UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

BREAST CANCER IN CHILE (2002-2018): STUDYING PUBLIC-PRIVATE GAP THROUGH STOCHASTIC SIMULATION.

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MATEMÁTICAS APLICADAS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

BENJAMÍN FELIPE MADARIAGA ESPINOZA

PROFESORA GUÍA:
SUSANA MONDSCHEIN PRIETO

MIEMBROS DE LA COMISIÓN:
ALEJANDRO MAASS SEPÚLVEDA
JORGE VERA ANDREO

Este trabajo ha sido parcialmente financiado por:
CMM ANID PIA AFB170001, CMM ANID BASAL ACE210010,
CMM ANID BASAL FB210005 y ANID PIA/APOYO AFB180003.

SANTIAGO DE CHILE
2022

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MATEMÁTICAS APLICADAS
MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL MATEMÁTICO
POR: **BENJAMÍN FELIPE MADARIAGA ESPINOZA**
FECHA: 2022
PROF. GUÍA: SUSANA MONDSCHEIN PRIETO

CÁNCER DE MAMA EN CHILE (2002-2018): ESTUDIANDO LA BRECHA PÚBLICO-PRIVADA A TRAVÉS DE SIMULACIÓN ESTOCÁSTICA.

Este trabajo está enfocado en el estudio del cáncer de mama en Chile y las inequidades entre los sistemas de salud público y privado. Se divide en dos partes. Primero, se calcularon incidencia, mortalidad y sobrevida usando bases de datos públicas del Ministerio de Salud y usando el estimador de Kaplan Meier y el modelo de riesgos proporcionales de Cox. Estas estadísticas fueron estudiadas a nivel nacional y regional, y por sistema de salud, encontrándose una diferencia importante en sobrevida entre los sistemas de salud público y privado. En segundo lugar, se desarrolló un proceso de simulación estocástica basado en modelos de crecimiento tumoral que considera políticas de tamizaje. Este fue usado para estudiar las diferencias de sobrevida antes mencionadas y entender en qué medida está relacionada con diferencias en el tamizaje con mamografías. A pesar de que el tamizaje mostró un impacto relevante en los resultados de sobrevida, éste logra explicar solo una fracción de la diferencia encontrada entre los sistemas de salud público y privado.

Palabras clave: cáncer de mama, incidencia, mortalidad, sobrevida, inequidad en salud, sistema de salud en Chile, simulación estocástica

ABSTRACT OF THE THESIS FOR
THE MASTER'S DEGREE IN ENGINEERING SCIENCES
WITH MENTION IN APPLIED MATHEMATICS
THESIS FOR THE DEGREE OF
MATHEMATICAL CIVIL ENGINEER
BY: **BENJAMÍN FELIPE MADARIAGA ESPINOZA**
DATE: 2022
GUIDE PROF.: SUSANA MONDSCHEIN PRIETO

BREAST CANCER IN CHILE (2002-2018): STUDYING PUBLIC-PRIVATE GAP THROUGH STOCHASTIC SIMULATION.

This work is focused on breast cancer in Chile and the inequities between the public and private health care systems. It is separated in two parts. Firstly, incidence, mortality and survival rates were calculated making use of public databases provided by the Health Ministry and using the Kaplan Meier estimator and the Cox proportional hazards model. These statistics were studied at a national and regional level, and by health care system, finding an important difference in survival between the public and private system. Secondly, a stochastic simulation process based on tumor growth models that considers screening policies was developed. This was used to study the above mentioned difference in survival and to understand in which degree it is related to differences in screening through mammography. Although screening showed an important impact in survival outcomes, this impact only explains a fraction of the difference found between the public and private system.

Keywords: breast cancer, incidence, mortality, survival rates, health care inequalities, Chilean health system, stochastic simulation.

A mi familia.

Agradecimientos

En primer lugar quiero agradecer a mis padres, Aldo y Bruna por todo su apoyo, amor y sabiduria que me entregan día a día. Gracias a mis hermanos Alduca, Yiyi y Nene por todas las alegrías y momentos especiales, me inspiran a seguir creciendo.

Gracias a Susana por su constante apoyo y guia, por cada reunion semana a semana, donde siempre surgian nuevas ideas y cosas por hacer. Gracias también a Soledad por apoyarnos en innumerabes reuniones con su experiencia y conocimiento.

Quiero agradecerle a todos mis amigos y amigas, que van surgiendo en el camino y que son especiales en distintas etapas. Gracias a los amigos y amigas de Los Andes por infinitas juntas y buena onda, en especial a los Cajas. A todos mis amigos y amigas de la U, en especial a los de la última etapa en el DIM, los Chacras: Azócar, Barri, Bruno, Choco, Javo, Magali, Majo, Nico, Pelela, Pavlito, Vicho y Vivi.

Finalmente agradecer muy especialmente a mi partner de la vida, Valentina Moreno. Gracias infinitas por todo tu amor y comprensión, por tus consejos, por tanto que hemos vivido juntos. Gracias por motivarme a seguir adelante.

Table of Content

Tables Index

Ilustrations Index

Introduction

Breast cancer (BC) is one of the most common cancers in women worldwide, with 2.26 million new cases and 685 thousand deaths globally in 2020, and 7.8 million women alive with breast cancer diagnosed in the past 5 years [1]. In 2020, the world age standardized incidence and mortality rates were 47.8 and 13.6, respectively [2]. It is also the cancer with highest mortality for women in Chile [3].

Although the breast cancer treatments have advanced over the last years, chances of improvement highly depend on the early diagnosis of the tumor. Tumors detected in an early stage can increase the survival rate from 27 % to 98 % compared to those in late stages [4]. This is why screening mammography programs have been implemented, mainly in developed countries, aiming to reduce breast cancer mortality [5].

In Chile this problem is addressed in different ways depending on the health care provider. The Chilean health care system is a hybrid of public and private providers and insurances consisting of three main players: the National Health Fund (FONASA) for 78 % of the Chilean population, the private health care insurers (ISAPREs) for 14 % of the Chilean population, and, finally, the Military and Police Forces' health system for approximately 3 % of the population. Privately insured patients can solely access private providers (with a variety of coverage). On the other hand, FONASA patients – paying a lower monthly price – might access public and private hospitals, depending on their income level with different amounts of co-payments. Thus, FONASA has four different segments of patients (A,B,C, and D), with higher co-payments for higher income patients, being A people with no monthly income, B people with monthly income lower than USD 409, C people with monthly income between USD 409 and 597 and D people with monthly incomes higher than USD 597.

The selection of the health care provider is determined mainly based on the individual economic income [6], [7]. As a consequence, health care indicators are significantly better for those with private health insurance. For this reason, in 2004, Chile implemented a profound health reform, the Explicit Health Guarantees (GES), originally known as AUGE plan, that aimed to achieve a more equitable and fairer system [8]. The objective of this reform was to ensure access, quality, opportunity and financial protection to those who require care for a set of pathologies to all Chilean citizens. It started in 2005 with 25 pathologies, including

breast cancer (for people aged 15 years and over with suspected, diagnosed or recurrent breast cancer), and has gradually grown to currently cover 85 pathologies.

There is an important bulk of evidence that the GES plan has reached the goals of opportunity [9] and is constantly monitoring the quality of health care centers [10]. However, to the best of our knowledge, there is no evidence showing that the effectiveness of the GES plan on breast cancer, through leveling out these guaranties for all citizens, has resulted in a reduction (or elimination) of the gap on health outcomes, such as mortality and survival rates among Chilean women.

According to the World Health Organization (WHO) [11], a screening program is not just a single test but rather a pathway that starts by identifying the people who are eligible for screening and stops when the outcomes are reported. Screening should invite people who do not have symptoms to undergo testing. Despite having preventive mammographies, neither public nor private health systems have a screening program in accordance with the previous definition.

Preventive mammographies in the public system were included through the Preventive Medicine Exams program (EMP), but this is not properly implemented as a screening program, for this only considers financial support and there is no follow up or active search of women. The EMP initially considered a single mammography applied at the age of 50, which was then enlarged to the realization of mammographies every 3 years between the ages of 50 and 54 in the year 2009. Now a days, and since 2013, women between the ages of 50 to 59 years old have right to preventive mammographies every 3 years.

The Chilean private health system not only does not have a screening program, but neither does it established any mammographies guideline. Nevertheless, physicians in the private system tend to recommend yearly mammographies to adult female patients. Thus, according to the National Socioeconomic Characterization Survey (CASEN) performed by the Ministry of Social Development, 50 % of women over 35 years old from the private sector have had at least one mammography in the last year, and 71.4 % in the last three years, while the same values are 31.9 % and 55.1 % for the public system [12]. This exhibits a important difference in terms of breast cancer prevention between the public and private health care systems.

In order to study the impact of this disease and to take effective measures to improve the populations health, such as screening programs, it is necessary to have reliable and up to date statistics over the incidence, mortality and survival rates.

In Chile, due to the lack of a national cancer registry, breast cancer crude and age standardized incidence rates are estimated based on projections on the number of breast cancer cases diagnosed from 1998 to 2012 with follow ups until 2015, from four regional population-based registries (Antofagasta and Los Ríos regions, and Concepción and Biobío provinces). The estimates for the rest of the country and from 2013 onwards are based on statistical models,

with assumptions such a constant fatality rate across regions and time, and, in some instances, using neighboring countries data [13, 14]. The figures reported by the Chilean Ministry of Health (Minsal) for 2018 correspond to 44.1 and 11.3 for breast cancer age standardized incidence and mortality rates, respectively [14, 15]. Thus, knowing the significant differences in ethnicity, urbanization, socioeconomic composition and health coverage between regions [12, 16, 17, 18], assumptions such as a constant fatality rate across the country might lead to imprecise estimations and therefore, inadequate public policies. A similar effect might have the non-consideration of advances in breast cancer treatment over time.

On the other hand, mortality rates are much more reliable. This because of the national death registry which includes a series of demographic variables along with the cause of death of every decease in Chile. Studies such as [19] by Icaza, Nuñez and Bugueño exploit the potential of this registry by studying breast cancer mortality including ecological analysis by socio-demographic variables.

Other important health parameters to measure are the survival rates. Survival rates for Chile were found in the clinical study of Del Castillo et al [20], who only considers women from the public health system, showing a average 5 year survival of 75.1 %. This survival was calculated considering 5,119 medical records and the study also presents survival rates separating cases by cancer stage and other medical variables. Another study that presents Chilean survival rates is [21], where such rates are estimated for a more extensive set of countries. Here Allemani estimates the Chilean five year survival rate and its 95 % CI to be 77.1 % and [70.4 %,83.8].

The main objective of this thesis is to study breast cancer in Chile, to estimate key statistics segregating by demographic variables and by health care provider, and to study to which extent this results are influenced by the use of preventive screening mammographies.

In order to achieve such objectives, the study is separated in two main parts. The first part will estimate breast cancer incidence, mortality, and survival rates using publicly available data from 2002 to 2018, at national and regional levels, by age group and health care insurer (public vs private), with the aim of measuring the impact on breast cancer health care outcomes in different segments of Chilean women. To meet this objective, a complete anonymized public database of national registry of hospital discharges was used, which includes information and diagnosis at the patient level, and the national death registry, with primary and secondary causes of death through a unique id classification, compiled by the Department of Health Statistics and Information (DEIS) of the Ministry of Health. To the best of our knowledge the public database of hospital discharges has not been used to study breast cancer.

Then, the second part will develop and use a stochastic model to evaluate the impact of preventive mammographies in the survival of women and compare such outcomes between different screening programs. It will be of particular interest to estimate the differences between no screening, the screening applied in the public and private systems, and a best case

scenario. Such findings will be compared with those of the first part.

Ethics statement

This work used publicly available data at the Chilean Ministry of Health through the DEIS. All data are protected, and personal information is anonymized. Therefore, no consent from participants was required.

Chapter 1

Analysis of Incidence, Mortality and Survival Rates

In this first part of the study, the authors are interested in evaluating the state of breast cancer in Chile and its differences between public and private health care systems. With this purpose, breast cancer incidence, mortality and survival rates will be estimated. This will be done by analysing two databases provided by the Department of Health Statistics and Information (DEIS) of the Ministry of Health.

1.1. Data

The two public anonymized databases provided by DEIS are the national death registry and the hospital discharges database. The national death registry includes 2,549,800 deaths from January 1990 to December 2018. For each death entry, the patient's id (identifying code), date of birth, date of death, gender, town and region of residence, health insurance, marital status, occupation and a cause of death code according to the International Statistical Classification of Diseases and Related Health Problems (CDI-10) were available.

The second database includes all discharges from all public and private healthcare facilities in the country, which consists of 32,443,591 registries from January 2001 to December 2020. Each registry has 39 fields, such as the patient's id (same as the national death database), date of birth, gender, town and region of residence, ethnicity, health insurance, benefit range, length of stay, condition at discharge, and primary and secondary diagnoses according to CDI-10 classification among others.

1.1.1. Inclusion and exclusion criteria

Although the digitized entries for the hospital discharges starts in 2001, the first year contained a large amount of inconsistencies and missing data, with 21 % fewer breast cancer registries than the rest of the years. Furthermore, data from 2019 onward is not reliable due to the political and social uprising in Chile (October - December 2019) [22] and COVID 19 pandemic, which largely influenced hospital discharges and deaths. Therefore, this study considers data from 2002 to 2018. Accordingly, the same period of time will be considered for the national death registry.

A discharge or death registry is considered to directly correspond to breast cancer if its CDI-10 diagnostic code belongs to the categories C50 and D05, or to the subcategory D486.

In order to calculate breast cancer mortality, all female registries whose cause of death was directly associated with breast cancer were selected from the 2,549,800 available in the deceases database. After this, 18 death registries with missing ids were removed, reaching a total of 22,149 breast cancer deaths.

For the calculation of incidence rates, all female patients whose primary diagnosis was breast cancer were selected from the 32,443,591 registries available in the discharges database. Additionally, other registries were included based on diagnoses associated with breast cancer for female patients who appeared in the death registry having breast cancer as the cause of death (see Appendix A for a detailed explanation).

Finally, after correcting inconsistencies and missing ids in the hospital discharge database, there were still 6,156 ids in the death register without an entry associated in the discharge database. In order to improve incidence estimations, a discharge was added to these deaths assuming that they had a uniform random survival time between one and five years. With this method, only 4,649 of the 6,156 ids had a discharge inside the 2002-2018 period and where therefore added to the discharges database. These discharges were only considered for the calculation of incidence and that other methods were tested, resulting in similar incidence rates.

The resulting discharge database for breast cancer patients consists of 149,181 registries, corresponding to 80,957 patients (some patients require several hospitalizations). Figures 1.1 and 1.2 summarize the inclusion and exclusion criteria for the death and discharge databases, respectively.

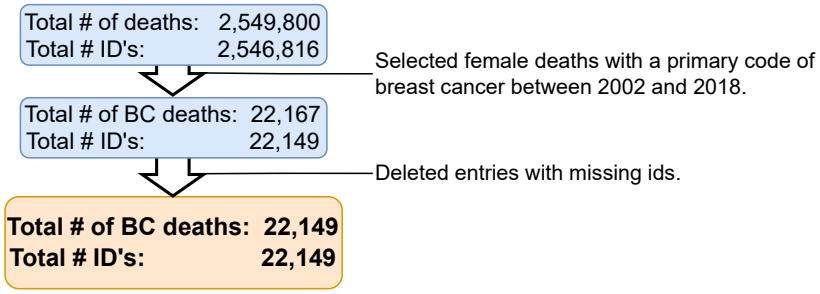

Figure 1.1: Inclusion and exclusion criteria for the breast cancer death database (2002–2018).

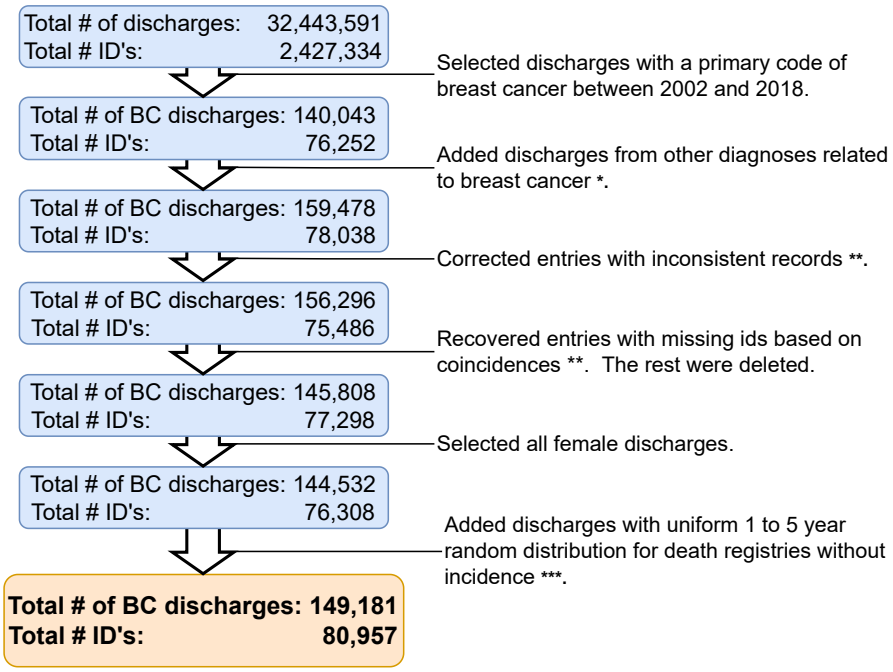

Figure 1.2: Inclusion and exclusion criteria for the breast cancer hospital discharge database (2002–2018).

