higher contribution of the aging factor was observed in the high scenario.

In terms of cancer incidence, the lowest impact of aging appears to be cervix uteri. However, the fact that only the aging factor continues to increase while the trend and net changer of the other factors continue to decrease suggests that some caution is needed in the older population. In the case of pancreatic cancer and lung cancer, we found a large increase in the long-term trend, and in the case of other major cancers, the increase in incidence due to the aging factor seems to be close to doubling or more than doubling compared to the baseline year (Table13-1).

The contribution of aging factors to deaths continues to increase over the long-term trend for all cancers, with pancreatic cancer showing the largest increase, but the increase in lung, liver, and colorectal cancers cannot be ignored when considering the proportion of total deaths. Non-Hodgkin's lymphoma also shows

a high increase along with other major cancers, as shown in the previous incidence increase (Table13-2).

The net change in incidence varied significantly by cancer type, with liver and cervix uteri showing a decrease in incidence over the long-term trend, and cervix uteri showing fewer cases than in the baseline year. The largest increase was in pancreatic cancer, and the other major cancers, lung and colorectal, also showed high increases. Non-Hodgkin's lymphoma was also one of the cancers that showed a significant increase in addition to the major cancers, and like men, it was one of the cancers that showed a significant increase in the long-term trend, suggesting the need for appropriate surveillance in older age groups (Table14-1).

The net change in deaths showed two different patterns: liver and cervix uteri continued to decrease, as did the net change in incidence, and stomach cancer continued to decrease similarly. pancreas cancer and non-Hodgkin's lymphoma showed a large increase, like the previous net change in incidence. The remaining cancers showed an increase, but the increase was insignificant. However, in the case of colorectal cancer and lung cancer, which are representative major cancers, the increase is not negligible in terms of the total number of deaths because they account for a high proportion of total deaths. Therefore, it is recommended that continuous surveillance of these cancers is necessary (Table14-2).

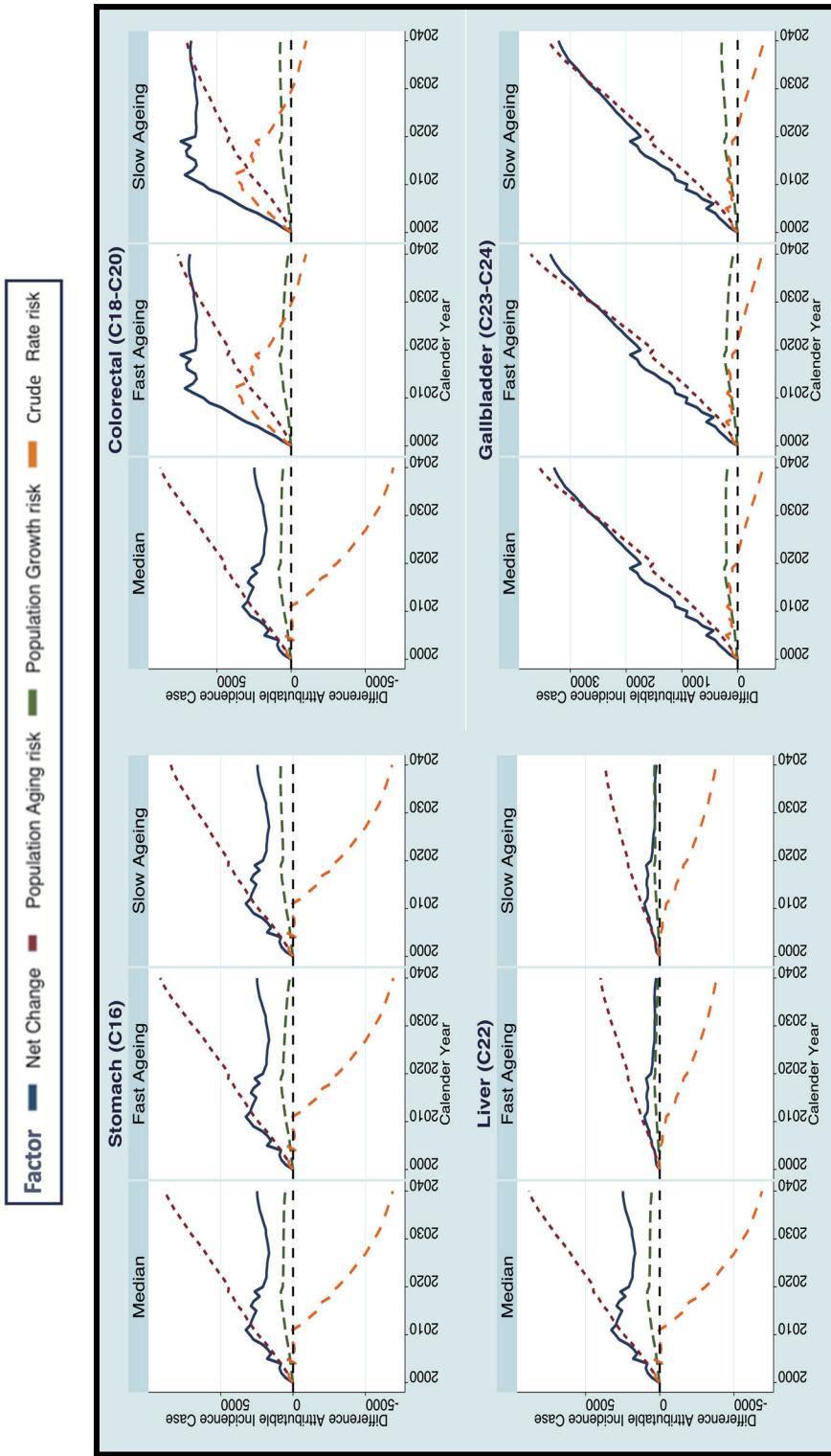


Figure 12-1. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer)

