CUADERNOS
DE LA FUNDACIÓN

C / 222

Individual Cancer Mortality Prediction

Kailan Shang

Fundación
MAPFRE

Insurance and Social Protection Area

Individual Cancer Mortality Prediction

Kailan Shang

Fundación MAPFRE

Paseo de Recoletos, 23
28004 Madrid (España)

www.fundacionmapfre.org

ISBN: 978-84-9844-665-4
Depósito Legal: M-32620-2017
Maquetación y producción editorial: Arias Montano Comunicación
www.ariasmontano.com

ACKNOWLEDGEMENTS

The author would like to thank Ramon Alemany for his guidance, reviews, encourages and full support throughout this research. The author also wants to thank Lingyan Jiang for her insightful inputs, especially on the actuarial implication of this research. The author is grateful for funding sponsored by FUNDACIÓN MAPFRE[1].

[1] https://www.fundacionmapfre.org/fundacion/en/gettoknowus/who-we-are/

PRESENTATION

Since 1975 FUNDACIÓN MAPFRE has been carrying out general interest activities for society in different professional and cultural areas, in addition to many actions aimed at improving the economic and social conditions of the most disadvantaged people and sectors of society.

We work within the Insurance and Social Protection Area with the aim of promoting and disseminating knowledge on, and the cultures of, Insurance and Social Protection.

In terms of the activities aimed at society in general, we create free and universal content on insurance, which we make available via the *Seguros y Pensiones para Todos (Insurance and Pensions for everyone)* website. We organize training courses and educational and awareness-raising activities for teachers, as well as workshops for schoolchildren and free group visits to the Insurance Museum. We also publish informative guides to help people learn the basics of insurance.

In addition to this informative work, we support research by preparing reports on insurance markets and other subjects of interest, awarding research grants related to insurance and social protection, publishing books and surveys on insurance-related matters and organizing conferences and seminars. Our commitment to knowledge is embodied in our specialized Documentation Center that supports all our activities and is open to the general public.

These activities include the publication of this book, as the result of the research entitled *"Individual Cancer Mortality Prediction"* by Kailan Shang and with Ramon Alemany from the University of Barcelona as tutor of the work. This research was sponsored by FUNDACIÓN MAPFRE under the Research grant Ignacio H. de Larramendi in 2015.

All information regarding our activities is available online and accessible to users around the world quickly and efficiently via our website: www.fundacionmapfre.org

Fundación MAPFRE

Kailan Shang[2] is a Fellow of the Society of Actuaries (FSA), Chartered Financial Analyst (CFA) and Professional Risk Manager (PRM). He has more than ten years' experience on insurance, predictive modeling and risk management. He works at Swin Solutions Inc. as Managing Director providing actuarial and risk consulting services to insurance companies. Before that, he worked at Manulife, AIG and AEGON with various roles in the actuarial and risk management field. Kailan is also an enthusiastic researcher with more than 20 publications and won several research prizes including the Emerging Issues Research Prize from the Casualty Actuarial Society in 2011 and the Best Paper for Practical Risk Management Application from Joint CAS/CIA/SOA Risk Management Section in 2016. Kailan received his Master's degree in Economics from York University in Canada.

[2] Kailan Shang, FSA, CFA, PRM, SCJP, of Swin Solutions Inc., can be reached at kailan.shang@swinsolutions.com.

CONTENTS

EXECUTIVE SUMMARY

It is difficult for cancer patients, whether they are under treatment or in recovery, to get life insurance coverages. High mortality risk leads to low approval rate of life insurance applications. The premium rate is also very high. On the other hand, cancer incidence rate is quite stable while the overall mortality rate has been decreasing and more patients have a longer survival period. Better early detection, prevention, medical treatment in the future are likely to lead to lower mortality rates and longer survival periods. However, cancer patients may still face the difficult of getting life insurance coverages at a reasonable price even though they need the insurance protection.

To improve cancer patients and survivors' access to life insurance, it is necessary to evaluate the mortality risk of a cancer patient on an individual basis. Mortality and survivorship prediction based on data can be used to supplement doctor opinion. Detailed information such as age, gender, race, age at diagnosis, cancer type, tumor size, histological information, and received treatment can be used to assess the risk of the patient or survivor. Predictive models can be calibrated to either experience data or doctors' estimate, with adjustments for future changes such as medical development. With more accurate underwriting and pricing for individual patients, it could lead to a higher acceptance rate of insurance applications from cancer patients and lower premium rates for applicants with low risk.