$*$ Several diagnosis were included considering health problems that could arise due to breast cancer progression or its treatment, and therefore, those diagnosis were included under certain specific conditions. See appendix A for further explanation.
$**$ Inconsistent records and missing ids were processed based on the consistency of some characteristics such as gender, date of birth and region of residence. See appendix B for further explanation.
$***$ A discharge registry was created for all breast cancer deaths whose ids did not appear in the discharges database. This estimation was made assuming that these patients had an uniform random survival duration between 1 and 5 years.

1.2. Methods

Incidence and mortality rates were calculated, as crude rates (CRs) and age standardized rates (ASRs), for breast cancer at a country and region level. Population estimates and projections provided by the Chilean National Institute of Statistics (INE) [23, 24] were used. For the age standardized incidence and mortality rates, the standad population from the International Agency for Research on Cancer (IARC) was considered [25]. Incidence and mortality rates were also calculated for each year from from 2002 to 2018 by health insurer (public and private systems), by region of residence, and by age groups (crude rates).

The data provided by FONASA in its yearly statistical bulletin, which is only available online from 2009 onwards [26], was used to obtain the female population beneficiary of the public health system. Prior to this, FONASA bulletins had incomplete information, not segmented by age intervals or by gender. Therefore, linear regression was used to estimate 2002 to 2008 population from the available data.

Information from private health system beneficiary population (ISAPRE) was obtained using the data provided by the Health Superintendence in its yearly statistics publication for ISAPREs' beneficiaries [27].

For the survival analysis, the standard estimator of the survival function proposed by Kaplan and Meier [28] was used, namely the Product-Limit estimator, considering right-censored data. In our application, censored data consists of breast cancer patients that either died of other causes during the timeline under study, or survived from then on. In order to estimate survival, the variable of interest will be time-on-study, and not age, due to the similarity with respect to a clinical trial, this is, survival is measured from the time from the first breast cancer diagnosis to death.

With out loss of generality, time is measured in months, and therefore, the survival time was recorded as months difference, this is, a patient diagnosed by 1/1/2002 that died on 01/31/2002 has a 0 month survival, while a patient diagnosed the 1/31/2002 that died on 2/01/2002 had a 1 month survival. From the 80,957 resulting patients in the discharge database, there were 4,914 whose first diagnosis only registered the year of diagnosis. The month of diagnosis was randomly generated for such patients, using a uniform distribution during the year of diagnosis.

Finally, the Cox proportional hazards model [29] was used to study the effects of covariates. Although this model does not assume any particular *survival model*, it does assume that the effects of the covariates on the survival function are constant over time. Information concerning health insurance, age, year of discharge, region of residence and FONASA benefit segments were used as covariates. A variable indicating if the registry was before or after the implementation of the GES plan was also included. These variables were processed, transfor-

ming some of them into dummy variables, resulting in a total of 27 covariates from where to select those relevant for the model. Selection was based on the Akaike's information criterion [30] and on the p-value for each variable's significance. See Appendix C for a description of the model selection procedure.

1.3. Results

Table 1.1 presents a summary of the resulting new cases and deaths by year, with the average age and standard deviation, for the period under study. It can be observed that the number of new breast cancer diagnosis increased in 43.2 % from 3,771 in 2002 to 5,399 in 2018. Total breast cancer deaths increased in 47.3 % from 1,050 to 1,547 in the same period. There is a mild increasing trend in the mean age of death, with a 3.4 % growth, from 64.9 to 67.1, while there is no significant change in the mean age at diagnostic. It is important to take into account that the female population in Chile grew 19.3 % from 7,971,000 to 9,506,921 in the same period.

	New cases		Deaths	
Year	Cases	Mean age (std)	Cases	Mean age (std)
2002	3,771	57.8 (15.0)	1,050	64.9 (15.4)
2003	4,275	58.1 (14.4)	1,071	64.4 (15.9)
2004	3,595	58.9 (14.9)	1,096	64.6 (15.4)
2005	4,571	57.6 (15.2)	1,170	65.0 (15.6)
2006	4,062	57.3 (15.4)	1,147	64.8 (15.6)
2007	4,469	58.2 (15.1)	1,158	65.4 (15.6)
2008	4,256	58.5 (14.8)	1,228	65.3 (15.6)
2009	4,400	58.1 (14.6)	1,337	65.8 (15.6)
2010	4,778	58.7 (14.6)	1,298	64.9 (15.6)
2011	5,113	58.7 (14.5)	1,349	66.2 (15.5)
2012	5,275	58.1 (14.4)	1,371	66.4 (15.6)
2013	5,419	58.6 (14.3)	1,391	65.8 (15.6)
2014	5,197	59.0 (14.2)	1,424	66.3 (16.0)
2015	5,492	58.3 (14.0)	1,512	66.5 (15.6)
2016	5,588	57.9 (14.2)	1,492	66.4 (15.3)
2017	5,297	58.3 (14.1)	1,508	66.8 (15.3)
2018	5,399	57.9 (14.0)	1,547	67.1 (16.0)
Total	80,957	58.2 (14.5) 22,149	65.8 (15.6)	

Table 1.1: Breast cancer new cases and deaths by year for period 2002 – 2018.

1.3.1. Incidence

Figure 1.3 shows both crude and age adjusted incidence rates (new cases/100,000 women). CRs are higher than ASRs and have a growing trend over time, with an increase of 20.1 %, from 47.3 in 2002 to 56.8 in 2018. ASRs show a constant trend with an average rate of 42.3 and a standard deviation of 2.5 during the period under study.

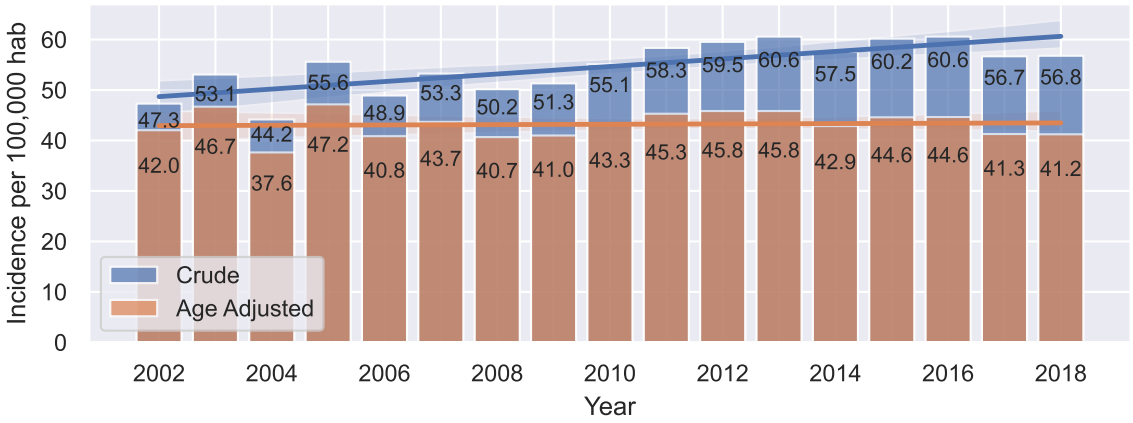

Figure 1.3: Crude and age standardized incidence rates by year (cases/100,000 women).

Figure 1.4 shows crude incidence rates by age groups for periods (2002–2007), (2008–2012), and (2013–2018). As expected, women younger than 29 years old and between 30 and 39 years old had the lowest incidence rates with average crude rates of 2.3 and 25.0, respectively. There are greater fluctuations in incidence rates over time for older age intervals than for younger age intervals, this may be because improvements in health care and prevention have a greater influence in older women's outcome.

There is a clear decreasing trend in the incidence rate for women older than 80 years old, evolving from having the highest rate, with a mean of 208.6 in the (2002–2007) period to being the third interval with higher incidence in the period (2013–2018) with a mean of 132.5 cases per 100,000 women. The latter is lower than the incidence for women in age group from 60 to 79 years old from any of the periods. A mild increasing trend can be observed for women aged 30 to 49 and women aged 70 to 79. There are no clear trends for women between 50 and 69 years old.

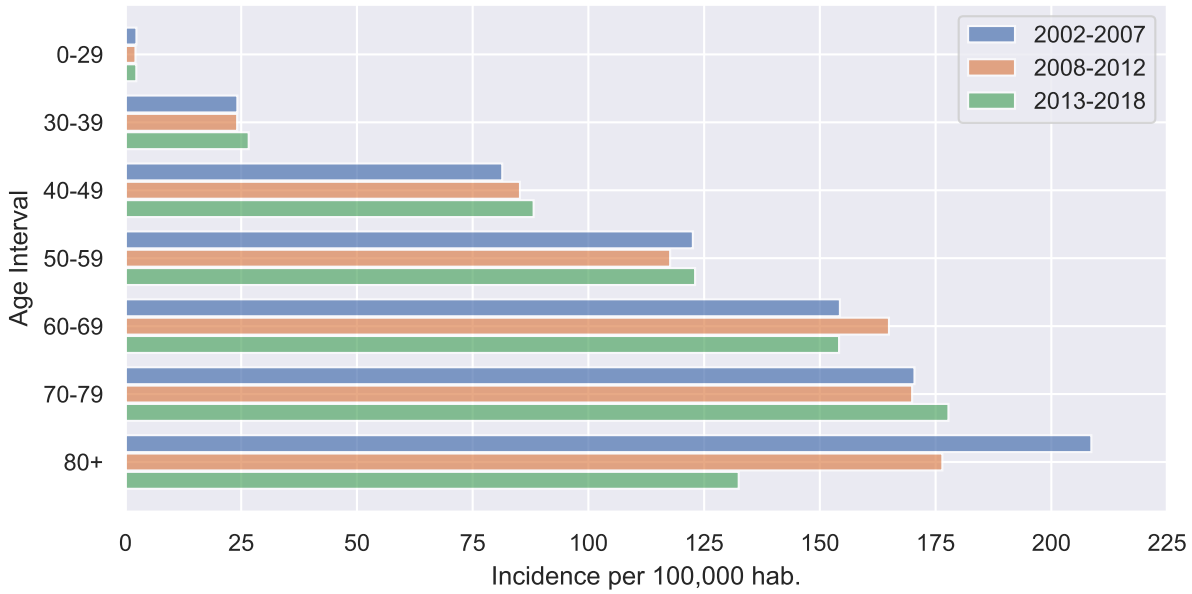

Figure 1.4: Mean crude incidence rates by age interval.

As shown in Figure 1.5, despite having considerable differences in incidence rates among regions, there is no clear geographical trend for the crude incidence rates. The higher incidences are found in the northern (Arica and Parinacota region) and in the center part of the country corresponding to Valparaiso region and Metropolitan area, with crude incidence rates of 67.7, 66.5 and 62.8 respectively. Interestingly, the regions with the lowest rates are also found in the northern area, Tarapacá and Atacama regions, with incidences of 34.3 and 36.9, respectively, followed by two southern regions, Los Lagos and Aysén, with crude incidence rates of 36.0 and 37.0, respectively. The rest of the regions have crude incidence rates between 39.6 and 57.1.

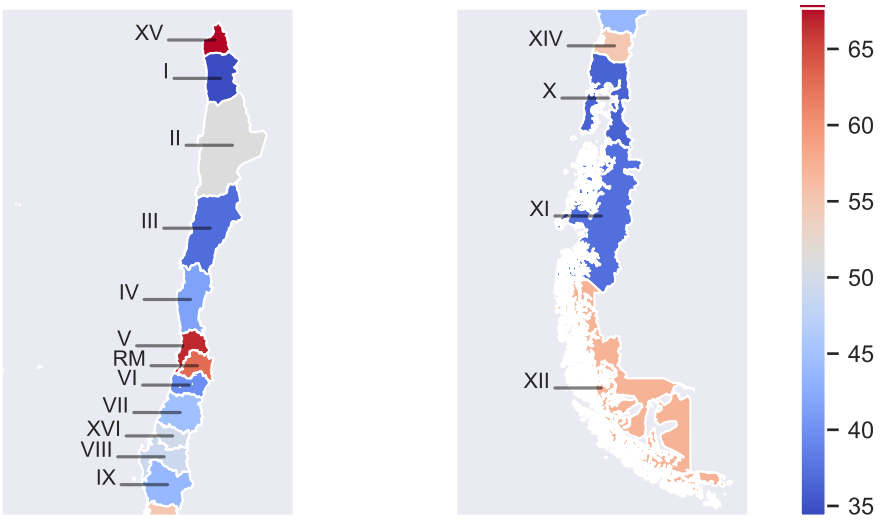

Figure 1.5: Map of mean crude incidence rates during (2002–2018) per 100,000 women by region.

Incidence values can be found in Table D.1 in Appendix.

Table 1.2 shows incidence rates by year for women affiliated to the private and public health insurance systems. For this analysis, only 69,374 out of the 80,957 registries were considered. This difference is due to discharges associated to other health care plans, from which 2,045 women belonged to the armed forces health insurance, 1,191 women had no health insurance, and 8,347 women had missing information. As digital registries started in 2001, in 2002 there is a much higher number of registries with missing health insurance information corresponding to the 44.5 % of that year's registries. Thus, mean and standard deviation were calculated without considering year 2002. Missing health care information corresponds to the 8.6 % of registries in the period 2003 to 2018 and has a decreasing tendency.

Women affiliated to a private provider (ISAPRE) have higher incidence rates in the period under study, having an average age adjusted incidence rate of 61.9, while women affiliated to the public provider (FONASA) have an average age adjusted incidence rate of 36.8. Neither of the systems had a clear trend (nor increasing nor decreasing). The private health system had a higher fluctuation, with a standard deviation of 9.8 in the age adjusted rates, while FONASA had a standard deviation of 1.8.

	ISAPRE		FONASA	
Year	Crude	Age adjusted	Crude	Age adjusted
2002*	20.5	28.8	30.9	28.0
2003	72.4	82.2	40.7	35.9
2004	47.2	51.8	39.2	33.0
2005	66.2	70.0	48.2	39.8
2006	58.0	62.3	43.5	35.2
2007	45.2	45.2	47.7	37.9
2008	42.2	42.3	46.3	36.3
2009	65.6	64.3	46.9	37.2
2010	69.4	68.8	47.9	36.0
2011	66.8	63.6	52.0	38.1
2012	65.3	61.9	53.0	38.7
2013	62.3	58.4	54.1	38.5
2014	68.8	63.9	50.6	35.1
2015	69.4	63.6	55.4	38.8
2016	81.0	72.1	53.7	37.2
2017	69.7	59.9	51.6	35.6
2018	70.9	59.8	52.1	36.2
Mean (Std)**	63.8 (10.6)	61.9 (9.8)	48.9 (4.8)	36.8 (1.8)

Table 1.2: Crude and age adjusted incidence rates for women with private and public health insurance.

∗ There is a high proportion of ids without health insurance information.
∗∗ Mean and std for 2003 to 2018.

1.3.2. Mortality

Figure 1.6 shows both crude and age adjusted mortality rates for the period under study. Crude mortality rates increased 23.5 %, growing from 13.2 in 2002 to 16.3 in 2018. However, age adjusted mortality rates decreased in 10.9 % during the same period, from 11.0 to 9.8.

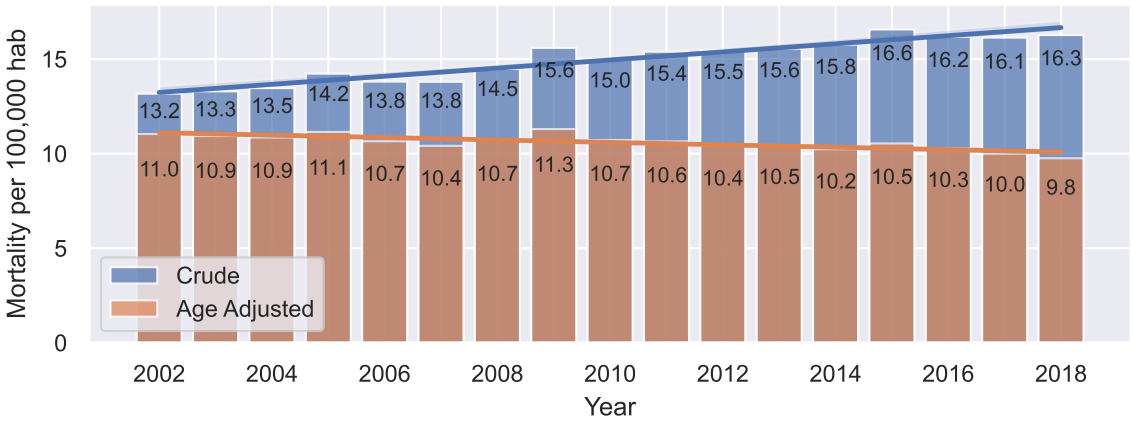

Figure 1.6: Crude and age adjusted mortality rates by year.

Figure 1.7 shows crude mortality rates by age group for periods (2002–2007), (2008–2012), and (2013–2018). There is a significant decrease in crude mortality rates in women older than 79, from 134.9 in (2002–2007), to 116.6 in (2013–2018). Women aged 50 to 69 years old experienced a smaller decrease in mortality. On the other hand, women aged 70 to 79 years old had a small increase in their mortality rates, from 59.7 in (2002–2007) to 63.4 in (2013–2018). Women younger than 50 years old did not experience any significant change in their mortality rates.