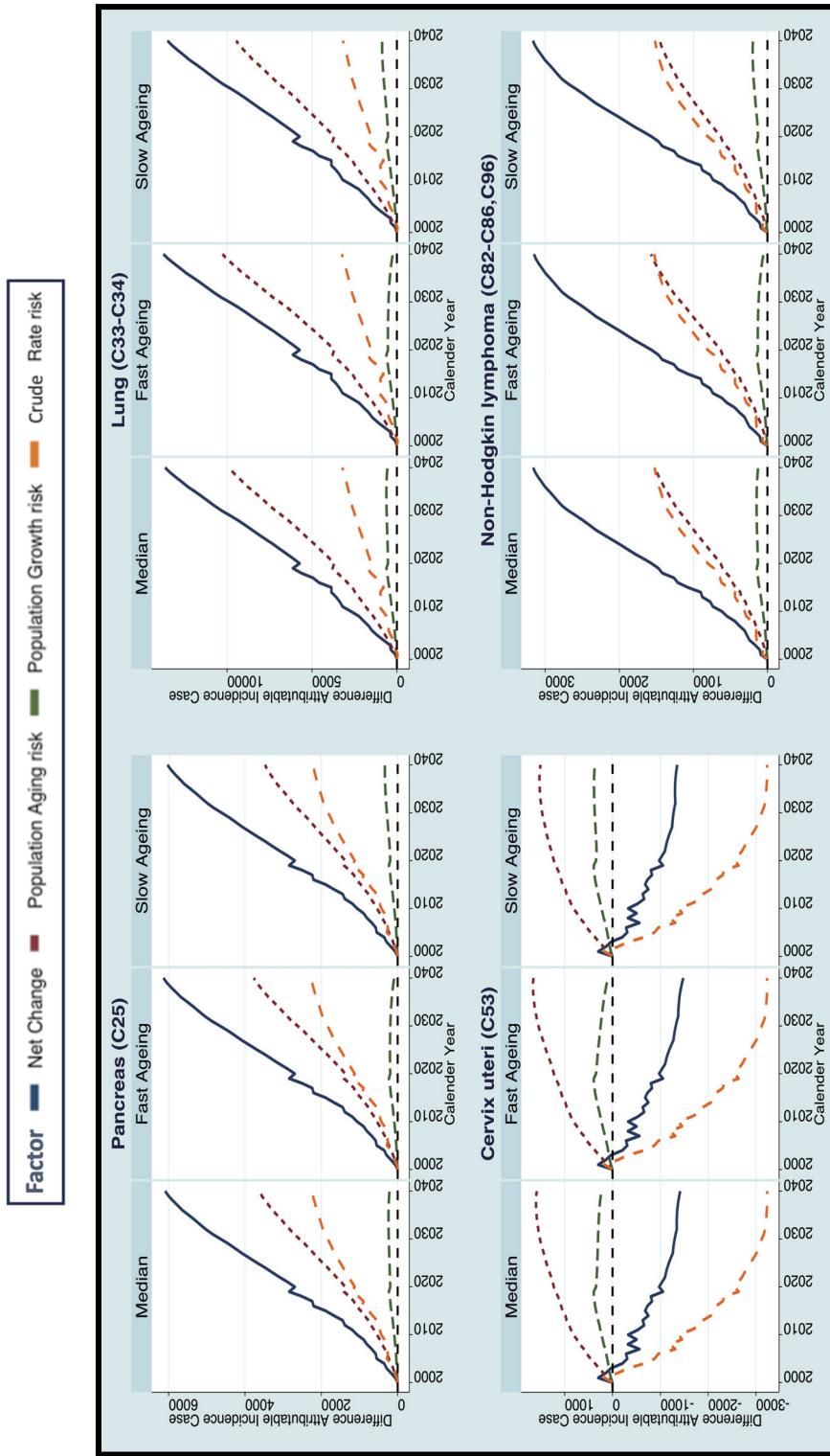


Figure 12-2. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Pancreas, Lung, Cervix Uteri Cancer, Non-Hodgkin lymphoma)