Existing relevant literatures focus on the chance of survival. However, the term structure of mortality is also important for insurance product pricing. For a term life product, the term structure of mortality affects not only the amount but also the timing of death benefit and premium income. In addition, studies on the volatility of cancer mortality are rare which makes the risk assessment difficult. This report explores cancer mortality term structure and its volatility using diagnostic, medical treatment and demographic information.

DATA SOURCE

The Surveillance, Epidemiology, and End Results (SEER) research data submitted in November 2014 is used in the study. The dataset includes U.S. cancer incidence and population data from 1972 to 2012 including age, gender, race, year of diagnosis, geographic location, tumor size, tumor location, tumor type (benign or malignant), histological information, diagnostic result, medical treatment, survival period, and so on. The dataset has 8.7 Million data records and 146 variables.

Aggregate mortality rates by tumor site, gender and age are constructed using the SEER incidence data. They can be used as a benchmark and help us understand the general trend. For cancer patients, the mortality rate is very high in the early years after diagnosis, especially for respiratory and other digestive sites. The mortality rate becomes lower thereafter. A simple indication is that for cancer patients who survived 2-3 years after diagnosis, the mortality risk is much smaller and more insurable.

DATA PROCESSING

Before predicting the mortality rates of individual cancer patients, the raw data is analyzed and adjusted to prepare a clean input for predictive models. Some data fields are only applicable for a small portion of the total dataset. They are either removed or combined with others to cover the entire study period using the same set of coding system.

Data records with missing values are removed from the dataset if they are only a small portion of the total dataset. Alternatively, an unwanted or missing value is replaced by a specific value or a random value based on the distribution of other records. Some data fields have a constant value for all data records. They are not useful for explaining the variance of the explained variable and are removed from the analysis as well.

Most variables in the SEER data are categorical variables. These categorical variables are converted into dummy variables with the original categorical variable

removed to avoid multicollinearity. Given the large amount of data used in this report, collinearity is checked and each pair of variables with a correlation coefficient higher than 0.85 are reduced to one variable. This reduces some redundancy of the explanatory variables and improves the robustness of the model estimation.

MODEL

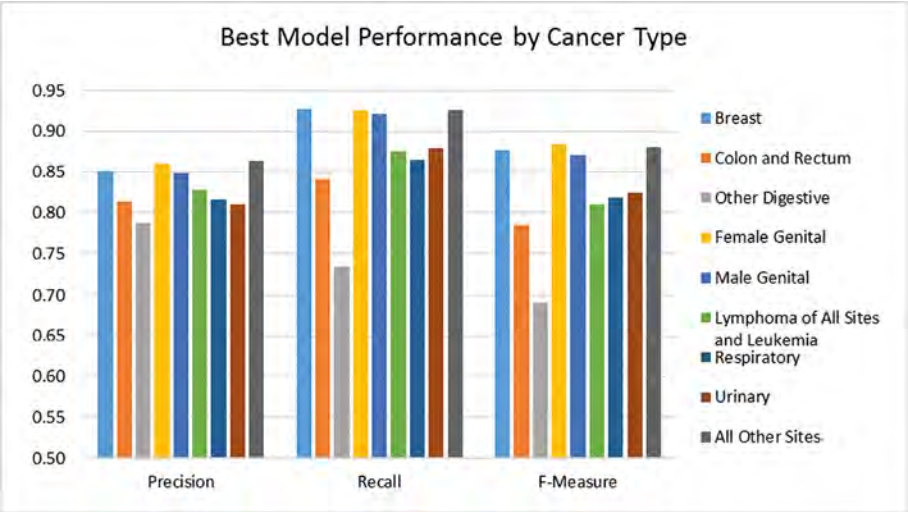
Unlike predicting the most likely survival period of a cancer patient, the probabilities of staying alive during each future year need to be predicted separately. 28 models are trained based on 28 data subsets for each cancer type. The first 25 models predict the annual mortality for the first 25 years after diagnosis. The remaining 3 models predict the 5-year mortality rates starting from the 26th, 31st and 36th year after diagnosis.

Six predictive models are trained to predict the mortality rates by cancer type. They include K-nearest neighbors (KNN), linear regression, Logistic regression, classification and regression tree (CART), random forest (RF), artificial neural network (ANN).

To evaluate the model accuracy consistently, precision, recall and F-measure based on the confusion matrix are used. Precision measures the type I error and recall measures the type II error. F-measure is the harmonic average of precision and recall. From the perspective of an insurer, a "false alive" is a big risk for life insurance products. False alive means that a patient who died was predicted to be alive by the model. Therefore, a low precision means an underestimation of cancer patient mortality risk.