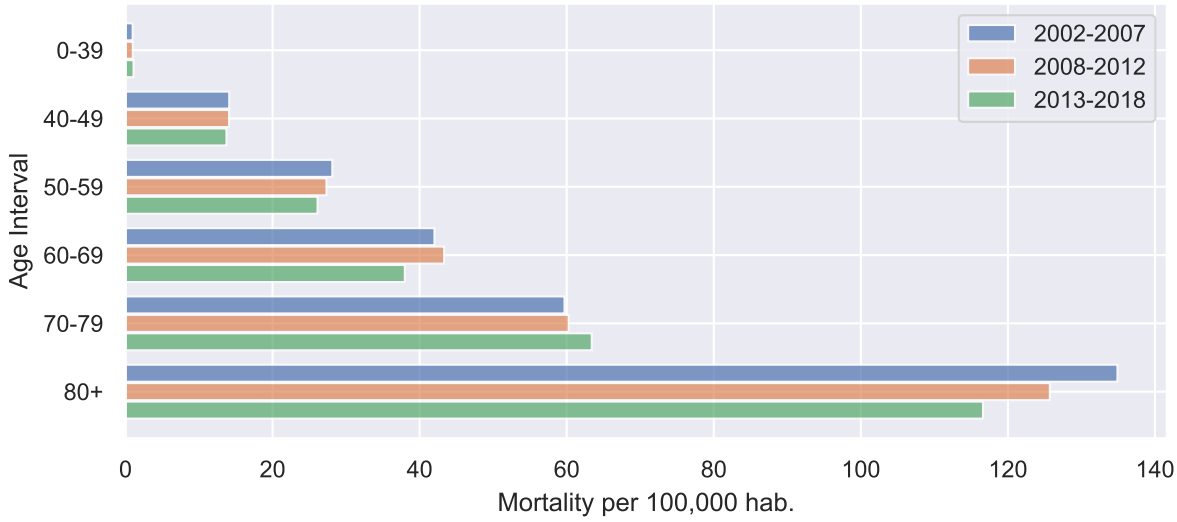

Figure 1.7: Crude mortality rate by age interval for the periods (2002–2007), (2008–2012) and (2013–2018).

Figure 1.8 shows that the southern region of Magallanes y la Antártica Chilena has the highest crude mortality rate of 19.7, followed by Valparaiso region with an average crude mortality of 18.6. On the other hand, the region with the lowest mortality rate is the southern

region of Los Lagos with a crude mortality rate of 11.3, followed closely by three northern regions (Atacama, Antofagasta and Tarapacá) with crude mortality rates of 11.4, 11.6 and 11.9 respectively. The rest of the regions have similar mortality rates, averaging between 12.1 and 16.0 deaths by 100,000 women.

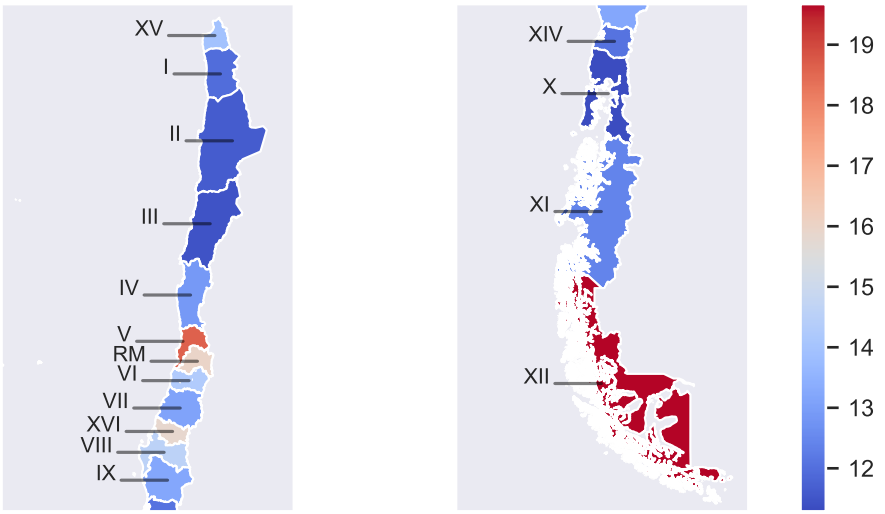

Figure 1.8: Map of mean crude mortality over the years 2002-2018 per 100,000 women by region.

Mortality values can be found in Table D.1 in the Appendix.

1.3.3. Survival rates

As discussed in the exclusion and inclusion criteria, the survival analysis will consider only 76,246 out of the 80,957 new cases registered during the period under study. The difference is due to the lack of data in the discharge database and inconsistencies such as records with discharge dates later than their death date.

1.3.3.1. Kaplan Meier estimations

The Kaplan Meier estimator was computed considering 76,246 observations, from which 63,591 were right-censored observations. Figure 1.9 shows the survival rates when considering all patients under the study period, and also separated by health care system. It is important to clarify that when referring to all patients under the studied period it is referred to the survival rate for Chile, this is, it refers to women affiliated to FONASA, ISAPRE, other health insurance systems and with no health insurance. The log rank test confirmed that the survival curves for the private (ISAPRE), the public health systems (FONASA) and all women (Chile) are statistically different ($p<0.001$).

The estimated one year survival rate 95 % confidence intervals are [0.918 ± 0.002] for FONASA, [0.967 ± 0.003] for ISAPRE patients, and [0.929 ± 0.002] when considering all women. The 95 % confidence interval for the five year survival rate are [0.781 ± 0.004], [0.890 ± 0.006], and [0.805 ± 0.003] for FONASA, ISAPRE and all women, respectively. This is, women in ISAPRE system have a five year survival rate 0.109 higher than women in FONASA.

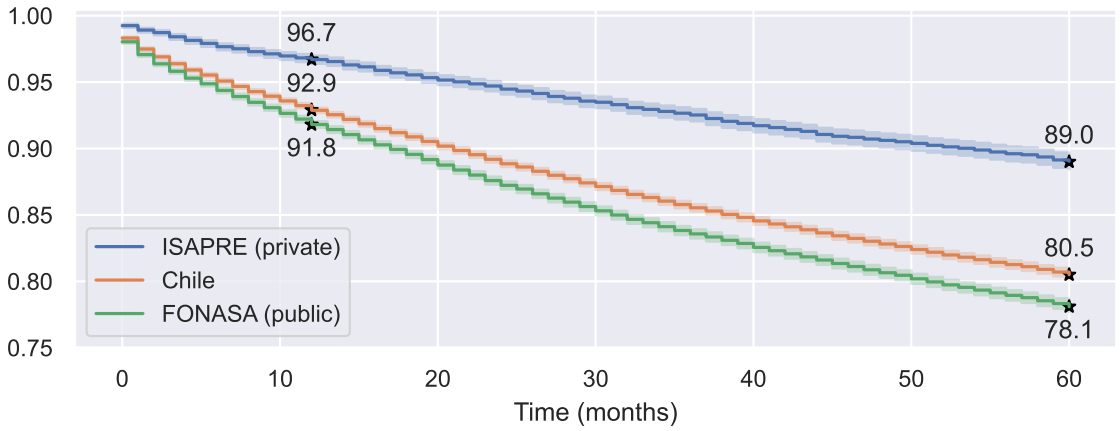

Figure 1.9: Kaplan Meier survival curves for Chilean women, and separated by health insurance system (public vs private).

A similar result if obtained when analyzing the survival curves for patients in each segment within the public health system shown in Figure 1.10. Women in the benefit range C and D have similar survival rates, which are significantly higher than those for women in FONASA segments A and B. Women in benefit range A have the worst survival of the four groups. Their 5-year survival rates are 0.82, 0.81, 0.78 and 0.75 for benefit ranges D, C, B and A, respectively. All these survival rates are lower than those for women in the private system.

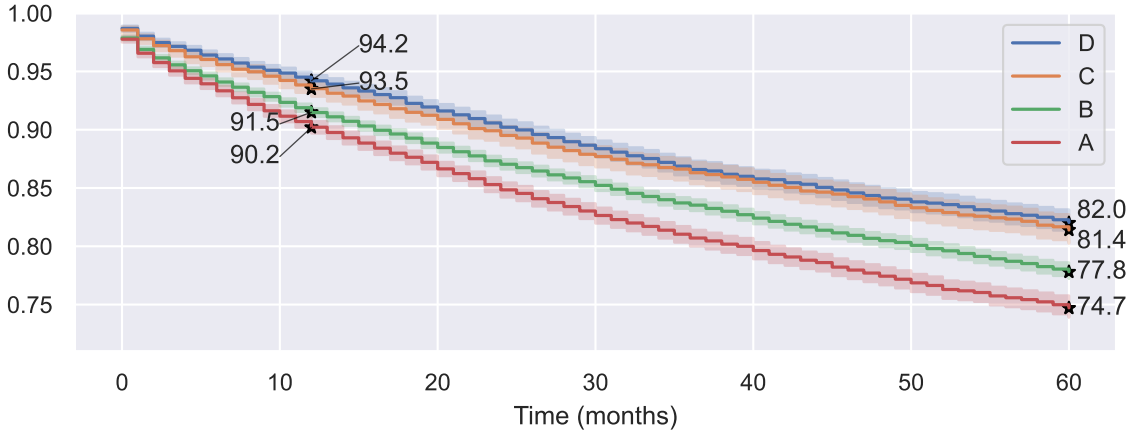

Figure 1.10: Kaplan Meier survival curves for FONASA patients by benefit range.

Figure 1.11 shows the survival curves as a function of year of diagnosis. The estimated 5 year survival for the population steadily increases over time, increasing from 0.71 for women diagnosed in 2002 to 0.85 for those diagnosed in 2014. This increment is not constant over time and seems to be slowing down in the last years.

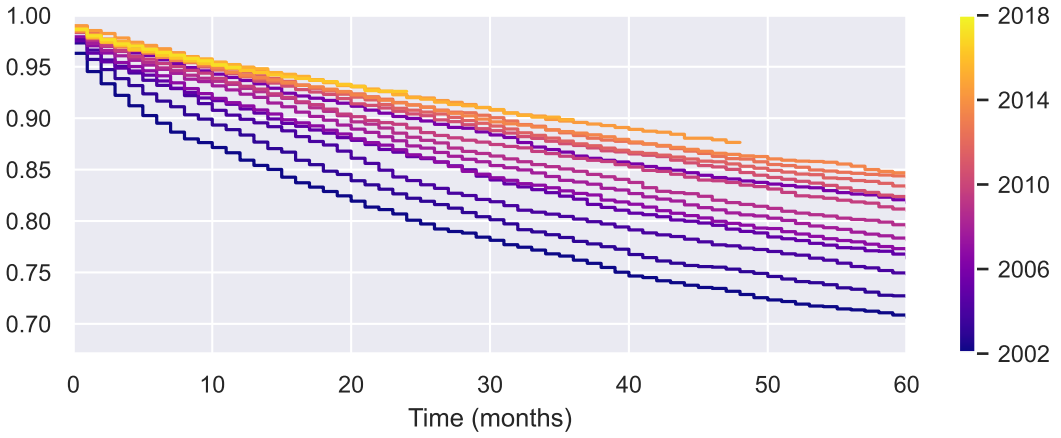

Figure 1.11: Kaplan Meier curve separated by year of diagnostic.

1.3.3.2. The Cox proportional hazards model

A Cox proportional hazards model with 11 selected variables out of 27 considered was optained using the procedure described in appendix C. The results are summarized in Table 1.3, where the first column contains the coefficient with its associated confidence interval and the second column shows the p value for the null hypothesis corresponding to equality of the base and the affected covariable.

The Cox model allows to evaluate the survival rate of a patient identified through these 11 selected covariates, such as dummy variables for private and public health insurance, benefit ranges A and B, age, squared age, year of diagnosis and the existence of GES at the time of diagnosis, among other variables. The Cox model also allows to compare the survival curves by modifying selected variables while maintaining the rest constant. For the results below, the variables that remain unmodified are set as the median of the data.

Variable	Coefficient	p value
Year of diagnosis	-0.049 ± 0.005	<0.001
Age	-4.782 ± 0.692	<0.001
Squared age	5.531 ± 0.578	<0.001
FONASA	0.194 ± 0.072	<0.001
ISAPRE	-0.347 ± 0.080	<0.001
FONASA beneficiary A	0.297 ± 0.054	<0.001
FONASA beneficiary B	0.100 ± 0.050	<0.001
GES	-0.130 ± 0.059	<0.001
Region RM	-0.126 ± 0.037	<0.001
Region XV	-0.457 ± 0.166	<0.001
Region VI	0.156 ± 0.085	<0.001

Table 1.3: Results for the Cox proportional hazard regression model.

Hazard ratios (or odds ratios) between two sets of characteristics can be evaluated through the Cox regression. Table 1.4 shows the hazard ratio for some specific variables. It can be observed that women of 40 and 50 years old have almost the same hazard, and that women of 60 years old have slightly higher hazard of dying due to breast cancer than women of 40 years old. The hazard ratios for 10 and 20 years of difference increase for older ages, thus the hazard of 60 year old women is 1.14 times that of 50 year old women and the hazard of 70 year old women is 1.46 times that of 50 year old women. The hazard of FONASA beneficiaries from the benefit range C-D is 1.72 times that of ISAPRE beneficiaries. FONASA patients form benefit ranges B and A have even higher hazard, being 1.11 and 1.35 times higher than that of FONASA patients from the benefit range C-D, respectively. The presence of the GES plan also produces a difference in hazard ratios, having women without the GES a hazard 1.14 times higher than that of women with GES. The presence of GES also influences the hazard progression through the years of diagnosis. Women diagnosed on 2018 have less than half the hazard (0.4) than a women diagnosed on 2002, this due to the variables year of diagnosis and presence of GES.

Coefficient	Hazard ratio
Age 40 → Age 50	1.02
Age 40 → Age 60	1.16
Age 50 → Age 60	1.14
Age 50 → Age 70	1.45
ISAPRE → FONASA Beneficiary C-D	1.72
FONASA Beneficiary C-D → FONASA Beneficiary B	1.11
FONASA Beneficiary C-D → FONASA Beneficiary A	1.35
With GES → Without GES	1.14
Year 2002 → Year 2004	0.91
Year 2004 → Year 2006	0.80
Year 2002 → Year 2018	0.40
Region RM → Region XV	0.72
Region RM → Region VI	1.33
Region RM → Any other region	1.13

Table 1.4: Hazard ratios obtained through the Cox model.

Figure 1.12 presents the Cox survival curves for women affiliated to public and private health insurance systems. The survival rates have a similar behavior than those obtained by the Kaplan-Meier estimations, with a survival curve for the patients in the private system that is significantly better than that for patients affiliated to the public health insurance. The Cox regression estimates the one year survival rates to be 0.965 and 0.940 for the private and public sector, respectively, on the other hand, the Cox regression estimates the five year survival rates to be 0.901 and 0.836, respectively. The Cox survival rates for ISAPRE patients are similar to those found by the Kaplan Meier estimations; the one year Cox survival rate lies in the confidence interval found with the Kaplan Meier method. The difference between the survival rates estimated by the Cox regression and the Kaplan Meier method is greater for women in FONASA than for women in ISAPRE. The survival rates for FONASA affiliates calculated by the Cox regression are 0.022 and 0.055 higher than those estimated by Kaplan Meier method for the one year and five year survival rates, respectively. Then the difference in five year survival rates between ISAPRE and FONASA patients calculated by the cox regression is of 0.06.

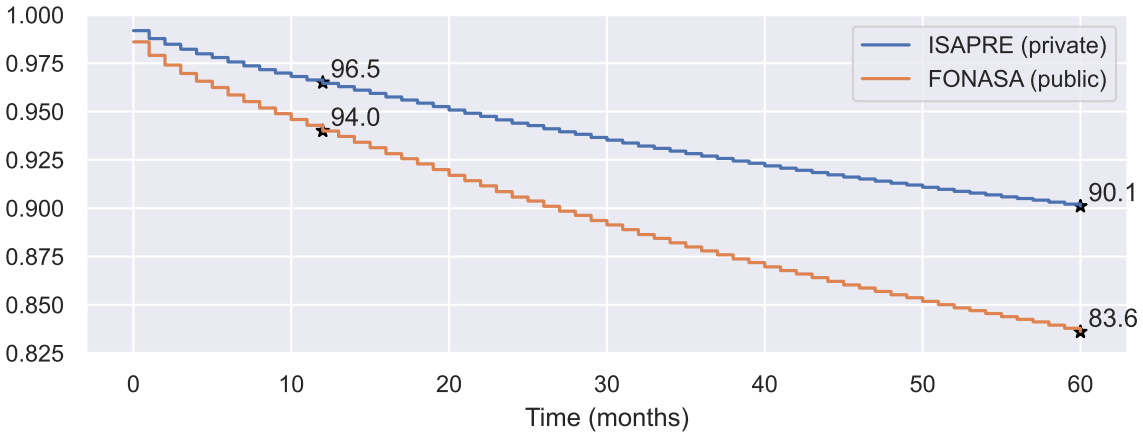

Figure 1.12: Cox survival curves adjusted by health insurance: public (FONASA) vs private (ISAPRE) health plans.

Figure 1.13 shows us the particular effect of the GES program. The one year survival of a women passes from 0.932 to 0.94 when evaluating her outcomes without and with the GES plan, respectively. The five year survival rate for a women without GES is 0.816, meanwhile the same rate for a women with GES is 0.836.