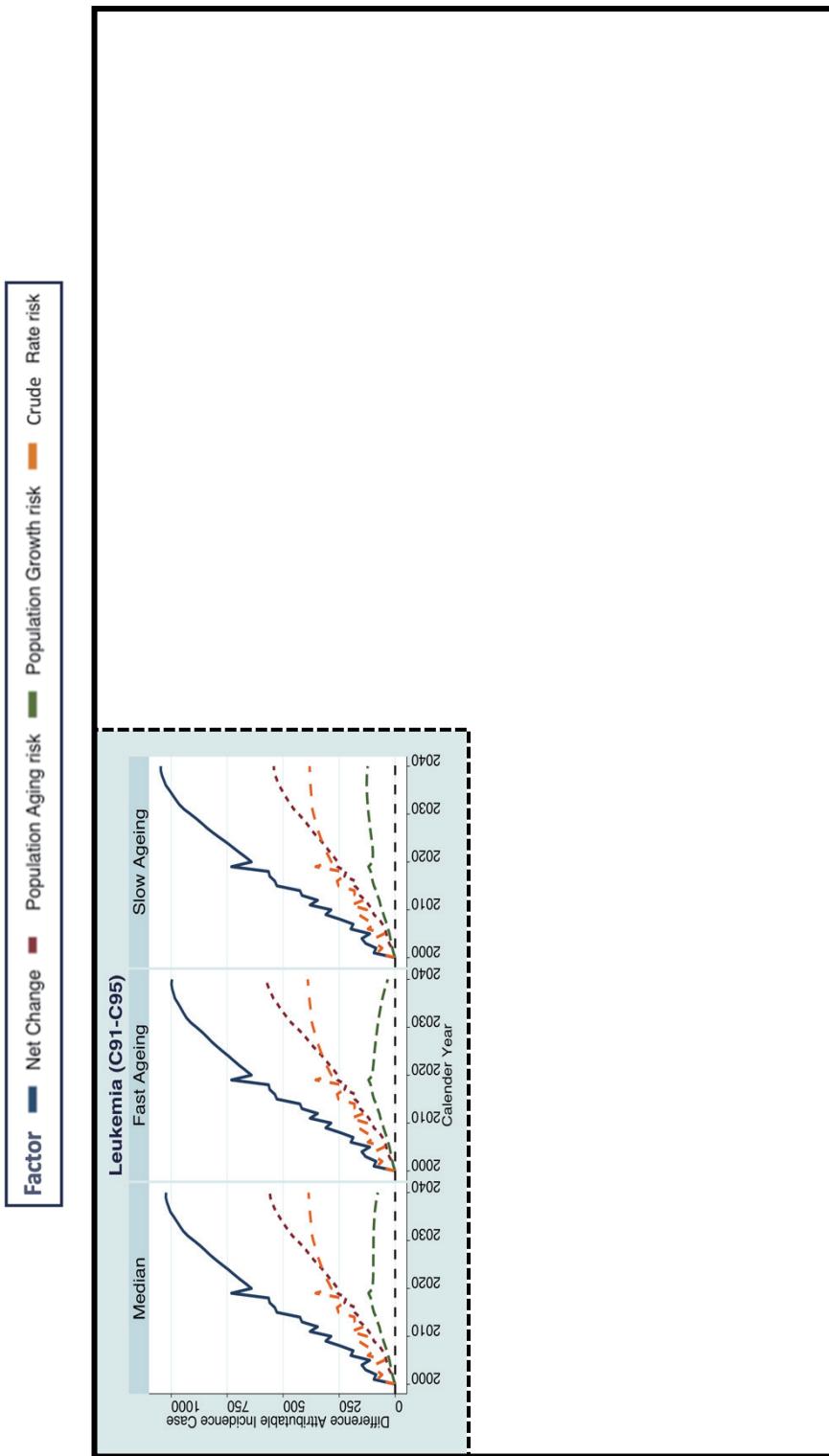


Figure 12-3. Comparison decomposition for female incidence in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Leukemia)

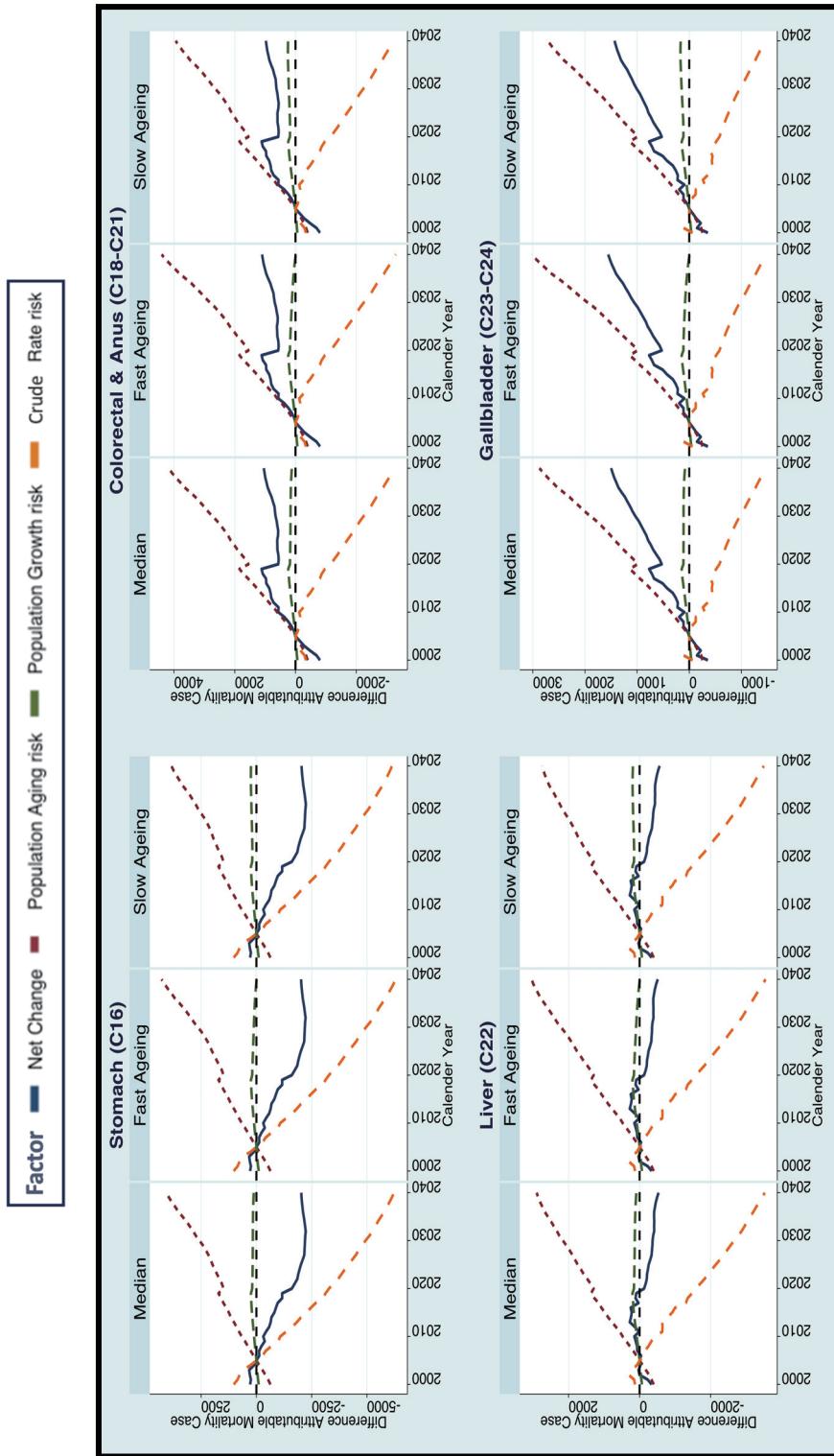


Figure 13-1. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Stomach, Colorectal, Colorectal & Anus, Liver, Gallbladder Cancer)