The chosen criteria are used to compare and select the best models based on the validation result. For example, in the prediction of the breast cancer mortality rate during the 25th year after diagnosis, the Logistic regression model has a high precision rate but a low recall rate. The RF model marginally outperforms other models. The ANN models generally underperform other models. However, the difference in the performance is not material among models. A deep learning model with multiple hidden layers does not necessarily outperform a single layer ANN as well.

The model accuracy varies by cancer type. For some cancer type, mortality rate prediction is more credible based on the validation results. The following graph compares the precision, recall and F-Measure of the best models among cancer types.



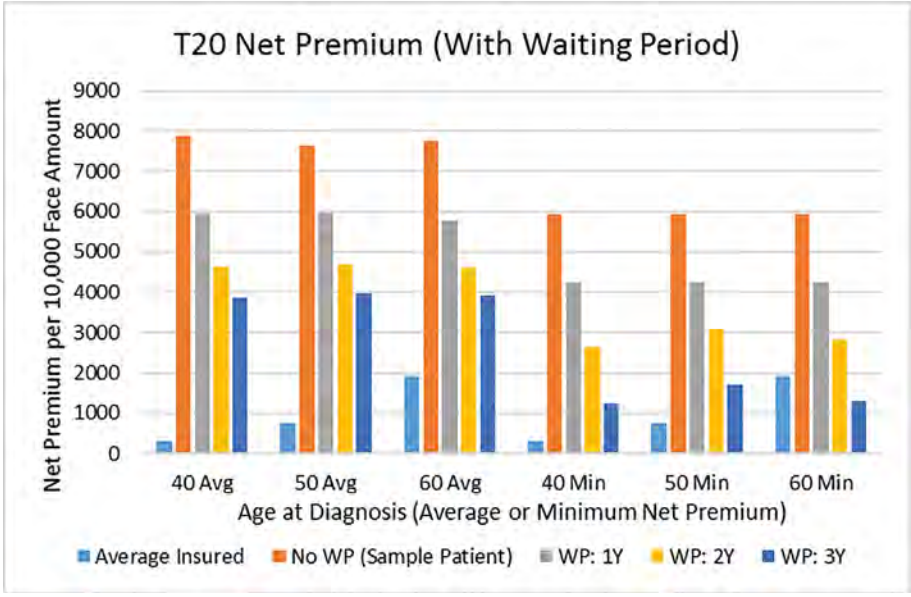
Unsupervised learning can be used to reduce the data dimension and shorten the time needed for supervised training. Unsupervised training techniques are tested to evaluate their impact on the prediction results. Using the dataset for studying the 25th year's breast cancer mortality rate, principal component analysis (PCA) is conducted to reduce 374 explanatory variables to 50 principal components, which explain 99.36% of the total variance. The first 50 principal components are then used as explanatory variables to predict the mortality rate. The PCA pre-training improves the validation results of logistic regression and artificial neural network models slightly. Deep Belief Network (DBN) models are also tested using two layers of restricted Boltzmann machines (RBMs). However, the validation results are much worse than a pure supervised training.

Caner type and year after diagnosis are used to separate the dataset for model training. Other key variables include age at diagnosis, age at 2012, cancer stage/grade information, surgery procedure, tumor size, primary site and lymph nodes. Age at 2012, the study cut-off time, reflects the trend of cancer mortality as medical treatments evolved over time. The key variables are quite stable among cancer types and prediction years.

INDIVIDUAL MORTALITY ASSESSMENT

After model building and assessment, a predictive model can be chosen based the weighted performance to predict the mortality rates for an individual cancer patient. The best model may vary by cancer type and year after diagnosis. A cancer patient's information can be fed into the calibrated models to calculate the mortality rate (probability of death) for each year after diagnosis. The predictive models can identify less risky patients who may be denied for insurance coverage before because of insufficient risk assessment.

A confidence interval can be constructed for the predicted mortality rate as well. Nonparametric methods can be used to estimate the interval using the results of similar cancer patients. By finding the nearest neighbors, the variance of their estimates can be used to derive the prediction interval in a practical way. The prediction interval can be used to measure the robustness and credibility of individual mortality rate prediction. With the individual mortality rate prediction, premium rate of life insurance products can be calculated and compared among individuals. The following graph illustrates the net single premium of a 20-year term life (T20) product for some sample breast cancer patients. Both the average and minimum premium of the sampled cancer patients are shown.



Insurance companies can also design a combined product including both life insurance and annuity. Life insurance and annuity have a natural hedge of mortality risk. With the knowledge of individual mortality rates, pricing and risk management of annuity products are feasible as well. It could further improve the insurability of cancer patients and the affordability of insurance coverages.