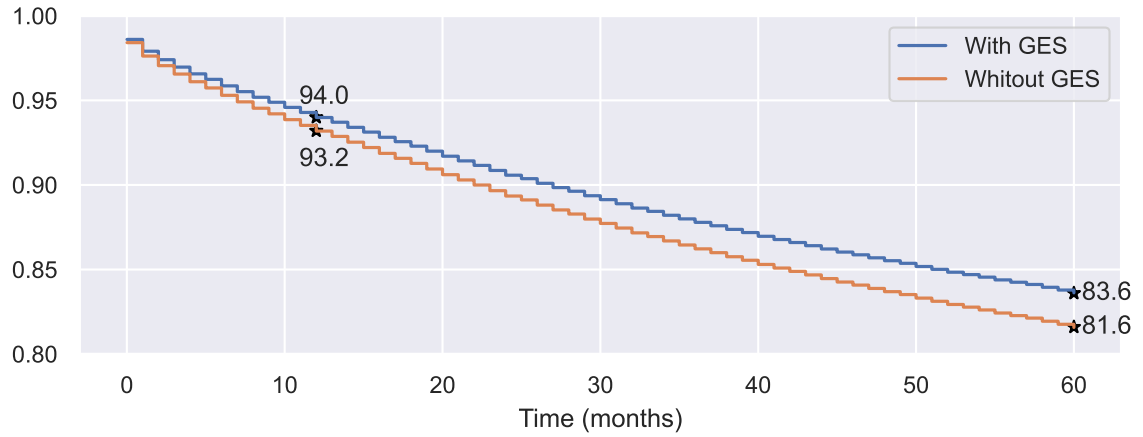

Figure 1.13: Cox survival curves adjusted by presence of GES: With GES vs Without GES.

Figure 1.14 shows the survival curves for different ages for both health insurance systems. The difference in survival rates between ISAPRE and FONASA patients increases for older people. The five year survival rates for women aged 40 are 0.91 and 0.85 for ISAPRE and FONASA systems respectively, this is a survival difference of 0.06. On the other hand, the five year survival rates for women aged 80 are 0.82 for ISAPRE and 0.71 for FONASA, with a survival difference of 0.11. The five year survival rates are shown in table 1.5

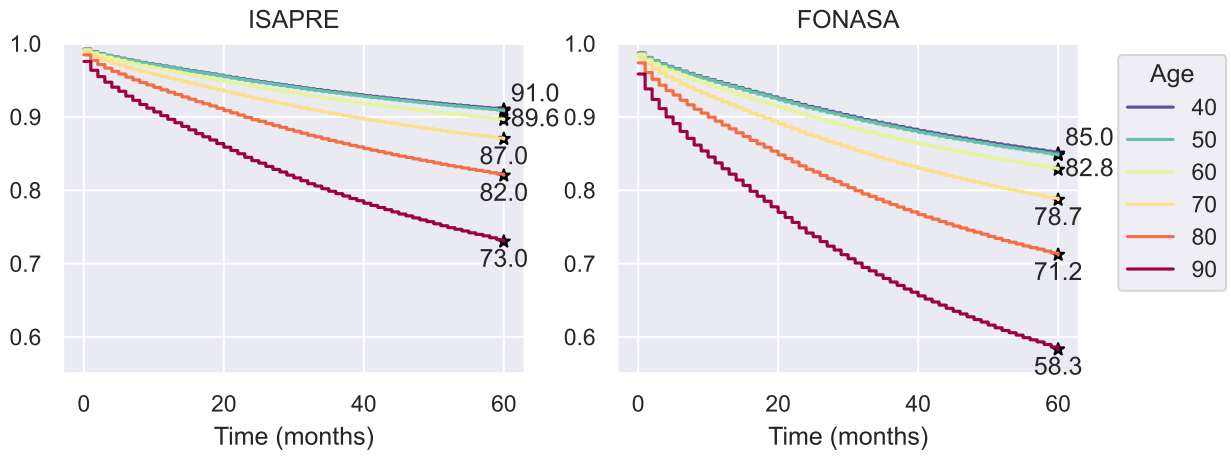

Figure 1.14: Survival curves obtained by the Cox regression, adjusted by age, for each health insurance.

Age	40	50	60	70	80	90
ISAPRE	0.91	0.908	0.896	0.870	0.820	0.730
FONASA	0.85	0.848	0.828	0.787	0.712	0.583

Table 1.5: Five year survival rate predicted by Cox model by age and health care system.

Chapter 2

Assessing the public-private survival gap

The following section will aim to evaluate the differences in survival produced by the early detection policies from the public and private health care systems.

In order to do that the progression of breast cancer will be simulated, evaluating the outcomes of such simulation over different screening policies.

2.1. Methods

2.1.1. Breast cancer natural progression

This section briefly presents a breast cancer tumor growth model developed by Plevritis et al. [31] which is used as the basic model for the natural evolution of the disease. The model allows to obtain transition probabilities between cancer stages, as well as to detected stages. There are different kinds of tumor classifications that use characteristics such as tumor size, lymph nodal involvement, hormonal receptors and metastatic state. A classification based on lymph nodal involvement will be used, where a tumor is said to be in *local* stage if the cancer is confined within the breast; *regional* if the lymph nodes, primarily in the armpit, are involved; and *distant* if the cancer is found in other parts of the body as well. Besides this, ductal carcinoma in situ (*DCIS*) is considered a non invasive tumor stage. This last stage will be considered in the following section. In what follows, the following notation will be used:

- $V(t)$: Volume of the tumor at time t.
- V_0: Initial volume of the tumor. The modeling of the natural progression of the disease is measured from the moment the tumor reaches this size V_0.
- DT: Volume doubling time, i.e., time that takes for a tumor to double its volume.
- $o(t)$: Little-o notation. $\lim_{t\to\infty} \frac{o(t)}{t} = 0$.
- T_{det}: Random variable representing the time of clinical detection. The latter is defined as the ability to feel a tumor through palpating the breast, or the manifestation of other symptoms such as swelling, skin irritation, or pain.
- T_{reg}: Random variable representing the time for the tumor to evolve from local to regional stage.
- T_{dist}: Random variable representing the time for the tumor to evolve from regional to distant stage.
- R: Gamma distributed random variable representing the inverse growth rate of a tumor. The higher the value of R, the slower the progression of the tumor.

The continuous tumor growth model proposed by Plevritis et al. [31] describes the natural history of breast cancer and considers invasive tumors in local, regional and distant stages. The model assumes that tumors are spheroidal and grow exponentially with a constant volume doubling time. Thus, tumors grow from an initial volume $V_0 = \frac{4}{3}\pi\ mm^3$, corresponding to a sphere of diameter 2 mm. Thus, the volume for a tumor at time t is given by equation (2.1), where the inverse growth rate R is a gamma distributed random variable with shape and size parameters α and $1/\beta$, respectively, which are empirically estimated. Plevritis et al. [31] notes that R is such that $\mathbb{E}(R) = \alpha/\beta$ and the tumor volume doubling time is given by $DT = \ln(2) \cdot R$.

$$V(t) = V_0 e^{t/R} \tag{2.1}$$

This model assumes that the time of clinical detection T_{det}, measured from the moment the tumor volume reaches the value V_0, depends on the current tumor volume through a hazard function as shown in equation (2.2).

$$\mathbb{P}(T_{det} \in [t, t+dt) | T_{det} > t) = \gamma V(t)dt + o(dt) \tag{2.2}$$

The model also assumes that tumors start at a local stage with a volume of V_0. The time that it takes to evolve from local to regional stage, T_{reg}, is expressed by (2.3).

$$\mathbb{P}(T_{reg} \in [t, t+dt) | T_{reg} > t) = \eta V(t) dt + o(dt) \tag{2.3}$$

Similarly, the disease may progress to a distant stage only from a regional stage, and this occurs at a random time T_{dist}. Its hazard function is expressed in equation (2.4).

$$\mathbb{P}(T_{dist} \in [t, t+dt) | T_{dist} > t, T_{reg} = t_{reg}) = \begin{cases} \omega V(t) dt + o(dt) & t > t_{reg} \\ 0 & t \leq t_{reg} \end{cases} \tag{2.4}$$

Similar growth models to the one developed by Plevritis et al. [31] can be found in [32, 33, 34, 35]. The parameters $(\gamma, \eta, \omega, \beta, \alpha)$ were estimated by maximum likelihood methods using data provided by the Surveillance, Epidemiology and End Results (SEER) program . This dataset contains registries of the tumor size and stage of breast cancer for female patients who were clinically detected with invasive tumors between 1975 and 1981 [36].

The likelihood function used by Plevritis et al. [31] does not change its value when the first four parameters are escalated by a constant. Therefore, they maximize the likelihood constraining to $\alpha = \beta$, and as a consequence, $\mathbb{E}(R) = 1$. For further details on the modeling and the source of parameters please refer to [31].

Lets recall that the value of R is related to the rate at which tumors grow. Plevritis explains that in order to modify the scale of time at which the events occur, the expected value of R can be modified. This produces a re escalation of the original parameters.

The assumption of $\mathbb{E}(R) = 1$ will be modified following Plevritis' advice. By doing this, the natural progression found in [31] remains unmodified, but the time scale at which these transitions occur is changed. As stated above, the expected value of R is related to the tumor doubling time by $\mathbb{E}(DT) = \ln(2) \cdot \mathbb{E}(R)$. The average tumor volume doubling time calculated by MacInnes et al. [37] of 167 days with 95 % CI [151,186], this is, 0.469 years 95 % CI [0.424,0.522]. Then, the expected value of R is modified to $\mathbb{E}(R) = 0.469/\ln(2)$ and the parameters found by Plevritis were escalated as specified in appendix F. A consequence of this escalation is that the time of this transitions is measured in years.

Using the tumor growth model developed above, a stochastic process will be formulated. In order to obtain this stochastic process, the above equations must be reinterpreted as probabilities of passing from one cancer stage to other in a time dt . States C_1, C_2 and C_3 correspond to the presence of breast cancer in local, regional and distant stages, respectively. Similarly, states D_1, D_2 and D_3 represent breast cancer tumors detected in local, regional and distant stages, respectively. Thus, a continuous non homogeneous Markovian process is obtained, defined by its Q matrix shown in equation (2.5).

$$Q(t) = \overbrace{\begin{bmatrix} (-\gamma-\eta)V(t) & \eta V(t) & 0 & \gamma V(t) & 0 & 0 \\ 0 & (-\gamma-\omega)V(t) & \omega V(t) & 0 & \gamma V(t) & 0 \\ 0 & 0 & -\gamma V(t) & 0 & 0 & \gamma V(t) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}^{\begin{matrix} C_1 & C_2 & C_3 & D_1 & D_2 & D_3 \end{matrix}} \tag{2.5}$$

Given that the Q matrix can be factorized as $Q(t) = V(t)W$, an explicit expression for the transition probabilities can be obtained, namely the matrix $P(t, \Delta t)$, where element $[P(t, \Delta t)]_{i,j}$ is the probability of being in state j at time $t + \Delta t$, given that at time t the disease is on state i. For details see [38].

$$\begin{aligned} P(t, \Delta t) &= e^{\left(\int_t^{t+\Delta t} V(\tau)d\tau\right)W} = e^{\bar{V}(t,\Delta t)W} \\ &= Ae^{\bar{V}(t,\Delta t)D}A^{-1} \end{aligned} \tag{2.6}$$

where $\bar{V}(t, \Delta t) = \int_t^{t+\Delta t} V(\tau)d\tau = V_0R(e^{(t+\Delta t)/R} - e^{t/R})$ and $W = ADA^{-1}$. The explicit values of this transition probabilities matrix as well as details on the parameters used can be found in Appendix F.

This explicit probability matrix allows the computation of transition probabilities for any time interval. Notice that this stochastic process starts when a tumor reaches the local stage. This is important because as it is a non-homogeneous Markov chain, the transitions not only depend on the width of a time interval but also on the current time.

As the states C_1, C_2 and C_3 are transient, and D_1, D_2 and D_3 are absorbent, notice that for a big enough Δt the matrix $P(t, \Delta t)$ will reach a steady state distribution that is concentrated in the D_1, D_2 and D_3 components. Then, this steady state D_1, D_2 and D_3 components provides us with the probability of detecting the tumor in stages local, regional and distant, respectively. In particular, notice the importance of the first row when evaluating the matrix $P(0, \Delta t)$ for a big enough Δt, for this is the probability distribution of detecting a tumor that started in state C_1.

The current model was preferred to others for integrating tumor growth with stage progression (local, regional, distant) in a population that is not undergoing breast cancer screening. The presence of stages was desired because it provides us with a direct way to estimate survival rates. For example, an alternative model by Plevritis, Sigal, Salzman, Rosenberg and Glynn [33], which also includes nodal progression through local, regional and distant stages, was not considered because it includes a parameter independent of time in the hazard fun-

ction that introduces a great degree of difficulty on finding an explicit form of the transition matrix P, without major improvements in the goodness of fit.

2.1.1.1. Adjustment to Chilean records

As mentioned above, the parameters of the natural history of breast cancer were estimated using the SEER database with population from the USA. In this data tumors were staged 50 % of local, 45 % regional, and 5 % distant at the moment of diagnosis. This stage distribution corresponds to data from 1975 to 1981 when there was a negligible level of screening mammography. A study by Prieto for the Chilean public system [39] reports for the year 2000 a distribution of 44.8 %, 35.3 % and 19.9 % for local, regional and distant tumors, respectively. The mean distribution for years 2000 to 2003 is 60.9 %, 27.8 % and 11.3 %, respectively. Before the implementation of GES in 2004, there was a negligible level of preventive mammographies in Chile, and therefore the tumors detected in this period correspond symptomatic detection. Although the reality of USA between 1975 to 1981 and Chile between 2000 to 2003 agree in having little to non screening mammographies, there are other factors, where the two countries do not match, that influence the detection of breast tumors such as education and the population's awareness of the disease and its consequences, among others.

This difference between the observed tumor stage distribution in USA and Chile might imply that the modeled natural progression will most likely not resemble the Chilean case. Although differences have been found in incidence rates and risk factors between Hispanic and non-Hispanic women in the US [40], it will be assumed that the major difference between the tumor stage distributions reported by Prieto and SEER come from preventive measures, such as self examination and educational campaigns. Therefore, this difference will be reduced by modifying the γ parameter. The γ parameter reflects the rate at which women self-detect their tumors, and therefore, changes the transitions of the $P(t, \Delta t)$ matrix.

Therefore, a parameter γ^* that reflects the Chilean reality was estimated. This is estimated by minimizing the absolute error between the stage distribution from [39] and the stage distribution from $P(0, \Delta t)$ for a big enough Δt, so that it reaches the steady state distribution. This minimization offers two options, to do it with respect to the distribution reported by Prieto for the year 2000, or to consider the mean of the years 2000 to 2003. The year 2000 could be a better reference, for there was less prevention, but to take a single year makes the measure subject to errors. Both options were implemented. When minimizing the absolute error with respect to the 2000 distribution, a γ^* parameter of $\gamma_1^* = \exp(-9.429)$ was achieved. This way, the steady state distribution for detected states using such γ_1^* are 44.8 %, 48.14 % and 7.06 % for local, regional and distant stages, respectively. On the other hand, when minimizing with respect to the mean of the years 2000 to 2003, a γ^* parameter of $\gamma_2^* = \exp(-8.775)$ was achieved. This γ_2^* parameter generates a steady state distribution for detected states of 60.94 %, 36.29 % and 2.77 % for local, regional and distant stages, respectively.

Both methods tend to adjust the distribution to the proportion of women detected in local stages. It is important to notice that women diagnosed in the regional and distant stages have higher impact on the survival outcomes, especially the distant stage, whose survival may be even a third of the survival of local stages [41]. The authors chose to use the parameter found when adjusting to the distribution of stages reported for the year 2000 for it produces higher values for the distant stage, which is closer to the Chilean reality.

2.1.2. Breast cancer model including screening mammograms

In Subsection 2.1.1, the natural evolution of breast cancer without any intervention but the possibility of clinical detection is described. Furthermore, it considered the tumors evolution from the time it reaches a initial volume, V_0. In this subsection, a discrete Markov process is presented, which integrates the natural evolution of the disease with screening programs, the inclusion of death by other causes, and the transition probabilities from healthy to sick stages. The following additional notation will be used.

- $(Z_n)_{n\geq 0}$: Discrete non homogeneous Markov process representing the health state of a woman.
- $\rho(n)$: Probability of a woman to die by a cause different than breast cancer at age n.
- $\phi_0(n)$: Probability of a healthy woman to develop breast cancer at age n. It can be estimated using breast cancer incidence rates.
- $\phi_{0,1}$: Probability that a breast cancer in DCIS stage evolves to a local stage.
- λ: Probability of a women with breast cancer in DCIS stage to detect such tumor through clinical detection.
- $\hat{N}$: Random variable representing the age in months, when a woman first reaches state C_1.
- q_0: Sensitivity of a mammography for a tumor at stage DCIS.
- $q(V)$: Sensitivity of a mammography for an invasive tumor of volume V.
- P: Transition probability matrix from the disease model.
- N_0: Initial age in months. All women start the process at this age.
- N_f: Maximum age in months. A women stops the process if she reaches this age.