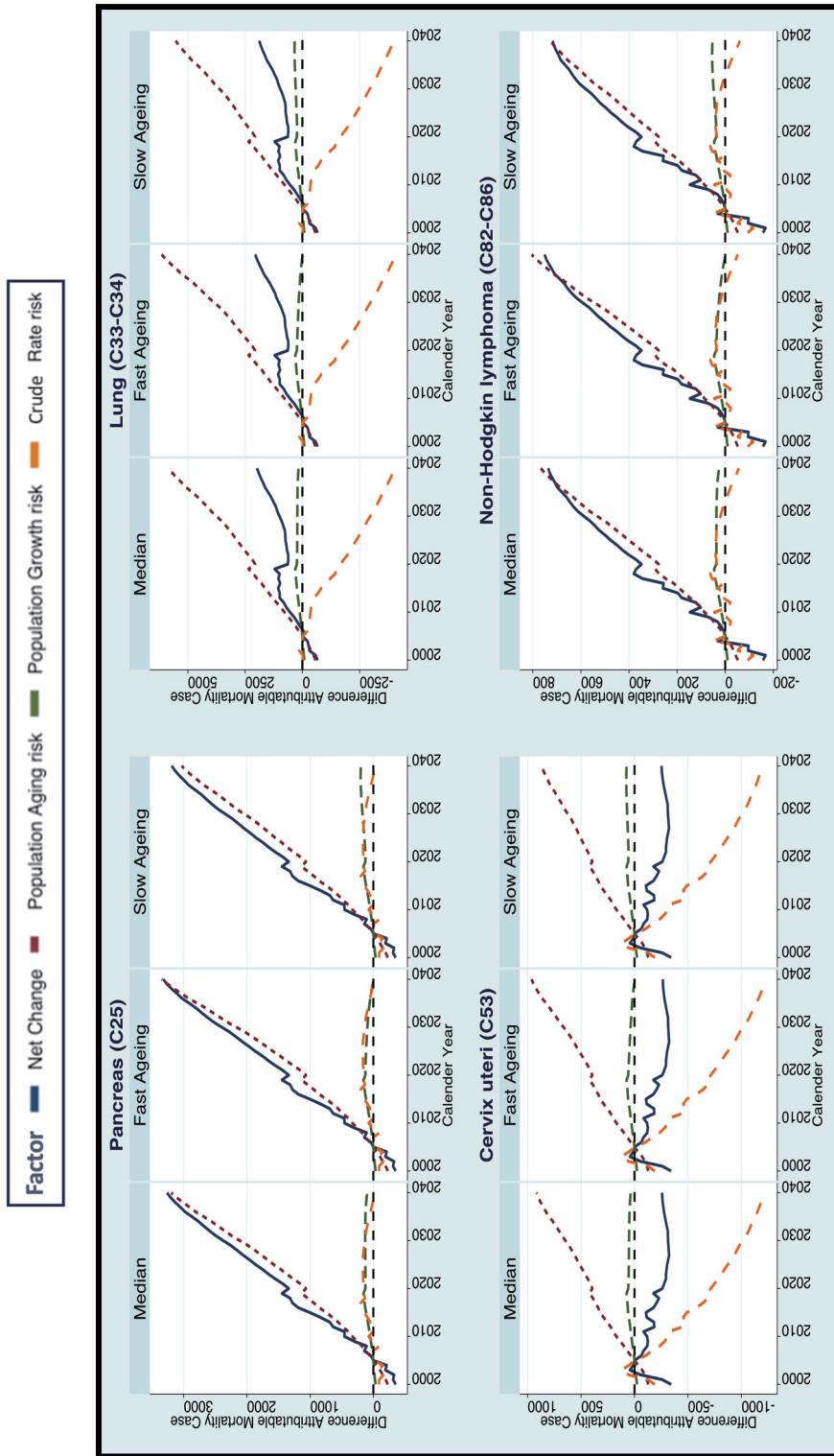


Figure 13-2. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing Slow Ageing Scenarios (Pancreas, Lung, Cervix Uteri Cancer, Non-Hodgkin lymphoma)

Factor ■ Net Change ■ Population Aging risk ■ Population Growth risk ■ Crude Rate risk

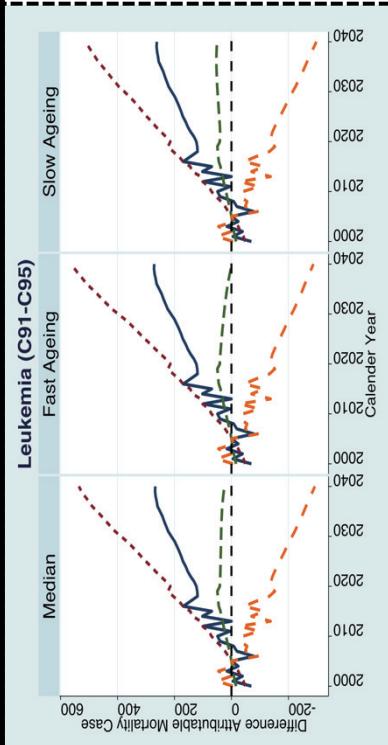


Figure 13-3. Comparison decomposition for female mortality in 2000 to 2040 using NP by Median, Fast Ageing, and Slow Ageing Scenarios (Leukemia)