In order to integrate the natural disease evolution with incidence, mortality by other causes and breast cancer screening mammograms, a discrete time finite state stochastic process $(Z_n)_{n\geq 0}$ is considered, representing the health state of a woman. Some transitions may depend

on the age of the patient and, therefore, the process corresponds to a non-homogeneous Markov chain.

- **Non-homogeneous Markov chain states :**
  - H: Healthy state; women do not have breast cancer.
  - C_0: Ductal carcinoma in situ (DCIS), which corresponds the first stage in the progression of breast cancer.
  - C_i: Invasive cancer states according to those described in section 2.1.1, $\forall i = 1 - 3$.
  - D_i: Diagnosed states. Each of these states considers patients in state C_i that have been diagnosed due to screening or by a clinical exam. Patients who enter these states are considered to begin an appropriate treatment, and therefore, their survival rates can be estimated. $\forall i = 0 - 3$.
  - R: Removed state, where patients move when dying for causes different than breast cancer or reaching the time horizon.

  It is important to notice that H and $C_i, \forall i$, are transient and R and $D_i, \forall i$, are absorbent states. Also notice that there is no "death by cancer"state, because it is assumed that nobody dies due to breast cancer without being previously diagnosed. This is a reasonable consideration for countries with relatively good health coverage such as Chile where this could happen only to a negligible amount of women.

- **Transition epochs:** In real life, transitions between states can take place at any time during the evolution of the disease. Therefore, although mammograms screening decisions usually take place on a yearly basis, it is closer to reality to use shorter periods of time to consider, for example, that a tumor might evolve from local to regional and later to distant in less than a year. Monthly epochs will be used for considering them small enough to capture the nature of transitions. Nevertheless, to choose months as time intervals is arbitrary and can be changed to smaller intervals such as weeks. Thus, the process (Z_n) is indexed by age in months. Each person starts the process at age of N_0 months and continues until she dies, gets a breast cancer diagnosis or reaches the end of the time horizon N_f months.

  Without lost of generality, the following sequence of events during each epoch (month) will be defined: i) at the beginning of each month, a person might reach the time horizon N_f or die due to causes different than breast cancer and transition to R state. ii) If not, and she is in H state, she might develop DCIS and evolve to C_0. iii) Subsequently, a decision is taken of whether or not a screening mammogram is performed. Women in cancerous states might pass to detected states through screening detection, but the transitions of women in healthy states is not affected by screening. It will be assumed that there might be false negatives, but no false positives. Further details on screening and its sensitivity are given below. iv) If the person is in some cancer state that was not detected through screening, she might pass to a detected state due to the progress of the cancer to a clinical stage or pass to a higher cancerous state.

  All progressions to detected states and to higher stages occur only if the person did not passed to the R state. This sequence of events is shown in Figure 2.1

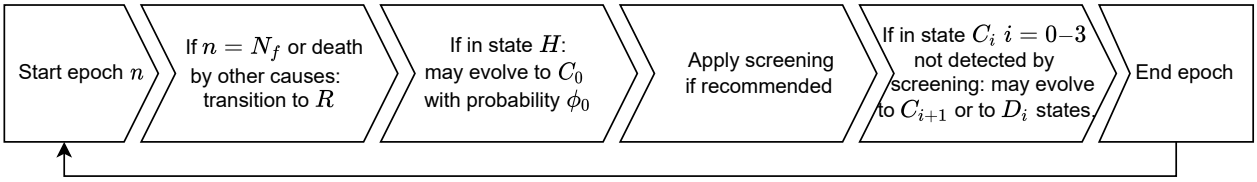

Figure 2.1: Diagram representing the sequence of events within an epoch.

- **Screening mammograms:** Mammograms are considered as the standard breast cancer screening test. When the mammogram is suspicious for a malignant tumor, further exams (ultrasounds and biopsies) are performed to confirm or not the diagnosis. Thus, there might be false negative results, but there are not false positives results. At the beginning of each epoch, a decision is made of whether to perform a test on a certain person. This will be denoted by $M(n) \in \{0, 1\}$, this is, $M(n) = 1$ if a woman is going to have a mammogram at the age of n; and $M(n) = 0$ if a woman is not going to have a mammogram at the age of n.

  Mammograms have a sensitivity of q_0 for diagnosing DCIS. Sensitivity for invasive tumors is considered to depend on the tumor volume, this is $q(V)$. It is important to remember that the volume in the disease model is measured from the instant the tumor first reaches local cancer state C_1 and that the time in the disease model is measured in years. Let $\hat{N}$ be the instant, measured as age in months, when a women first reaches the C_1 state, i.e., $\hat{N} = \inf\{n \geq 0 | Z_n = C_1\}$. Then, the volume of an invasive tumor at time n is $V((n-\hat{n})/12)$, with V as in equation (2.1). Once the volume is calculated, the sensitivity of the mammographies can be evaluated by $q(V)$.

- **Transitions:** Every person in states $\{H, C_0, ..., C_3\}$ has a probability of $\rho(n)$ of dying by a cause different than breast cancer and pass to state R. Additionally, a healthy person has a probability of $\phi_0(n)$ of developing DCIS and moving to C_0 state. This probability is a function of the disease incidence rate at age n. It is important to notice that both ρ and ϕ_0 depend on the persons age, and therefore, on time.

  All transitions between cancerous to detected states are affected by the screening policy. A person in C_0 state, given that she did not died from other causes and that was not screened, has a probability λ of advancing to a detected state D_0, and a probability $\phi_{0,1}$ to progress to C_1. For simplicity, parameters λ and $\phi_{0,1}$ where considered as constants.

  Similarly to the volume calculation for the mammogram's sensitivity, notice that in order to make use of the transition matrix P from the disease model from Section 2.1.1 the instant $\hat{N}$ must be considered. Then, transitions between cancerous states $C_i,\ i = 1, 2, 3$ and detected states $D_i,\ i = 1, 2, 3$ from time n to $n+1$, given that she did not died from other causes and that was not screened, is determined by the stochastic matrix $P(t, \Delta t)$ with $t = (n - \hat{N})/12$ and $\Delta t = 1/12$. Thus, for $X, Y \in \{C_1, C_2, C_3, D_1, D_2, D_3\}$

$$\mathbb{P}\Big(Z_{n+1} = Y | Z_n = X, \hat{N} = \hat{n}, M(n+1) = 0\Big) = P\Big((n - \hat{n})/12, 1/12\Big)_{X,Y}$$

  The above is well defined since the transitions between cancerous states $C_i,\ i = 1, 2, 3$ and the detected states $D_i,\ i = 1, 2, 3$ can take place only after a woman has reached

the local tumor stage, this is, such transitions can only be evaluated for instants n such that $\hat{N} < \infty$ and $n > \hat{N}$.

Figure 2.2 summarizes the process described above.

This modelling allows for explicit transition probabilities. The transition probabilities matrix can be found in appendix G.

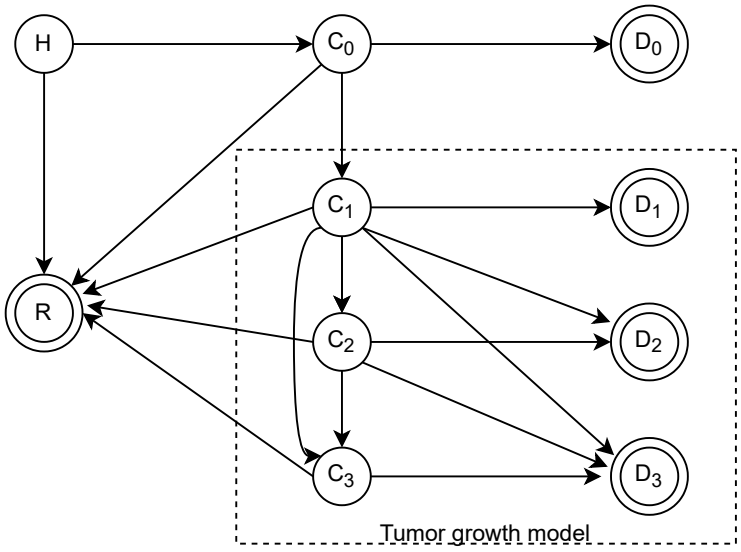

Figure 2.2: Diagram of the disease simulation process.

Arrows represent transition probabilities. For simplicity, transitions to stay in the same state were omitted.

2.2. Computational experiments

In this section, computational experiments are designed and applied to study the impact of breast cancer screening policies on the etapification, and therefore on survival, of diagnosed patients. As was seen in section 1.3.3, there are significant differences in the survival rates of Chilean women with breast cancer, depending on health insurance, which motivated the current study to determine the weight of screening in such differences.

All the computational studies are carried out in Python environment on a Dell desktop (Core 2.80GHz and RAM 8GB).

2.2.1. Screening policies

In Chile there is no screening program as defined by WHO [11]. Currently, the ministry of Health provides financial support for the realisation of 4 mammograms every 3 years between the ages of 50 and 59 [42]. However, what screening is following each woman in Chile is unknown, and is highly dependent on their health care system. Data shows that women in the private system get more preventive mammograms than those in the public system. Table 2.1 shows the amount of screening mammograms in each system by age interval, elaborated with data from the CASEN survey [12].

Age	[35,40)	[40,45)	[45,50)	[50,55)	[55,60)	[60,65)	≥65
FONASA	34.7	54.9	65.4	71.3	70.2	68.7	40.2
ISAPRE	52.4	70.4	80.2	80.4	82.3	75.4	65.9

Table 2.1: Percentage (%) of women that had a mammography in the last three years, by age and health care system. Source: CASEN survey 2017 [12].

Five screening policies will be studied: the first three consider that all women comply with the recommendation, while the others consider that women have a probability of having a test in the recommended date. The screening policies are as follows:

1. **No screening**: This is the base case scenario where no mammograms are performed.
2. **GES**: This policy emulates the preventive mammograms recommended in the GES program. As this guideline has changed over the years, the current policy was chosen, in place since 2013, which instructs mammograms every 3 years between the ages 50 to 59 [42]. It is assumed that all women follow this guideline.
3. **US screening**: This policy apply the U.S. Preventive Services Task Force (USPSTF) screening guidelines. It will be assumed that all women follow this guideline. This guideline instructs to perform mammographies every two years from the age of 50 until 74 [43].
4. **Random FONASA**: Random Bernoulli assistance replicating FONASA. It will be considered that each year women randomly chose to perform a mammography with a probability p_F. This probability depends on her age and will be taken such that it replicates the behaviour shown on the first row of table 2.1. Further details are given below.
5. **Random ISAPRE**: Random Bernoulli assistance replicating ISAPRE. It will be considered that each year women randomly chose to perform a mammography with a probability p_I. This probability depends on her age and will be taken such that it replicates the behaviour shown on the second row of table 2.1. Further details are given below.

Women will be simulated under these policies in order to obtain the etapification of diagnosed women.

It is important to determine an appropriate amount of simulations to obtain confident results. The sample size formula for infinite population will be used $n = \frac{t^2 \cdot p(1-p)}{e^2}$, where e is the margin error, t the critical value of the normal distribution for a desired confidence level and p the sample proportion [44]. As the objective is to obtain the distribution for more that 2 sub populations whose proportions p are unknown, the worst case will be considered, which is $p = 0.5$. Therefore, 9604 women diagnosed by cancer are needed to reach a error of 1 % with a confidence level of 95 %. Thus, there will be as many simulated women as needed to obtain such number of women diagnosed by cancer.

Once a policy is simulated the necessary amount of times to obtain a etapification, a survival rate will be estimated based on such etapification. In order to do this, it will be considered that once a woman in state C_i is diagnosed and moves to state D_i, she begins appropriate treatment, and therefore, her expected five year survival rate can be estimated. Let $r_i(n)$ be the expected five year survival rate for a woman diagnosed with a tumor in stage C_i at the age of n months. These rates are such that they satisfy $r_i(n) \geq r_{i+1}(n),\ i = 0, 1, 2$ for a given age n. As a simplification, it is assumed that the survival of a women depends only of her age and stage of cancer at the moment of diagnosis. For estimation purposes, survival rates from the SEER Explorer [41] will be used, which are calculated using a large database from women in USA. These rates and their 95 % confidence intervals can be found in appendix G. Finally, the estimated five years survival rate for the whole population $\mathcal{R}$ is estimated by the mean survival of all the simulated women with breast cancer diagnosis.

2.2.2. Data

This section shows all parameters used in the simulation and their source.

In order to simulate the policies above stated, the authors selected as parameters $N_0 = 30 \cdot 12$ and $N_f = 100 \cdot 12$ months.

The incidence rate ϕ_0 was extracted from the results of the first part of this study. The mean incidence rates for the 2002-2018 period were used. These rates are age specific and are calculated for bins of 5 years. The annual probability obtained was then passed to a monthly probability.

The decease rates depending on age were calculated using the national death registry and population estimates and projections provided by the Chilean National Institute of Statistics (INE). The number of deaths for each age and year in the period 2002 to 2018 were calculated

by selecting all women who died from causes different than breast cancer (identified by the C50 category) from the death registry. Dividing these number of deaths by the population for each age and year, a probability of death is obtained. Finally, the mean was taken over the years 2002 to 2018 to obtain a yearly probability of decease for each age. This yearly probability was then passed to a monthly probability.

Values for the parameters λ and $\phi_{0,1}$ were not found in literature as age dependent. Therefore they will be taken as constant.

The transition rates from the DCIS stage were obtained from [45]. The probability of transitioning from DCIS to Local stage was calculated considering the mean sojourn time before progressing to invasive cancer, which the study reports to be 2.4 months. Assuming an exponential transition, its rate would be 1/2.4, and then the probability of performing a transition in a month is 0.34. Tan et al [45] also calculates the probability to symptomatically detect DCIS before the progression to invasive breast cancer to be 0.03. Assuming the detection transition is also exponentially distributed, and considering a race of exponential distributions, then the detection transition has a rate of 0.013. This means that the probability of a women with DCIS to detect her tumor in a month is of 0.013.

This study [45] also provides us with mammography sensitivity values for DCIS and for invasive cancers depending on their size. The sensitivity for invasive tumors is presented as depending on the tumors diameter, separated the bins <10mm, [10,20), [20,50) and ≥50mm. Calculating the corresponding volumes, the sensitivity rates are found as depending on volume for the bins $< \frac{10^3}{6}\pi$, $[\frac{10^3}{6}\pi, \frac{20^3}{6}\pi)$, $[\frac{20^3}{6}\pi, \frac{50^3}{6}\pi)$ and $\geq \frac{50^3}{6}\pi$ mm^3.

The survival rates for invasive breast cancer in local, regional and distant stages where found in the SEER explorer application [41]. These rates where found for four age bins, <50, [50,65), [65,75) and ≥75 years old. Given that the survival rate for DCIS has to be such that $r_1 \leq r_0 \leq 1$ and that DCIS is usually considered as not life-threatening [46], the authors decided to set r_0 as constant $r_0 = 1$.

All parameters used in the simulation are shown in table 2.2.

Parameter	Source	Value (95 % CI)
ϕ_0	Table D.2	-
ρ	National death registry	-
$\phi_{0,1}$	[45]	0.34
λ	[45]	0.013 $(-)$
q_0	[45]	0.88 (± 0.05)
$q(V)$	[45]	0.90 (± 0.03) $if\ V < \frac{10^3}{6}\pi\ mm^3$ 0.91 (± 0.03) $if\ V \in [\frac{10^3}{6}\pi, \frac{20^3}{6}\pi)\ mm^3$ 0.92 (± 0.03) $if\ V \in [\frac{20^3}{6}\pi, \frac{50^3}{6}\pi)\ mm^3$ 0.93 (± 0.03) $if\ V \geq \frac{50^3}{6}\pi\ mm^3$
$r_1(age)$	[41]	0.972 (± 0.002) $if\ age < 50$ 0.985 (± 0.002) $if\ age \in [50, 65)$ 0.999 (± 0.017) $if\ age \in [65, 75)$ 1 $(-)$ $if\ age \geq 75$
$r_2(age)$	[41]	0.874 (± 0.005) $if\ age < 50$ 0.877 (± 0.005) $if\ age \in [50, 65)$ 0.870 (± 0.008) $if\ age \in [65, 75)$ 0.738 (± 0.015) $if\ age \geq 75$
$r_3(age)$	[41]	0.393 (± 0.019) $if\ age < 50$ 0.298 (± 0.013) $if\ age \in [50, 65)$ 0.271 (± 0.017) $if\ age \in [65, 75)$ 0.186 (± 0.019) $if\ age \geq 75$
r_0	Given the restriction $r_1 \leq r_0 \leq 1$	1 $(-)$

Table 2.2: Parameters and their sources.