Table 13-1. Female incidence case difference based on population ageing risk

Cancer	Scenarios (%) (Ref 2000)		
	Median	Fast Ageing	Slow Ageing
Stomach (C16)			
2019	62	62	62
2030	94	96	92
2040	123	128	118
Colorectal (C18-C20)			
2019	92	92	92
2030	129	131	126
2040	159	165	152
Liver (C22)			
2019	67	67	67
2030	97	99	95
2040	119	124	114
Gallbladder (C23-C24)			
2019	101	101	101
2030	164	167	159
2040	226	235	215
Pancreas (C25)			
2019	124	124	124
2030	227	232	222
2040	320	333	307
Lung (C33-C34)			
2019	105	105	105
2030	188	192	184
2040	275	285	263
Cervix Uteri (C53)			
2019	27	27	27
2030	35	36	34
2040	36	38	35
Non-Hodgkin lymphoma (C82-C86,C96)			
2019	74	74	74
2030	140	143	137
2040	186	193	178
Leukemia (C91-C95)			
2019	29	29	29
2030	48	49	48
2040	62	64	60

Table 13-2. Female mortality case difference based on population ageing risk

Cancer	Scenarios (%) (Ref 2005)		
	Median	Fast Ageing	Slow Ageing
Stomach (C16)			
2019	45	45	45
2030	67	69	65
Colorectal & Anus (C18-C21)	106	112	101
2019	66	66	66
2030	98	101	95
Liver (C22)	150	158	142
2019	51	51	51
2030	80	82	77
Gallbladder (C23-C24)	107	112	101
2019	69	69	69
2030	115	118	111
Pancreas (C25)	174	183	165
2019	74	74	74
2030	138	142	134
Lung (C33-C34)	209	219	199
2019	64	64	64
2030	102	105	99
Cervix Uteri (C53)	160	168	151
2019	40	40	40
2030	60	62	58
Non-Hodgkin lymphoma (C82-C86)	85	90	81
2019	61	61	61
2030	112	115	108
Leukemia (C91-C95)	161	169	153
2019	34	34	34
2030	60	61	58
2040	83	87	79

Table 14-1. Net change by female incidence case

Cancer	Scenarios (%) (Ref 2000)		
	Median	Fast Ageing	Slow Ageing
Stomach (C16)			
2019	37	37	37
2030	25	25	25
2040	35	35	34
Colorectal (C18-C20)			
2019	161	161	161
2030	140	141	140
2040	147	147	145
Liver (C22)			
2019	28	28	28
2030	11	11	10
2040	7	8	7
Gallbladder (C23-C24)			
2019	122	122	122
2030	165	166	163
2040	209	213	204
Pancreas (C25)			
2019	251	251	251
2030	409	411	407
2040	538	542	532
Lung (C33-C34)			
2019	170	170	170
2030	273	274	272
2040	380	382	376
Cervix Uteri (C53)			
2019	-24	-24	-24
2030	-30	-30	-29
2040	-33	-34	-31
Non-Hodgkin lymphoma (C82-C86,C96)			
2019	181	181	181
2030	318	317	318
2040	388	387	389
Leukemia (C91-C95)			
2019	81	81	81
2030	100	98	101
2040	113	111	116

Table 14-2. Net change by female mortality case

Cancer	Scenarios (%) (Ref 2005)		
	Median	Fast Ageing	Slow Ageing
Stomach (C16)			
2019	-30	-30	-30
2030	-57	-57	-58
2040	-53	-53	-53
Colorectal & Anus (C18-C21)			
2019	40	40	40
2030	24	25	23
2040	38	40	35
Liver (C22)			
2019	4	4	4
2030	-14	-13	-14
2040	-20	-19	-21
Gallbladder (C23-C24)			
2019	47	47	47
2030	63	64	61
2040	91	94	87
Pancreas (C25)			
2019	95	95	95
2030	155	157	154
2040	214	218	209
Lung (C33-C34)			
2019	33	33	33
2030	27	28	26
2040	54	56	52
Cervix Uteri (C53)			
2019	-17	-17	-17
2030	-29	-29	-29
2040	-24	-25	-24
Non-Hodgkin lymphoma (C82-C86)			
2019	78	78	78
2030	124	125	122
2040	155	158	151
Leukemia (C91-C95)			
2019	19	19	19
2030	33	33	32
2040	41	42	40

5.5. Summary

The purpose of this chapter is to compare the predicted values of cancer incidence and mortality and the effects of factors by sex and scenario. Our analysis revealed variations in the rate of increase and decrease of incidence and mortality by cancer type for both men and women across scenarios, as well as differences in the incidence and mortality in the projections for certain cancers that were affected by the age structure. We also observed changes in the effect of factors by cancer type within the same population projection scenario, particularly in the contribution to the aging factor and the contribution to population growth. Given that some cancer types showed changes in factor effects depending on the scenario, using the same scenario for each cancer type may introduce bias in the estimation.

Therefore, to provide a consistent demographic impact by cancer type, it is essential to perform a decomposition analysis on the predicted values and analyze the differences by scenario. If the effect on the same demographic structure is observed, the scenario based on the scenario can be used. However, if there are differences, appropriate data should be found by discussing the population estimation data to be used in the model, taking those differences into account. This approach can help ensure that the estimations generated by the models are more accurate and reflective of the potential demographic impacts on cancer incidence and mortality rates.

CHAPTER VI. CONCLUSION

In conclusion, this study explored various methods of measuring the burden of cancer in a future society where the cancer burden is anticipated to increase. We employed projection models to estimate cancer incidence and mortality rates and validated the selection of cancers associated with aging through APC-analysis. The results revealed the significant role of aging factors in cancer incidence and mortality. In the Korean context, research has been conducted on estimating cancer incidence and mortality using projection models. Although the importance of aging in cancer incidence and mortality has been emphasized in various studies, no research has combined these factors with projection models [73]. Our study aimed to quantify the risk associated with each factor by separating the incidence and mortality rates of each cancer type into three components: population aging, population growth, and rate change risk. We selected five representative models, including both demographic and non-demographic-based models, for projection. Decomposition analysis, which was performed only for the APC models, consistently identified aging as a major factor contributing to the increase in cancer incidence or mortality.

We observed varying degrees of robustness among the selected models, with the NP model proving to be the most reliable. As such, we utilized the NP model to construct demographic scenarios and measure incidence and mortality rates. While most cancer types exhibited a high number of cases or deaths in fast-aging

scenarios, the actual increase or decrease in the absolute number of cases or deaths was not significant. These findings suggest that caution is required when selecting scenarios for cancer type, as each type may be affected by different demographic factors.

Our study has following limitations:

Moller's study [14] recommends using data from at least 20 years when fitting the APC model. In addition, when using other time series models, the average observation period of the data used in other existing studies has been found to be between 20 and 50 years. Although long-term data has been found to have high predictive power when evaluated in several studies, it is not known how much the predictive power decreases when using short-term data. Therefore, there is a possibility that the predictive power of models built using somewhat short-term observations may be somewhat reduced.

The second is the observation limit. We expected our data to generate persistent underestimates in future model fits for mortality and incidence that were somewhat under-observed due to the coronavirus [31]. The inclusion of data from 2020 to 2022 during the pandemic, and subsequent data updates for 2023, 2024 ... are likely to result in lower estimates of the overall trend in cancer mortality incidence than we would have predicted using the 2000 to 2019 observation period. Therefore, to check the trend of cancer incidence and mortality, it is necessary to continuously secure the latest data and build a model for each year and continuously check the

trend.

The implications of changing health policies and social trends can introduce significant bias, particularly regarding cancer incidence and mortality rates. For example, aggressive screening policies for certain diseases like thyroid cancer may result in inflated incidence rates during specific periods, potentially making the model less reliable for predicting future rates [74]. On the mortality front, advancements in diagnostic techniques could gradually diminish uncertainties surrounding the classification of cancer types upon death. Shin and colleagues discovered that historical misclassification, such as with uterine and cervical cancer, had led to an overestimation of recorded fatalities [75]. It's plausible that similar overestimations may have occurred with other types of cancer in the past. This implies that even if the model is designed with accurate incidence and mortality rates, the predictions might still tend towards a higher level of realism because the model has been calibrated with previously overestimated data. These preliminary data-related concerns signal that any usage or interpretation of the model's results must be undertaken with a high degree of caution.

Notwithstanding these limitations, the present study can be deemed significant as it sheds light on the contribution of aging factors to the rise in incidents or deaths in projections through projection models, which have not been extensively explored in Korea. Furthermore, the study presents realistic projections by constructing scenarios to account for various future uncertainties and diversify

projections. These findings have practical implications for policymakers and health professionals in developing effective strategies to mitigate the impact of aging on cancer incidence and mortality. Thus, this study represents a valuable contribution on cancer trends and projections, despite its limitations.

This research has helped to guide our subsequent research. Our ultimate research goal is to derive a detailed quantification of the burden of cancer, not only from primary studies, but also from secondary and tertiary studies. To derive these outcomes, values such as QALYs and DALYs are needed, and these values can be obtained by building a prevalence model that considers both incidence and mortality. Therefore, future studies will develop it and use more data to build a prevalence model not only to check the prevalence of common cancers, but also to check the socioeconomic burden.

Ultimately, our study highlights the critical role of aging in cancer incidence and mortality and underscores the importance of considering demographic factors when predicting long-term trends in cancer burden.

BIBLIOGRAPHY

1. National Statistical Office, "Causes of Death Statistics", 2021, 2023.05.05, Number of deaths by cause of death (104 items)/gender/age (5 years old), mortality
2. Ministry of Health and Welfare, "Cancer Registration Statistics", 2020, 2023.05.05, the number and incidence of cancer by 24 carcinoma/gender/age (5 years old)
3. The importance of aging in cancer research. *Nat Aging*, 365–366 (2022).
4. OECD (2022), *OECD Economic Surveys: Korea 2022*, OECD Publishing, Paris.
5. HIRA, "Statistics on medical expenses for cancer patients, 2013-2019"
6. Moller H, Fairley L, Coupland V, Okello C, Green M, Forman D, Moller B, Bray F. The future burden of cancer in England: incidence and numbers of new patients in 2020. *Br J Cancer*. 2007;96(9):1484–8.
7. Nowatzki J, Moller B, Demers A. Projection of future cancer incidence and new cancer cases in Manitoba, 2006-2025. *Chronic Dis Can*. 2011;31(2):71–8.
8. Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction. *Stat Med*. 1997;16(20):2297–309.
9. Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Stat Med*. 1994;13(15):1513–23.
10. Stock C, Mons U, Brenner H. Projection of cancer incidence rates and case numbers until 2030: A probabilistic approach applied to German cancer registry data (1999-2013). *Cancer Epidemiol*. 2018;(57):110–9.

11. Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics*. 2005;6(4):576–89.
12. Engeland A, Haldorsen T, Tretli S, Hakulinen T, Horte LG, Luostarinen T, Schou G, Sigvaldason H, Storm HH, Tulinius H, et al. Prediction of cancer mortality in the Nordic countries up to the years 2000 and 2010, on the basis of relative survival analysis. A collaborative study of the five Nordic Cancer registries. *APMIS Suppl*. 1995;49:1–161.
13. Smith TR, Wakefield J. A review and comparison of age-period-cohort models for Cancer incidence. *Stat Sci*. 2016;31(4):591–610.
14. Møller, B., Fekjær, H., Hakulinen, T., Sigvaldason, H., Storm, H.H., Talbäck, M. and Haldorsen, T. (2003), Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Statist. Med.*, 22: 2751-2766.
15. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*. 2000 Feb 15;19(3):335-51.
16. Raicharoen, T., Lursinsap, C., & Sanguanbhokai, P. (n.d.). Application of critical support vector machine to time series prediction. *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03*.
17. Tong F. Primary visual cortex and visual awareness. *Nat Rev Neurosci*. 2003 Mar;4(3):219-29.
18. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.