Recall that parameters p_F and p_I are the p parameters of the random policies replicating FONASA and ISAPRE respectively. These values are taken such that the Bernoulli variable replicates the information shown in Table 2.1. Notice that the probability of a Bernoulli variable of parameter p of having at least one success in three experiments is $1-(1-p)^3$. Then, the p_F and p_I will take values depending on the woman's age as shown in equation 2.7.

$$p_F(age) = \begin{cases} 0 & if\ age < 35 \\ 0.132 & if\ age \in [35, 40) \\ 0.233 & if\ age \in [40, 45) \\ 0.298 & if\ age \in [45, 50) \\ 0.34 & if\ age \in [50, 55) \\ 0.332 & if\ age \in [55, 60) \\ 0.321 & if\ age \in [60, 65) \\ 0.158 & if\ age \geq 65 \end{cases}, \quad p_I(age) = \begin{cases} 0 & if\ age < 35 \\ 0.219 & if\ age \in [35, 40) \\ 0.334 & if\ age \in [40, 45) \\ 0.417 & if\ age \in [45, 50) \\ 0.419 & if\ age \in [50, 55) \\ 0.439 & if\ age \in [55, 60) \\ 0.373 & if\ age \in [60, 65) \\ 0.301 & if\ age \geq 65 \end{cases} \tag{2.7}$$

2.2.3. Results

Each policy was simulated the necessary amount of times in order to get 9604 diagnosed women. On average each policy was simulated 177,648 times. The results for each of these policies are shown in table 2.3 and in figure 2.3. This table shows the average amount of mammographies per women, the number of women diagnosed by screening and by clinical detection, the mean age at diagnosis, the distribution of diagnosed women by stage of diagnosis and the estimated five year survival rate $\mathcal{R}$ for each policy.

Policy	No screening	GES	US screening	Random FONASA	Random ISAPRE
N° of sim.	183,517	181,808	180,551	175,133	167,231
N° of tests/w	0.0	3.8	11.7	10.8	16.1
Screened	0	2011	5050	4948	6189
Clinical	9604	7593	4554	4656	3415
Age	66.8	66.1	65.0	66.1	65.7
DCIS	0.042	0.054	0.101	0.089	0.11
Local	0.438	0.529	0.66	0.656	0.702
Regional	0.457	0.368	0.21	0.225	0.166
Distant	0.062	0.049	0.029	0.03	0.022
$\mathcal{R}$	0.892	0.909	0.937	0.935	0.946

Table 2.3: Results obtained by simulating different policies.

All policies where simulated until 9,604 women where diagnosed. For each of the five simulated policies, the table shows the number of simulations, average number of test per women, tumors detected by screening, tumors detected by symptoms, proportion of tumors detected in DCIS, local, regional and distant stages, and the estimated five year survival rate $\mathcal{R}$ under such policy.

There is a clear tendency between the different policies, where the proportion of cases diagnosed in DCIS and local stages increase with more screening, while women diagnosed in regional and distant stages decrease, meaning that the more screening is performed, the better survival rate is achieved.

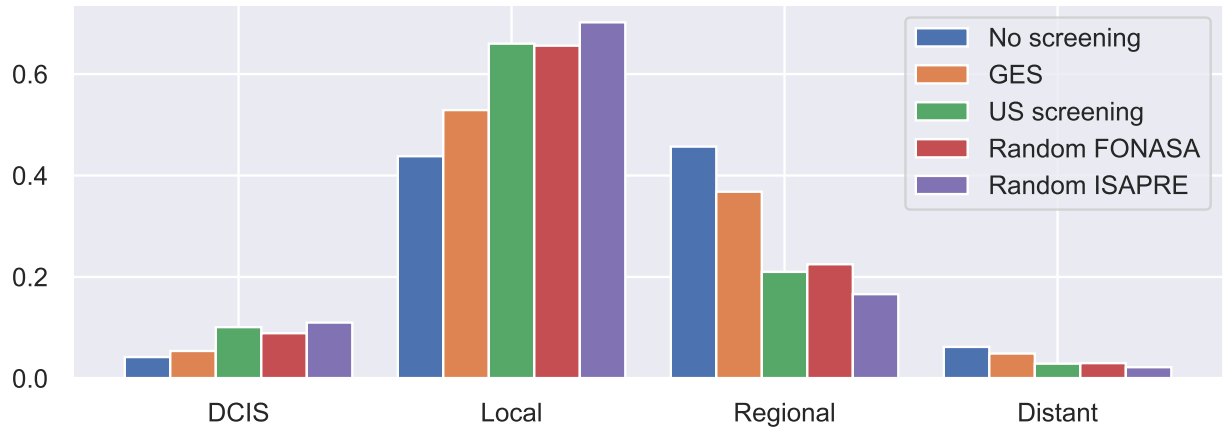

Figure 2.3: Distribution of women diagnosed in each stage for different policies.

Values are shown in table 2.3.

The no screening policy generated a detection staging composed by 4.2 % DCIS, 43.8 % local, 45.7 % regional and 6.2 % distant. The mean distribution for years 2000 to 2003 of 4.3 % DCIS, 58.2 % local, 26.6 % regional and 10.8 % distant reported by Prieto [39]. Then, the computational experiments' base case scenario does not replicate the Chilean situation from the years 2000 to 2003 where there was little to none screening.

The policy that generate the higher survival rate is the random ISAPRE policy. This policy also performs the most mammograms, with an average of 16.1 mammograms per woman, and is able to detect more tumors through screening, detecting 6189 through this exam and only 3415 by clinical detection.

The random FONASA policy scores similarly to the US screening policy, which is considered a great case scenario, with survival rates of 0.935 and 0.937, respectively. This policy that replicates the behaviour of FONASA patients performs an average of 10.8 mammograms per woman and detects 4948 tumors by screening. Again, this is similar to the US screening which performs an average of 11.7 mammograms per woman and detects 5050 tumors by screening.

The US screening policy generates the lower ages at diagnosis, with an average of 65 years old at diagnosis, which is 1.8 years less than the no screening policy.

The GES policy improved compared to no screening. Despite performing only 3.8 average screening test per women, the GES policy detected 2011 tumors, resulting in a reduction of 0.017 in the five year survival rate compared to no screening. Even though the GES policy performs far less mammograms that the random FONASA policy, they both achieve the same average age at diagnosis.

The estimated difference in five year survival rates between the policies replicating FO-NASA and ISAPRE is of 0.011.

Discussion

Incidence, mortality and survival rates were calculated using public anonimyzed databases. It is important to remark that the estimated incidence rates are based on hospital discharges and therefore they represent a lower bound of the real rates.

The Global Cancer Observatory (GLOBOCAN) estimated incidence and mortality ASRs of 37.4 and 10.2 for 2020, respectively [2]. The ASR mortality found, 9.8, is consistent with the measure of GLOBOCAN, while our estimation of ASR incidence, 41.2, is higher than that of GLOBOCAN. However, the ASRs found for incidence at national level show a constant trend, which is consistent with the world trend and differs from the decreasing trend found by the Ministry of Health [14].

The significant decreasing trend found for women older than 80 years old goes along with the milder increasing trends in the age intervals 30 to 49 and 70 to 79 to maintain the overall ASR constant trend. The authors believe that this change in the age distribution of incident cases is due to the increase in preventive measures, such as preventive mammograms, which help women obtain an earlier diagnosis.

It is important to take into account that as shown in [47], the population in Chile has grown older during the studied period. Therefore, although table 1.1 shows no decreasing trend for the age at diagnosis, this might be because the aging of the population is compensated by an earlier age of diagnosis.

In terms of mortality, the decreasing trend found for ASR is consistent with the decreasing trends found for several age intervals by Icaza, Nuñez and Bugueño [19]. On the other hand, unlike [19], this study shows an important, statistically significant, decreasing trend for the mortality of women older than 80 years, which may be related to the earlier diagnosis discussed above.

The survival rates found are similar to those reported by Del Castillo et al. [20] and by Allemani et al. [21]. Del Castillo et al. [20] carried out a clinical study of the public system in which they estimate the five year survival for the public system to be 0.751, which is only

0.026 below the confidence interval found in this study by the Kaplan Meier method for the public health care system. Allemani et al. [21] estimates survival rates for a more extensive set of countries. Our measure with the Kaplan Meier method of the five year survival rate for Chile of 0.805 lies inside the 95 % CI estimated by Allemani, which is [0.704,0.838].

The hazard ratios found show that there is an important age impact on survival, which grows for older ages. Thus, the hazard ratio for a 20 year difference between 40 and 60 year old women is of 1.16, while the same 20 year difference between 50 and 70 year old women produces a hazard ratio of 1.45. It is also relevant to highlight the hazard ratio between ISAPRE and FONASA affiliates of 1.72. This is one of the highest hazard ratios found, with far more relevance than the presence of GES, which has a hazard ratio of 1.14.

Most of the evaluations of the GES plan have focused on compliance of the explicit guarantees offered by the plan: access, opportunity, financial protection, and quality [9]. In terms of opportunity, in 2017 99.59 % of the services fulfilled these guarantee, being 92.76 % before the deadline. On the other hand, inspections by the General Contralory of the Republic (CGR) makes sure that a series of quality measures are fulfilled by the health care facilities.

However, ultimately, the effectiveness of the GES plan must be reflected on the improvement of health indicators, such as for example patients survival rates or quality-adjusted years of life. As discussed in section 1.3, this study shows that since the incorporation of the GES plan did have a modest improvement on the patients' survival rate, with a odd ratio of 1.14. Nevertheless, results show a marked difference in survival rates between insured patients in the public and private health systems. Thus, according to the Cox regression, at the age of 60 five year survival rates for FONASA and ISAPRE patients are 0.896 and 0.828, respectively, with a difference of 0.068. Meanwhile at the age of 90 the five year survival rates for FONASA and ISAPRE patients are 0.730 and 0.583, increasing the difference to 0.147.

Patiens from the public and private health care systems not only differ in the health insurer of their choice, but also conform two different sociodemographic groups in terms of economic income and education, and eventually comorbidities and age. For example, more than 85 % of the people from the five lower income deciles belong to FONASA, while only 25 % of people in the highest income decile belong to FONASA [12]. Likewise, income is also related to education, where the three lower income deciles have an average of less than 10 years of schooling, while people from the highest decile has an average of more than 15 years of schooling [48]. These differences might translate into disparities of opportunity to consult a doctor, adherence to treatments, treatment outcomes due to comorbidities, among others.

Hence, multiple reasons contribute to this inequality regarding health indicators between private and publicly insured patients. Therefore it is not clear that the differences in survival rates found between the public and the private health systems are due to the health systems themselves. Nevertheless, to measure them is a key step to study such differences and aim for improvements in the detection and treatment of breast cancer.

The computational experiments provide important insight over the situation of breast cancer in Chile, despite the fact that the no screening policy does not replicate the Chilean situation.

It is important to recall that the modeling of the natural progression of the tumors came from a study based on USA data, which may not represent the Chilean reality. This issue was addressed by modifying the parameter related with the natural detection of breast cancer, improving the fit to Chilean staging. Yet a complete fit could not be achieved for it also relied on other parameters related to the natural progression of breast tumors which were not modified. With the existence of a Chilean national registry of cancer, different and better models could be applied to improve results.

The survival rates obtained by the simulation model are considerably higher that those obtained in the first part of the study. This may be due to the disparity in staging above mentioned, or to other factors that influence the survival rates that are not considered here, such as quality and opportunity of breast cancer treatment and patient adherence.

It is also important to recall that it is unknown when people in FONASA and ISAPRE will get screened. This because they don't comply with a concrete guideline, but rather follow a random behaviour. Moreover, this random screening decisions cannot be fully replicated, for there is no available data with which such behaviour can be studied.

Although the distribution of diagnosed states does not completely explain the difference in survival between FONASA and ISAPRE affiliates, it significantly influences the outcome. Here we highlight that the screening method may provoke an increase in survival rates of 0.054 between no screening and the random policy replicating the ISAPRE behaviour.

We also highlight that women in FONASA experience an important improvement in their five year survival rates, this comparing the no screening policy with the GES and the random FONASA policies, having a five year survival rate raise of 0.017 and 0.043, respectively.

The difference in five year survival rates measured in the first part of this study between FONASA and ISAPRE affiliates was of 0.06 according to the Cox Regression. According to the simulation process, assuming that the random FONASA and random ISAPRE policies resemble reality of mammographies performed, only 0.011 of this gap is attributable to the differences in screening. Then, only a 18 % of the survival difference found in chapter 1 between the public and private health care systems is directly attributable to the screening policy.

There are several other differences between patients from FONASA and ISAPRE that may help explain the survival gap. One of these differences is the financial support. There is an additional coverage for catastrophic diseases (CAEC), which began in the year 2000 and

is now a days provided by 99.2 % of the ISAPREs [49]. This financial support program is defined to cover all medical expenses that go beyond the GES plan. This also makes impact in the quality of treatment. An example of this is the inclusion of a drug called trastuzumab, which has shown a great impact in the treatment of HER2+ breast cancer [50]. This drug was approved by the Public Health Insitute (ISP) in the year 2001 (registry B-1028/21), but it wasn't until 2016 with the implementation of the Ricarte Soto law [51] that the drug had financial support for the public sector. Meanwhile it was available in the private system to whoever could afford it.

Conclusions

This work studied the state of breast cancer in Chile, giving special emphasis in the differences by health care providers and how prevention through mammographies influences such differences. Incidence, mortality and survival rates were calculated using public anonimyzed databases.

The hospital discharges database provided by DEIS was used for the calculation of incidence and survival rates, obtaining values consistent with previous literature and international trends. Moreover, the use of this database, which to the best of our knowledge has not been used to study breast cancer, allowed to study incidence and survival rates by age, region of residence, year of diagnosis and health care provider. The use of this database presents an opportunity for the estimation of such key statistics without the costs of controlled trials.

The differences found in incidence and mortality among regions in section 1.3 cast doubt on estimations such as the ones made by the Minsal for national incidence which assumed constant fatality across regions [14]. This reinforces the utility of the use of the hospital discharges database.

Survival rates where analyzed by the Kaplan Meier method and by the Cox regression. The Kaplan Meier method estimates the breast cancer five year survival for chilean female population and its 95 % CI to be 0.805 and [0.802,0.808], respectively. The difference in breast cancer survival found between women affiliated to the public and private sectors is of key importance. The Cox regression, which besides the health care provider accounts for covariables such as age, region of residence and year of diagnosis, estimates a difference in the five year survival between the public and private health care systems of 0.06. Although it is well known that there are vast differences between the public and private health care systems, to the best of our knowledge this gap has not been previously quantified.

A stochastic model was developed based on previous work over the natural evolution of breast tumors. This allowed to perform computational experiments under different mammography screening policies. As expected, the computational experiments showed the high impact that screening mammographies have in the etapification and therefore in survival. Nevertheless, the difference in screening between the policies replicating the public and private

health care providers resulted in a difference of only 0.011 points of survival.

As discussed in section 2.1.1.1, the etapification with which the natural progression of the disease was originally calibrated may not represent the Chilean reality. Although this issue was addressed by adjusting the parameter associated with natural detection of symptoms, this may be a source of errors. For that matter, a sensibility analysis is yet to be performed to assure consistency of results.

Another liability may surge from the policies used to study the survival outcomes of public and private health care affiliates. As stated in section 2.2.1, the actual behaviour of these women in terms of screening adherence is unknown.

Possible extensions of this work are the use of the discharges database to analyze key statistics of other diseases. A similar methodology based on stochastic modeling and simulation could also be applied to other diseases to better understand differences between the public and private health care systems.

Further work may focus on a better understanding of the breast cancer survival gap between FONASA and ISAPREs systems. For this matter, it is yet to be studied the influence of treatment opportunity, quality of treatment, financial support, patients adherence, among others.

Bibliography

[1] World Health Organization, "Cancer," 2021. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 07 October 2021.

[2] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: Cancer today," 2020. https://gco.iarc.fr/today. Accessed 07 October 2021.

[3] Ministerio de Salud, "Plan nacional de cáncer 2018 - 2028," 2019.

[4] American Cancer Society, "Breast cancer facts & figures 2009-2010," *Atlanta: American Cancer Society, Inc.*, 2009.

[5] International Agency for Research on Cancer, "Breast cancer screening," 2016.

[6] K. Gallardo, L. Varas, and M. Gallardo, "Inequality of opportunity in health: evidence from chile," *Revista de Saúde Pública*, vol. 51, p. 110, Dec. 2017.

[7] C. Castillo-Laborde, X. Aguilera, M. Hirmas, I. Matute, I. Delgado-Becerra, M. Najera, A. Olea-Normandin, and C. González-Wiedmaier, "Health insurance scheme performance and effects on health and health inequalities in chile," *MEDICC review*, vol. 19, pp. 57–64, 07 2017.

[8] R. Lenz, *Proceso político de la reforma auge de salud en Chile: algunas lecciones para América Latina: una mirada desde la economía política.* Cieplan Santiago, 2007.

[9] Secretaria Técnica GES, Ministerio de Salud, "Monitoreo garantías explícitas en salud," 2018.

[10] Ministerio de Salud, *Cuenta Pública 2016 del Instituto Nacional del Cáncer mostró excelentes resultados en gestión y atención a pacientes oncológicos.* Santiago de Chile, 2017. https://www.minsal.cl/cuenta-publica-2016-mostro-excelentes-resultados-en-gestion-y-atencion-a-pacientes-oncologicos/. Accessed 14 March 2022.

[11] WHO Regional Office for Europe, "Screening programmes: a short guide. increase effectiveness, maximize benefits and minimize harm.," 2020.

[12] Observatorio Social, Ministerio de Desarrollo Social y Familia, *Encuesta de caracterización socioeconómica nacional. Resultados Salud Casen 2017.* Santiago de Chile, 2018. Accessed 14 October 2021.

[13] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. Parkin, M. Piñeros, A. Znaor,

and F. Bray, "Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, 2019.

[14] Ministerio de Salud, "Segundo informe nacional de vigilancia de cancer en chile. estimación de incidencia [internet]," 2019.

[15] Departamento de Epidemiología, Ministerio de Salud, "Informe de vigilancia de cáncer. análisis de mortalidad. década 2009-2018.," 2020.

[16] Observatorio Social, Ministerio de Desarrollo Social y Familia, *Encuesta de caracterización socioeconómica nacional. Resultados Ingresos Casen 2017.* Santiago de Chile, 2018. Accessed 14 October 2021.

[17] Observatorio Social, Ministerio de Desarrollo Social y Familia, *Encuesta de caracterización socioeconómica nacional. Resultados Pueblos Indigenas Casen 2017.* Santiago de Chile, 2018. Accessed 14 October 2021.

[18] Observatorio Social, Ministerio de Desarrollo Social y Familia, *Encuesta de caracterización socioeconómica nacional. Resultados Territorios Rurales Casen 2017.* Santiago de Chile, 2018. Accessed 14 October 2021.

[19] G. Icaza, L. Núñez, and H. Bugueño, "Descripción epidemiológica de la mortalidad por cáncer de mama en mujeres en Chile," *Revista médica de Chile*, vol. 145, pp. 106 – 114, Jan. 2017.

[20] C. d. S. Castillo, M. E. C. Cabrera, L. Derio P., F. Gaete V., and G. Cavada CH., "Resultados del tratamiento del cáncer de mama, Programa Nacional de Cáncer del Adulto," *Revista Medica de Chile*, vol. 145, pp. 1507 – 1513, Dec. 2017.

[21] C. Allemani, H. K. Weir, H. Carreira, R. Harewood, D. Spika, X.-S. Wang, F. Bannon, J. V. Ahn, C. J. Johnson, A. Bonaventure, R. Marcos-Gragera, C. Stiller, G. Azevedo e Silva, W.-Q. Chen, O. J. Ogunbiyi, B. Rachet, M. J. Soeberg, H. You, T. Matsuda, M. Bielska-Lasota, H. Storm, T. C. Tucker, and M. P. Coleman, "Global surveillance of cancer survival 1995–2009: analysis of individual data for 25.676.887 patients from 279 population-based registries in 67 countries (concord-2)," *The Lancet*, vol. 385, no. 9972, pp. 977–1010, 2015.

[22] New York Times, "Chile woke up: Dictatorship's legacy of inequality triggers mass protests," 2019. https://www.nytimes.com/2019/11/03/world/americas/chile-protests.html. Accessed 15 October 2021.

[23] Subdepartamento de Demografía y Estadísticas Vitales, Instituto Nacional de Estadistica, *Estimaciones y proyecciones de la población de Chile 1992-2050.* Santiago de Chile, 2017. https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion. Accessed 14 October 2021.

[24] Subdepartamento de Demografía y Estadísticas Vitales, Instituto Nacional de Estadistica, *Estimaciones y proyecciones 2002-2035, comunas.* Santiago de Chile, 2017. Accessed 14 October 2021.

[25] F. Bray, M. Colombet, L. Mery, M. Piñeros, A. Znaor, R. Zanetti, and J. Ferlay, eds., *Cancer Incidence in Five Continents, Vol. XI.* No. 166 in IARC Scientific Publication, Lyon, France: International Agency for Research on Cancer, 2021.

[26] Fonasa, *Boletín Estadístico 2017-2018, Boletín Estadístico 2015-2016, Boletín Estadístico 2013-2014, Boletín Estadístico 2011-2012, Boletín Estadístico 2009-2010*. Santiago de Chile. https://www.fonasa.cl/sites/fonasa/documentos. Accessed 15 October 2021.

[27] Departamento de Estudios y Desarrollo, Superintendecia de Salud, *Cartera de beneficiarios de isapres Año 2018 - Cartera de beneficiarios de isapres Año 2002*. Santiago de Chile. https://www.supersalud.gob.cl/documentacion/666/w3-propertyvalue-3742.html. Accessed 15 October 2021.

[28] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, no. 282, p. 457–481, 1958.

[29] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society*, vol. 34, no. 2, p. 187–220, 1972.

[30] H. Akaike, "Information theory as an extension of the maximum likelihood principle in petrov bn & csaki f.(eds.) second international symposium on information theory (pp. 267–281)," 1973.

[31] S. K. Plevritis, P. Salzman, B. M. Sigal, and P. W. Glynn, "A natural history model of stage progression applied to breast cancer," *Statistics in Medicine*, vol. 26, no. 3, pp. 581–595, 2007.

[32] L. Abrahamsson and K. Humphreys, "A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density," *Statistical Methods in Medical Research*, vol. 25, no. 4, pp. 1620–1637, 2016. PMID: 23839121.

[33] S. K. Plevritis, B. M. Sigal, P. Salzman, J. Rosenberg, and P. Glynn, "Chapter 12: A Stochastic Simulation Model of U.S. Breast Cancer Mortality Trends From 1975 to 2000," *JNCI Monographs*, vol. 2006, pp. 86–95, 10 2006.

[34] L. Abrahamsson, G. Isheden, K. Czene, and K. Humphreys, "Continuous tumour growth models, lead time estimation and length bias in breast cancer screening studies," *Statistical Methods in Medical Research*, vol. 29, no. 2, pp. 374–395, 2020. PMID: 30854935.

[35] G. Isheden, L. Abrahamsson, T. Andersson, K. Czene, and K. Humphreys, "Joint models of tumour size and lymph node spread for incident breast cancer cases in the presence of screening," *Statistical Methods in Medical Research*, vol. 28, no. 12, pp. 3822–3842, 2019. PMID: 30606087.

[36] Surveillance, Epidemiology and End Results (SEER) Program, "Public use data (1973–2000)," 2003.

[37] E. G. MacInnes, S. W. Duffy, J. A. Simpson, M. G. Wallis, A. E. Turnbull, L. S. Wilkinson, K. Satchithananda, R. Rahim, D. Dodwell, B. V. Hogan, O. Blyuss, and N. Sharma, "Radiological audit of interval breast cancers: Estimation of tumour growth rates," *The Breast*, vol. 51, pp. 114–119, 2020.

[38] K. S. Trivedi and A. Bobbio, *Non-Homogeneous Continuous-Time Markov Chains*, p. 489–508. Cambridge University Press, 2017.

[39] M. Marta Prieto, "Epidemiología del cáncer de mama en chile," *Revista Médica Clínica Las Condes*, vol. 22, no. 4, pp. 428–435, 2011.

[40] L. M. Hines, B. Risendal, M. L. Slattery, K. B. Baumgartner, A. R. Giuliano, C. Sweeney,

D. E. Rollison, and T. Byers, "Comparative analysis of breast cancer risk factors among hispanic and non-hispanic white women," *Cancer*, vol. 116, no. 13, pp. 3215–3223, 2010.

[41] Surveillance Research Program, National Cancer Institute, "SEER*Explorer: An interactive website for SEER cancer statistics [Internet]." https://seer.cancer.gov/explorer/. Accessed 16 March 2022.

[42] Ministerio de Salud, "Guia clínica auge: Examen medicina preventiva," 2013.

[43] A. L. Siu, "Screening for breast cancer: U.s. preventive services task force recommendation statement," *Annals of Internal Medicine*, vol. 164, no. 4, pp. 279–296, 2016. PMID: 26757170.

[44] W. G. Cochran, *Sampling Techniques*. Jhon Wiley & Sons, 3rd ed., 1977.

[45] K. H. X. Tan, L. Simonella, H. L. Wee, A. Roellin, Y.-W. Lim, W.-Y. Lim, K. S. Chia, M. Hartman, and A. R. Cook, "Quantifying the natural history of breast cancer," *British Journal of Cancer*, vol. 109, p. 2035–2043, 2013.

[46] J. Stallard, "What is ductal carcinoma in situ (dcis), and how do i decide on the right treatment? [internet]," *Memorial Sloan Kettering Cancer Center*, 2021.

[47] C. Albala, "El envejecimiento de la población chilena y los desafíos para la salud y el bienestar de las personas mayores," *Revista Médica Clínica Las Condes*, vol. 31, no. 1, pp. 7–12, 2020.

[48] Observatorio Social, Ministerio de Desarrollo Social y Familia, *Encuesta de caracterización socioeconómica nacional. Resultados Educación Casen 2017.* Santiago de Chile, 2018. Accessed 13 April 2022.

[49] Superintendencia de Salud, "Más de 15 mil afiliados de isapres han utilizado la Cobertura Adicional para Enfermedades Catastróficas (CAEC)," 2009. http://www.supersalud.gob.cl/prensa/672/w3-article-5485.html. Accesed 8 April 2022.

[50] M. Jahanzeb, "Adjuvant trastuzumab therapy for her2-positive breast cancer," *Clinical Breast Cancer*, vol. 8, no. 4, pp. 324–333, 2008.

[51] T13, "Las 11 enfermedades que cubre la Ley Ricarte Soto que comienza a regir," December 2015. https://www.t13.cl/noticia/nacional/Hoy-comienza-a-regir-la-Ley-Ricarte-Soto-Revisa-cuales-son-las-patologias-que-cubre. Accessed 15 Abril 2022.

Annexes

Annexed A

Other breast cancer diagnosis

There is a considerable number of breast cancer death registries that do not have a any discharge registry associated with breast cancer (8,292 deaths). However, there are other discharge registries, different from breast cancer, which are not typically included in a breast cancer incidence analysis. These are health problems that can arise due to the progression of the disease or its treatment (surgery, chemotherapy, radiotherapy), and therefore, are included under certain specific conditions. Thus, these diagnostics were taken into consideration only when they belonged to a patient who died because of breast cancer. These diagnostics were grouped in three classes:

1. Discharge registries directly attributable to breast cancer and its treatment. This item includes examination, treatment by chemotherapy and radiotherapy, and different breast related diseases and issues, which may be due to breast cancer misdiagnosis. These diagnostics are: D486, D24X, Z123, Z803, Z853, Z031, Z080, Z081, Z082, Z087, Z088, Z089, Z129, Z400, Z510, Z511, Z512, Z515, Z809, Z859, Z860.

2. Diagnostics of other cancers and malignancies associated with breast cancer. This item includes secondary tumors and tumors of unspecified places. These diagnostics are: C798, C782, C795, C793, C787, C786, C412, C800, D382, C792, C709, C383, C799, D059, C770, D420, C796, C414, C500, C413, C771, C728, C967, C779, D383, C399, C773, C781, C783, C700.

3. Other diagnostics, which might be attributable to symptoms of breast cancer and its treatment. They are considered only if such discharge is close enough to the death registry. Each diagnostic has a different period of time to be associated with the death and are shown in Table A.1. Absence of period means the diagnostic is always included independent of its gap from the death registry.

Diagnostic Code	Relation period
D649	2
G039	1
G540	2
G939	1
G952	2
I495	-
I891	1
I972	-
J80X	2
J90X	2
J948	2
J960	2
J969	1
J984	2
J989	2
M532	4
M544	4
M546	4
M549	4
M808	4
M844	4
N63X	-
N645	-
N648	-
N649	-
N850	-
R060	1
R17X	1
R18X	1
R51X	1
R53X	1
S220	4
S320	2
S323	4
S325	4
S327	4
S328	4
S423	4
S720	4
S721	4
S722	4
S723	4
S724	4
S728	4
S729	4
T08X	4
T12X	4
T142	4
T932	4

Table A.1: CDI-10 codes and relation periods (years) of the health problems to be considered as the breast cancer debut.

In addition, discharges for breast tumors of unknown characteristics (D24X) were also included whenever followed by a malignant or in situ tumor because it corresponds to the real cancer debut.

Annexed B

Failure data processing

It will be assumed that some characteristics such as gender and date of birth should be invariant for each patient. There were several patients who had different registries with more than one gender, date of birth or year of birth (10,157). It is important to note that this can only be checked for registries with known ids and that not all registers had the patient's date of birth. For the patients that do not have any registry with a full date of birth, invariance was checked for their year of birth. It is also assumed that the death database had no errors in these fields, and therefore gender and date of birth were corrected for the people who had a death register by any cause. For the people who had no death register, gender, date of birth and year of birth were corrected by taking the mode of such characteristics over all registries available. In case that the complete date's mode was not conclusive, mode was taken by day, month and year. All registers who had no death associated in the decease database and that its patient's gender, date of birth or year of birth were not conclusive in terms of mode were deleted (2,552).

Besides these invariant characteristics, there were other characteristics that presented little variance through the discharges registries. Less than 4 % of the people with known ids changed their region of residence and less than 6 % their commune of residence. Also, only 3.7 % ever changed their health insurance and 8.5 % changed their FONASA beneficiary classification. Therefore, it was considered that a register with no id corresponds to the id of another register if they match in the characteristics gender, date or year of birth, commune of residence, health insurance and FONASA beneficiary classification, being the only match. It was also considered that a register with no id corresponded to an id reported in a later date on the death registry if it had no other discharge associated and they match in the characteristics gender, date or year of birth and commune of residence, being this the only match. Finally, it was considered that a register with no id belonged to a person not considered in the rest of the breast cancer discharges database if the register had no match considering gender, date or year of birth and region of residence. These last registries were given a new id different from those already in the database. Thus, it was possible to recover 4,487 ids, while the rest of registries with missing ids were deleted (12,793).

Annexed C

Implementation of Cox regression

The variables were processed previous to the selection. Squared age was added as a new variable to add more flexibility to the hazard's modelling as a function of age. Both age and squared age were normalized to a maximum age of 100 years. Categorical variables (health insurance, region of residence and FONASA benefit ranges) were transformed into dummy variables. A GES variable was added indicating if the diagnosis was from the year 2005 on, and therefore the GES plan was already implemented. The above transformations of variables resulted in a total of 27 covariables to be studied, this is, 3 health insurance dummy variables, 4 benefit range dummy variables, 16 region dummy variables, year of diagnose, normalized age, squared normalized age and a dummy variable indicating the existence of GES. The selection of relevant variables among these 27 covariables was done following a greedy procedure based on the Akaike's information criterion and on the p-value for each variable's significance, this is, variables were sequentially selected by taking the ones that mostly improved the Akaike's information criterion, while maintaining all variables' p-value as statistically significant.

More formally, given a set of n variables from where to choose which ones generate the best model, based on each variable's p-value and the models akaike information criterion the selection was performed as follows.

```
Variable_selection:
    Set the current best variables as an empty list.
    Set the current best akaike value as infinity.
    Set the n variables as a pending list.
    While there are still variables in the pending list, do:
        For each variable from the pending list, generate a cox model with it and the selected
    ↪ variables.
        For each such model, check if all their variables have significant p-values (≤0.05). If
    ↪ none of the models comply with it, break.
        For each model that meets the above:
            Calculate the Akaike information criterion.
```

```
10          If the calculated Akaike value is lower than the current best, set such value as the
    ↪ current best akaike value and add such variable to the current best variables,
    ↪ removing it from the pending list.
11        If none of the above models achieves an akaike value better than the current best,
    ↪ break.
12    Return the model generated with the current best variables.
```

Code C.1: Pseudo-code for greedy variable selection for Cox model.

For the 27 variables before mentioned, the previous process selected the following 11 variables: normalized age, square normalized age, private health insurance, public health insurance, discharge year, A FONASA beneficiary classification, B FONASA beneficiary classification, residence in the Metropolitan region, residence in the Arica and Parinacota region, residence in the Libertador O'higgins region and existence of GES.

Annexed D

Incidence and Mortality Results

D.1. Breast cancer incidence and mortality by geographical region

The mean incidence and mortality for each of the 16 Chilean regions over the studied period (2002-2018) are found in table D.1. Both are presented as age adjusted and crude rates. They are sorted from north to south.

		Incidence		Mortality	
Region	Region Name	Crude	Age adjusted	Crude	Age adjusted
XV	Arica y Parinacota	67.7	56.0	13.9	10.4
I	Tarapacá	34.3	32.3	11.9	10.7
II	Antofagasta	51.1	47.2	11.6	10.0
III	Atacama	36.9	32.4	11.4	9.5
IV	Coquimbo	41.5	33.4	12.9	9.3
V	Valparaíso	66.5	48.3	18.6	11.7
RM	Metropolitana de Santiago	62.8	49.5	16.0	11.2
VI	Libertador General Bernardo O'Higgins	39.6	31.3	14.2	10.2
VII	Maule	44.6	35.0	13.1	9.4
XVI	Ñuble	49.5	37.2	15.8	10.6
VIII	Biobío	48.9	38.6	14.5	10.4
IX	La Araucanía	43.7	35.0	13.2	9.4
XIV	Los Ríos	54.9	42.0	12.1	8.4
X	Los Lagos	36.0	29.8	11.3	8.4
XI	Aysén del General Carlos Ibáñez del Campo	37.0	33.6	12.4	10.5
XII	Magallanes y de la Antártica Chilena	57.1	43.5	19.7	13.6

Table D.1: Average age adjusted and crude incidence and mortality over the period 2002-2018 by region (cases/100,000 women).

D.2. Breast cancer incidence by age group

Table D.2 presents age specific incidence rates for each age interval for each year from 2002 to 2018.