19. Zhang, P. (2007). *Multiple Imputation: Theory and Method*. International Statistical Review, 71(3), 581–592.
20. Choi Y, Kim YJ, Shin HR, Noh DY, Yoo KY. Long-term prediction of female breast cancer mortality in Korea. *Asian Pac J Cancer Prev*. 2005 Jan-Mar;6(1):16-21. PMID: 15780025.
21. Park HM, Won YJ, Kang MJ, Park SJ, Kim SW, Jung KW, Han SS. **Trend Analysis and Prediction of Hepatobiliary Pancreatic Cancer Incidence and Mortality in Korea**. *J Korean Med Sci*. 2022 Jul;37(28):e216.
22. Pak S et.al., Incidence and mortality projections for major cancers among Korean men until 2034, with a focus on prostate cancer. *Investigation Clinical Urology*. 2022 Mar;63(2):175-183.
23. Son M, Yun JW. Cancer Mortality Projections in Korea up to 2032. *J Korean Med Sci*. 2016 Jun;31(6):892-901.
24. Kwak, K., Cho, S., & Paek, D. (2021). Future Incidence of Malignant Mesothelioma in South Korea: Updated Projection to 2038. *International Journal of Environmental Research and Public Health*, 18_(12), 6614.
25. Hong, Y et.al., Projection of Cancer Incidence and Mortality From 2020 to 2035 in the Korean Population Aged 20 Years and Older. *Journal of Preventive Medicine and Public Health*. Korean Society for Preventive Medicine.
26. Jung KW, Won YJ, Kang MJ, Kong HJ, Im JS, Seo HG. Prediction of Cancer Incidence and Mortality in Korea, 2022. *Cancer Res Treat*. 2022 Apr;54(2):345-351.

27. Paul de Boer & João F. D. Rodrigues (2020) Decomposition analysis: when to use which method? *Economic Systems Research*, 32:1, 1-28.
28. Cheng, Xunjie, et al. "A new method to attribute differences in total deaths between groups to population size, age structure and age-specific mortality rate." *PLoS One* 14.5 (2019): e0216613.
29. Wong, I. O., Cowling, B. J., Law, S. C., Mang, O. W., Schooling, C. M., & Leung, G. M. (2010). Understanding sociohistorical imprint on cancer risk by age-period-cohort decomposition in Hong Kong. *Journal of Epidemiology & Community Health*, 64(7), 596-603.
30. Wang, Lijun, et al. "Lung cancer mortality trends in China from 1988 to 2013: new challenges and opportunities for the government." *International journal of environmental research and public health* 13.11 (2016): 1052.
31. Wells, Chad R., and Alison P. Galvani. "Impact of the COVID-19 pandemic on cancer incidence and mortality." *The Lancet Public Health* 7.6 (2022): e490-e491.
32. Dongyoung, Noh. "The Clinical and Histopathological Characteristics of Male Breast Cancer Patients" *Journal of Breast Cancer* 10, no.3 (2007) : 211-216.
33. Tsoi, Kelvin KF, et al. "Cancer burden with ageing population in urban regions in China: projection on cancer registry data from World Health Organization." *British Medical Bulletin* 121.1 (2017): 83-94.
34. Fay, Michael P., et al. "Estimating average annual percent change for disease rates without assuming constant change." *Biometrics* 62.3 (2006): 847-854.

35. Gupta, Prithwis Das. Standardization and decomposition of rates: A user's manual. No. 186. US Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1993.
36. Bashir, S. A., and J. Estève. "Analysing the difference due to risk and demographic factors for incidence or mortality." *International journal of epidemiology* 29.5 (2000): 878-884.
37. Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. Future of cancer incidence in the United States: burdens upon an aging, changing nation. *J Clin Oncol.* 2009;27:2758-2765.
38. Carstensen B. Age-period-cohort models for the lexis diagram. *Stat Med.* 2007;26:3018- 45.
- 39, Dyba T, Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. *Stat Med.* 2000;19:1741-52.
- 40, Breslow N, Day N. Statistical methods in cancer research. Volume 2: The design and analysis of cohort studies. Lyon (FR): IARC Scientific Publications no. 82; 1987.
41. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med.* 1987;6:469-81.
42. Moller B, Fekjaer H, Hakulinen T, Tryggvadottir L, Storm HH, Talback M, Haldorsen T. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev Suppl* 2002; 11: S1-S96
43. Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *J R Stat Soc Ser C, Appl Stat.* 2002;51:151-64.

44. Robert CP, Casella G. Monte Carlo Statistical Methods. New York: Springer-Verlag; 1999.
45. Lizotte, Daniel James. "Practical bayesian optimization." (2008).
46. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics*. 1983;39:311-24.47.
48. Plummer, Martyn. "Penalized loss functions for Bayesian model comparison." *Biostatistics* 9.3 (2008): 523-539.
49. Yang, Yang, and Kenneth C. Land. Age-period-cohort analysis: New models, methods, and empirical applications. Taylor & Francis, 2013.
50. O'Brien, Robert M. "Age period cohort characteristic models." *Social science research* 29.1 (2000): 123-139.
51. Bell, Andrew, and Kelvyn Jones. "The hierarchical age-period-cohort model: Why does it find the results that it finds?." *Quality & quantity* 52 (2018): 783-799.
52. Kupper, Lawrence L., et al. "Statistical age-period-cohort analysis: a review and critique." *Journal of chronic diseases* 38.10 (1985): 811-830.
53. Pickle, Linda W., et al. "A new method of estimating United States and state-level cancer incidence counts for the current calendar year." *CA: a cancer journal for clinicians* 57.1 (2007): 30-42.
54. Arbeev, Konstantin G., et al. "Mathematical models for human cancer incidence rates." *Demographic Research* 12 (2005): 237-272.