Age	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Mean (Std)
0-19	0.7	0.3	0.5	1.5	1.2	1.0	0.6	0.8	0.9	0.7	0.7	0.7	0.5	0.9	0.5	0.7	0.6	0.8 (0.3)
20-24	3.2	3.2	2.8	4.5	5.1	3.2	3.2	4.3	2.3	3.2	3.1	3.6	1.9	2.1	3.5	4.1	3.7	3.3 (0.8)
25-29	6.4	6.5	4.4	8.3	8.9	9.8	8.4	5.3	7.5	6.1	6.2	6.8	5.8	6.3	8.5	6.7	9.4	7.1 (1.5)
30-34	14.0	16.0	11.5	18.2	17.6	14.6	12.3	13.3	14.1	13.8	18.8	15.4	15.8	16.5	20.2	17.2	18.8	15.8 (2.4)
35-39	34.2	35.8	23.9	35.3	36.6	32.8	30.4	32.6	33.3	36.0	37.9	36.6	34.8	37.5	35.9	35.9	36.7	34.5 (3.3)
40-44	67.1	63.6	55.7	74.4	64.5	66.1	62.2	61.6	62.0	69.7	74.2	71.5	67.8	68.1	78.9	71.2	64.2	67.2 (5.6)
45-49	91.4	106.5	87.3	111.4	94.0	105.3	92.8	100.0	103.7	112.4	117.1	106.3	101.3	111.9	117.9	97.6	104.0	103.6 (8.7)
50-54	110.2	123.8	100.7	113.9	99.7	111.2	105.6	114.0	114.0	119.0	123.0	121.7	108.1	129.1	125.9	102.8	115.9	114.0 (8.6)
55-59	140.0	164.6	126.7	149.6	119.4	131.0	111.8	122.1	122.2	124.9	122.2	140.1	127.3	138.5	125.1	124.5	121.9	130.1 (12.5)
60-64	142.9	173.2	124.0	173.8	143.9	161.0	152.5	144.1	158.5	165.4	162.1	159.3	147.0	138.0	138.9	136.9	133.1	150.3 (14.0)
65-69	151.2	172.5	135.9	162.5	143.1	169.3	178.0	157.2	177.0	184.7	180.4	196.0	177.9	171.9	163.7	164.2	144.5	166.5 (15.6)
70-74	165.9	176.9	160.1	176.0	164.8	150.2	152.8	157.9	182.6	185.9	179.8	181.7	198.5	188.6	173.4	178.6	178.7	173.7 (12.8)
75-79	172.3	189.6	165.9	207.0	153.5	170.7	154.0	163.8	168.9	169.1	182.0	171.7	166.3	180.7	181.3	160.3	163.3	171.8 (12.9)
80-84	207.8	208.0	195.4	197.8	178.5	200.5	167.0	149.5	181.1	173.1	163.1	178.3	156.0	147.1	144.4	128.3	124.5	170.6 (25.6)
85+	230.0	224.4	234.2	225.8	194.2	228.7	213.5	179.3	189.5	204.1	161.6	156.4	149.5	118.9	111.7	100.1	75.1	176.3 (49.3)

Table D.2: Crude incidence rate by year and age group (cases/100,000 women).

D.3. Breast cancer mortality by age group

Table D.3 presents age specific mortality rates for each age interval for each year from 2002 to 2018.

Age	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Mean (Std)
0-19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0 (0.0)
20-24	0.0	0.3	0.2	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1 (0.1)
25-29	0.6	0.6	0.5	0.3	1.1	0.5	1.2	0.9	0.9	1.1	0.3	1.2	0.9	0.4	0.5	0.3	1.0	0.7 (0.3)
30-34	2.0	2.3	1.3	2.9	1.6	1.1	3.0	1.4	2.0	2.0	1.9	2.0	2.1	2.6	2.1	2.9	3.4	2.2 (0.6)
35-39	4.9	5.0	6.7	5.9	6.5	4.6	5.3	5.5	6.0	4.0	4.9	5.0	7.0	6.2	4.8	4.1	5.2	5.4 (0.8)
40-44	9.4	11.2	10.1	12.1	10.3	11.1	9.3	9.7	11.4	9.6	12.3	11.1	10.2	10.6	10.7	8.7	12.8	10.6 (1.1)
45-49	18.1	18.3	16.7	16.6	19.0	18.8	16.6	19.6	18.9	17.2	17.2	19.7	17.7	17.5	17.3	14.8	14.1	17.5 (1.5)
50-54	26.5	30.1	26.1	22.2	21.3	23.3	24.7	25.7	25.4	26.7	22.0	21.2	24.5	22.5	24.7	25.3	18.8	24.2 (2.6)
55-59	33.0	31.6	33.1	32.4	29.6	33.4	29.0	32.2	31.0	29.0	29.6	28.9	27.8	32.8	30.0	29.7	29.5	30.8 (1.7)
60-64	36.2	38.5	43.1	40.5	41.2	34.4	40.8	42.4	39.1	38.0	36.4	38.1	29.1	33.4	31.7	38.0	30.8	37.2 (4.0)
65-69	49.2	42.6	43.4	48.4	45.1	45.4	48.0	53.2	47.8	48.0	44.9	43.3	46.4	47.6	47.2	38.7	39.7	45.8 (3.5)
70-74	55.6	44.2	51.8	61.8	52.7	52.2	55.1	52.5	47.5	60.3	56.9	63.4	56.9	54.4	62.1	60.1	49.2	55.1 (5.2)
75-79	76.9	69.7	59.8	71.7	71.1	60.8	72.6	70.7	69.0	55.6	68.9	61.3	66.3	74.3	69.1	71.6	83.2	69.0 (6.5)
80-84	93.6	97.7	122.0	104.9	92.1	83.2	81.8	91.7	85.8	102.4	88.8	95.6	84.7	99.6	77.5	73.6	90.8	92.1 (11.2)
85+	198.3	198.8	164.2	181.2	169.9	185.3	175.1	192.0	153.2	170.4	173.4	139.1	164.8	150.5	139.6	142.1	144.3	167.2 (19.4)

Table D.3: Crude mortality rate by year and age group (cases/100,000 women).

Annexed E

Survival Records

The Kaplan-Meier curves for women in FONASA are made with the events observed and censored shown in table E.1. On the other hand, table E.2 shows the events observed and censored used to make the Kaplan-Meier curves for women in ISAPRE. The event table for both FONASA and ISAPRE women can be obtained by summing both tables.

Event tables for other Kaplan-Meier curves shown in results section are not included due to their extension.

Time (months)	Removed	Observed	Censored	At risk
0	1,191	1,073	118	54,642
1	907	527	380	53,451
2	760	375	385	52,544
3	659	302	357	51,784
4	578	277	301	51,125
5	589	224	365	50,547
6	604	262	342	49,958
7	622	241	381	49,354
8	590	227	363	48,732
9	568	207	361	48,142
10	616	219	397	47,574
11	532	218	314	46,958
12	550	206	344	46,426
13	515	189	326	45,876
14	510	189	321	45,361
15	500	194	306	44,851
16	487	184	303	44,351
17	523	168	355	43,864
18	495	180	315	43,341
19	538	186	352	42,846
20	553	197	356	42,308
21	449	176	273	41,755
22	542	177	365	41,306
23	501	191	310	40,764
24	494	166	328	40,263
25	486	127	359	39,769
26	474	162	312	39,283
27	465	145	320	38,809
28	430	135	295	38,344
29	471	147	324	37,914
30	465	140	325	37,443
31	400	138	262	36,978
32	470	140	330	36,578
33	446	108	338	36,108
34	436	125	311	35,662
35	407	120	287	35,226
36	434	107	327	34,819
37	383	96	287	34,385
38	446	102	344	34,002
39	420	95	325	33,556
40	401	119	282	33,136
41	442	99	343	32,735
42	420	92	328	32,293
43	416	90	326	31,873
44	359	91	268	31,457
45	391	99	292	31,098
46	406	86	320	30,707
47	365	92	273	30,301
48	388	82	306	29,936
49	321	70	251	29,548
50	365	96	269	29,227
51	375	84	291	28,862
52	342	80	262	28,487
53	363	66	297	28,145
54	315	72	243	27,782
55	323	72	251	27,467
56	355	65	290	27,144
57	321	59	262	26,789
58	341	78	263	26,468
59	289	70	219	26,127
60	25,838	81	25,757	25,838

Table E.1: Event table for the survival curve for patients in the public health system.

Time (months)	Removed	Observed	Censored	At risk
0	127	110	17	14,693
1	167	49	118	14,566
2	152	28	124	14,399
3	163	45	118	14,247
4	120	39	81	14,084
5	155	32	123	13,964
6	119	35	84	13,809
7	109	22	87	13,690
8	122	30	92	13,581
9	117	23	94	13,459
10	106	22	84	13,342
11	94	20	74	13,236
12	118	17	101	13,142
13	127	20	107	13,024
14	134	34	100	12,897
15	128	19	109	12,763
16	146	35	111	12,635
17	122	24	98	12,489
18	121	20	101	12,367
19	131	27	104	12,246
20	105	22	83	12,115
21	85	17	68	12,010
22	108	20	88	11,925
23	75	20	55	11,817
24	120	29	91	11,742
25	130	18	112	11,622
26	130	22	108	11,492
27	126	27	99	11,362
28	121	16	105	11,236
29	146	26	120	11,115
30	120	10	110	10,969
31	120	21	99	10,849
32	128	26	102	10,729
33	118	15	103	10,601
34	136	17	119	10,483
35	82	15	67	10,347
36	135	15	120	10,265
37	115	29	86	10,130
38	124	23	101	10,015
39	91	17	74	9,891
40	94	18	76	9,800
41	93	15	78	9,706
42	102	15	87	9,613
43	99	15	84	9,511
44	108	24	84	9,412
45	92	14	78	9,304
46	122	8	114	9,212
47	79	13	66	9,090
48	113	10	103	9,011
49	103	10	93	8,898
50	108	12	96	8,795
51	123	14	109	8,687
52	75	11	64	8,564
53	84	10	74	8,489
54	85	12	73	8,405
55	101	14	87	8,320
56	84	12	72	8,219
57	94	7	87	8,135
58	90	15	75	8,041
59	77	20	57	7,951
60	7,874	9	7,865	7,874

Table E.2: Event table for the survival curve for patients in the private health system.

Annexed F

Disease progression

This appendix presents further details on the tumor growth model presented in section 2.1.1.

It was shown that the stochastic process representing the tumors stages had a probability transition matrix P shown in equation 2.6.

For the diagonalization, the eigenvalues found for the W matrix are $\{-\gamma-\eta, -\gamma-\omega, -\gamma, 0\}$. The resulting transition matrix P is as follows.

$$M_1 = \begin{bmatrix} e^{(-\gamma-\eta)\bar{V}(t,\Delta t)} & \frac{\eta(e^{(-\gamma-\omega)\bar{V}(t,\Delta t)} - e^{(-\gamma-\eta)\bar{V}(t,\Delta t)})}{\eta-\omega} & \frac{\omega e^{(-\gamma-\eta)\bar{V}(t,\Delta t)} - \eta e^{(-\gamma-\omega)\bar{V}(t,\Delta t)}}{\eta-\omega} - e^{-\gamma\bar{V}(t,\Delta t)} \\ 0 & e^{(-\gamma-\omega)\bar{V}(t,\Delta t)} & e^{-\gamma\bar{V}(t,\Delta t)} - e^{(-\gamma-\omega)\bar{V}(t,\Delta t)} \\ 0 & 0 & e^{-\gamma\bar{V}(t,\Delta t)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} \frac{\gamma(1-e^{(-\gamma-\eta)\bar{V}(t,\Delta t)})}{\gamma+\eta} & \frac{\gamma\eta}{\eta-\omega}\left(\frac{e^{(-\gamma-\eta)\bar{V}(t,\Delta t)}}{\gamma+\eta} - \frac{e^{(-\gamma-\omega)\bar{V}(t,\Delta t)}}{\gamma+\omega}\right) + \frac{\gamma\eta}{(\gamma+\eta)(\gamma+\omega)} & \frac{\gamma}{\eta-\omega}\left(\frac{\eta e^{(-\gamma-\omega)\bar{V}(t,\Delta t)}}{\gamma+\omega} - \frac{\omega e^{(-\gamma-\eta)\bar{V}(t,\Delta t)}}{\gamma+\eta}\right) - e^{-\gamma\bar{V}(t,\Delta t)} + \frac{\omega\eta}{(\gamma+\eta)(\gamma+\omega)} \\ 0 & \frac{\gamma(1-e^{(-\gamma-\omega)\bar{V}(t,\Delta t)})}{\gamma+\omega} & \frac{\gamma e^{(-\gamma-\omega)\bar{V}(t,\Delta t)+\omega}}{\gamma+\omega} - e^{-\gamma\bar{V}(t,\Delta t)} \\ 0 & 0 & 1 - e^{-\gamma\bar{V}(t,\Delta t)} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P(t,\Delta t) = [M_1, M_2]$$

The parameters found by Plevritis et al [31] under the assumtion that $\mathbb{E}(R) = 1$ are shown in table F.1.

Estimate	Estimate	95 % CI
$\ln(\hat{\gamma})$	-9.602	$[-9.624, -9.580]$
$\ln(\hat{\eta})$	-9.636	$[-9.661, -9.610]$
$\ln(\hat{\omega})$	-11.765	$[-11.816, -11.713]$
$\ln\left(\hat{\beta}\right)$	-0.165	$[-0.187, -0.143]$
$\hat{\alpha}$	$\hat{\beta}$	-

Table F.1: Parameters found by Plevritis et al. [31] under the assumption of $\mathbb{E}(R) = 1$.

Then, considering $\mathbb{E}(R) = (167/365)/ln(2)$, parameters in table F.1 must be re escalated as shown in equation F.1. The parameters obtained are shown in table F.2.

$$\begin{aligned}
\hat{\gamma} &= \exp(-9.602)/\mathbb{E}(R) \\
\hat{\eta} &= \exp(-9.636)/\mathbb{E}(R) \\
\hat{\omega} &= \exp(-11.765)/\mathbb{E}(R) \\
\hat{\beta} &= \exp(-0.165)/\mathbb{E}(R) \\
\hat{\alpha} &= \hat{\beta}/\mathbb{E}(R) = \exp(-0.165)
\end{aligned} \tag{F.1}$$

Estimate	95 % CI
$\mathbb{E}(R)$	$[0.66 \pm 0.075]$
$\ln(\gamma)$	$[-9.186 \pm 0.116]$
$\ln(\eta)$	$[-9.22 \pm 0.117]$
$\ln(\omega)$	$[-11.349 \pm 0.125]$
$\ln(\beta)$	$[0.251 \pm 0.116]$
$\ln(\alpha)$	$[-0.165 \pm 0.022]$

Table F.2: Escalated parameters for the disease progression model.

Confidence intervals where estimated assuming the independence between variables and that the variables where normal distributed.

In addition, and as stated in section 2.1.1, the parameter λ was modified in order to improve the similarity with the Chilean reality. The parameter $\gamma^* = \exp(-9.429)$ was found by minimizing the absolute error between the distribution of stages from the P matrix at a steady state distribution, and the distribution of stages reported by [39].

Annexed G

Breast cancer model including screening mammograms

This appendix presents further details on the stochastic model from section 2.1.2.

A explicit form of the transition probability matrix P' will be shown. To simplify notation, define the following variables that depend on previous parameters.

$$\begin{aligned}
\widetilde{\rho} &:= \widetilde{\rho}(n) = 1 - \rho(n) \\
\widetilde{Mq_0} &:= \widetilde{Mq_0}(n) = 1 - M(n)q_0 \\
\widetilde{Mq} &:= \widetilde{Mq}(n) = 1 - M(n)q\left(V(\frac{n-\hat{N}}{12})\right) \\
\hat{P}_{S,S'} &:= \hat{P}_{S,S'}(n) = P_{S,S'}\left(\frac{n-\hat{N}}{12}, 1/12\right)
\end{aligned} \tag{G.1}$$

Then, the non homogeneus markov process (Z_n) with states $\{H, C_0, ..., C_3, D_0, ..., D_3, R\}$ is defined by its transition matrix $P'(n)$ shown in equation G.2. Figure 2.2 shows a diagram of this process.

$$P_1'(n) = \begin{bmatrix} 1-\widetilde{\rho}\phi_0-\rho & \widetilde{\rho}\phi_0 & 0 & 0 & 0 \\ 0 & \widetilde{\rho}\widetilde{Mq_0}(1-\phi_{0,1}-\lambda) & \widetilde{\rho}\widetilde{Mq_0}\phi_{0,1} & 0 & 0 \\ 0 & 0 & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_1,C_1} & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_1,C_2} & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_1,C_3} \\ 0 & 0 & 0 & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_2,C_2} & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_2,C_3} \\ 0 & 0 & 0 & 0 & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_3,C_3} \end{bmatrix}$$

$$P_2'(n) = \begin{bmatrix} 0 & 0 & 0 & 0 & \rho \\ \widetilde{\rho}(Mq_0+\widetilde{Mq_0}\lambda) & 0 & 0 & 0 & \rho \\ 0 & \widetilde{\rho}(Mq+\widetilde{Mq}\hat{P}_{C_1,D_1}) & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_1,D_2} & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_1,D_3} & \rho \\ 0 & 0 & \widetilde{\rho}(Mq+\widetilde{Mq}\hat{P}_{C_2,D_2}) & \widetilde{\rho}\widetilde{Mq}\hat{P}_{C_2,D_3} & \rho \\ 0 & 0 & 0 & \widetilde{\rho}(Mq+\widetilde{Mq}\hat{P}_{C_3,D_3}) & \rho \end{bmatrix}$$

$$P'(n) = \begin{bmatrix} P_1'(n) & P_2'(n) \\ 0_5 & I_5 \end{bmatrix} \tag{G.2}$$

Where 0_5 is a $5x5$ matrix with zeros and I_5 is the $5x5$ identity.