55. Maddams, J., M. Utley, and H. Møller. "Projections of cancer prevalence in the United Kingdom, 2010–2040." *British journal of cancer* 107.7 (2012): 1195-1202.
56. Lee, Terry CK, C. B. Dean, and Robert Semenciw. "Short-term cancer mortality projections: A comparative study of prediction methods." *Statistics in Medicine* 30.29 (2011): 3387-3402.
57. Zhu, Li, et al. "Predicting US-and state-level cancer counts for the current calendar year: Part II: evaluation of spatiotemporal projection methods for incidence." *Cancer* 118.4 (2012): 1100-1109.
58. Miller, Kimberly D., et al. "Updated methodology for projecting US-and state-level cancer counts for the current calendar year: part II: evaluation of incidence and mortality projection methods." *Cancer Epidemiology, Biomarkers & Prevention* 30.11 (2021): 1993-2000.
59. Ho, Siu Lau, and Min Xie. "The use of ARIMA models for reliability forecasting and analysis." *Computers & industrial engineering* 35.1-2 (1998): 213-216.
60. Bosq, Denis. *Linear processes in function spaces: theory and applications*. Vol. 149. Springer Science & Business Media, 2000.
61. Fairman, Frederick Walker. *Linear control theory: the state space approach*. John Wiley & Sons, 1998.
62. Moonen, Marc, et al. "On-and off-line identification of linear state-space models." *International Journal of Control* 49.1 (1989): 219-232.

63. Jerant, Anthony F., et al. "Age-related disparities in cancer screening: analysis of 2001 Behavioral Risk Factor Surveillance System data." *The Annals of Family Medicine* 2.5 (2004): 481-487.
64. Viswanath, Kasisomayajula, et al. "Cancer knowledge and disparities in the information age." *Journal of health communication* 11.S1 (2006): 1-17.
65. DePinho, Ronald A. "The age of cancer." *Nature* 408.6809 (2000): 248-254.
66. Rosenberg, Philip S., David P. Check, and William F. Anderson. "A Web Tool for Age–Period–Cohort Analysis of Cancer Incidence and Mortality RatesSoftware for Cancer Rates and Trends." *Cancer epidemiology, biomarkers & prevention* 23.11 (2014): 2296-2302
67. Zeaiter, M., J-M. Roger, and Véronique Bellon-Maurel. "Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods." *TrAC Trends in Analytical Chemistry* 24.5 (2005): 437-445.
68. Lee, JayHyung, and Z. H. Yu. "Tuning of model predictive controllers for robust performance." *Computers & chemical engineering* 18.1 (1994): 15-37.
69. Rahib, Lola, et al. "Estimated projection of US cancer incidence and death to 2040." *JAMA Network Open* 4.4 (2021): e214708-e214708.
70. Mistry, M., et al. "Cancer incidence in the United Kingdom: projections to the year 2030." *British journal of cancer* 105.11 (2011): 1795-1803.
71. Poirier, Abbey E., et al. "The future burden of cancer in Canada: long-term cancer incidence projections 2013–2042." *Cancer epidemiology* 59 (2019): 199-207.

72. National Statistical Office, Prospective population projection 2070, 2023.05.05, estimated population by scenario (estimated population by gender and 5 years old) / nationwide
73. Yancik, Rosemary. "Cancer burden in the aged: an epidemiologic and demographic overview." *Cancer: Interdisciplinary International Journal of the American Cancer Society* 80.7 (1997): 1273-1283.
74. Ahn, H. S., Kim, H. J., & Welch, H. G. (2014). Korea's Thyroid-Cancer "Epidemic" Screening and Overdiagnosis. *New England Journal of Medicine*, 371(19), 1765–1767.
75. Shin HR, Park S, Hwang SY, Kim JE, Jung KW, Won YJ, Hwang SS, Yim SH, Choi KS, Park EC, Park SY, Kim JW, Lee HP. Trends in cervical cancer mortality in Korea 1993-2002: corrected mortality using national death certification data and national cancer incidence data. *Int J Cancer*. 2008 Jan 15;122(2):393-7.

ACKNOWLEDGEMENT

The path to my master's degree was not a smooth one, and I did not accomplish it alone, but with the help of many people. I would like to use this chapter to thank them.

First and foremost, I would like to express my profound gratitude to my thesis advisor, Prof. Hyunsoon Cho, for her invaluable guidance, persistent encouragement, and insightful criticisms during the research and writing of this thesis. Her extensive knowledge and unwavering dedication inspired and propelled me throughout this journey.

I would like to thank Prof. Jin Kyung Oh for her generous advice on the direction of the research and the research methods, and for her willingness to chair the committee. I would also like to thank Prof. Chang Mo Oh for traveling a long way to be with me, for his insightful advice, and for helping me develop my research.

My special thanks go to my lab members, whose unwavering support and camaraderie made this endeavor not only bearable but enjoyable. In particular, I would like to thank to Sanghee Lee, who graduated earlier, for her help during my master's program. In addition, I would like to thank Dahhay Lee, Tran Thi Ngoc, Min-sun, Sung, Jongheok Jung, and Thao Minh TU, and I would like to thank Eun

Jung Kim and Eun Ji Kim, who were previously members of the lab. I would also like to thank Dahye Baek, although in a different lab.

I would like to extend my gratitude to the entire faculty and staff of the Department of Cancer AI & Digital Health at NCC-GCSP. Their expertise and assistance were fundamental to the completion of this project.

I must also acknowledge my family, who have provided me with unconditional support throughout my academic journey. Their faith in my abilities has been a driving force behind my pursuit of this degree.

Most of all, i give glory to God for taking us on this challenging and rewarding journey.