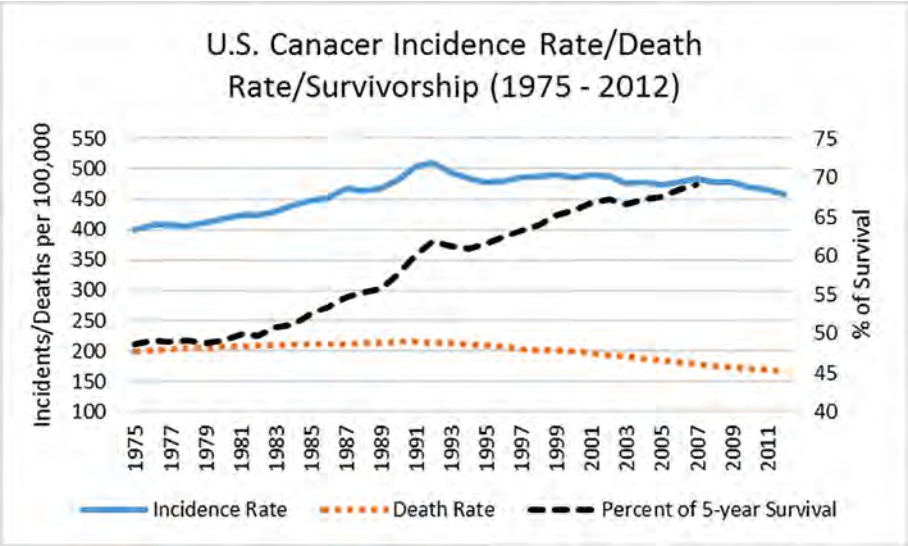
This study only uses part of the available data to predict the mortality rate of cancer patients. Other data such as genetic information and computed tomography scan of tumors can be used to improve the accuracy of prediction.

1. INTRODUCTION

For patients who are under medical treatment for cancer and people who have recovered from cancer, it is not very easy for them to get life insurance coverages. Due to the high risk, their insurance applications may be rejected by insurance companies according to the underwriting rules. There is no guarantee that a cancer patient or survivor can get life insurance regardless of his/her health conditions. Even if an application is approved, the cost of insurance may not be affordable. The most likely way for them to get life insurance coverages is through group life insurance such as an employer sponsored program without underwriting requirements for a minimum amount of protection.

The cancer incidence rate is quite stable while the overall mortality rate has been decreasing and more patients have a longer survival period. Figure 1 shows the cancer incidence rate, mortality rate and survivorship of the U.S. population from 1975 to 2012. The incidence rate increased from 400.42 per 100,000 in 1975 to 510.67 per 100,000 in 1992, and then gradually decreased to 457.19 in 2012. The mortality rate had a downward trend since 1991 with a rate of 215.22 per 100,000 decreasing to 166.52 per 100,000 in 2012. The percentage of cancer patients with a minimum 5-year survivor period increased from 48.69% in 1875 to 69.04% in 2007.

Figure 1. U.S. Cancer Statistics



Data: National Cancer Institute, NIH, DHHS, "Cancer Trends Progress Report."

Better early detection, prevention, medical treatment in the future are likely to lead to lower mortality rates and longer survival periods. However, cancer patients may still face the difficult of getting life insurance coverages at a reasonable price even though they need the insurance protection. In addition, the impact of cancer on mortality varies by individual. Not all the cases are of high risk. Without thorough evaluation on a case by case basis, there is a high chance that an insurable applicant is rejected or the premium rate is unnecessarily high for a low risk case.

To improve cancer patients and survivors' access to life insurance, it is necessary to evaluate the mortality risk of a cancer patient on an individual basis. Doctor's opinions can be used regarding the estimated survival period of a patient. However, it could be costly and sometimes inconsistent. Mortality and survivorship prediction based on data is another approach which can supplement Doctor's opinions. Detailed information such as age, gender, race, age at diagnosis, cancer type, tumor size, histological information, and received treatment can be used to assess the risk of the patient or survivor. Predictive models can be calibrated to experience data and doctor's estimate with results adjusted for future changes such as medical development. With statistically credible results, insurance companies can

use these models to assess the risk of cancer patients and survivors. Models can be customized using company-specific experience. With more accurate underwriting and pricing for individual patients, it could lead to a higher acceptance rate of insurance applications from cancer patients and lower premium rates for applicants with low risk.

Predictive models have already been widely used in cancer detection, diagnosis and prognosis using histological, clinical, family history and genomic data. They serve as a reference point and can help doctors make more informed decisions as well. William et al. (2006) used logistic regression models to predict the risk of breast cancer based on age, breast density, family history of breast cancer, and the use of hormone therapy. Park et al. (2013) compared three machine learning models for prognosis of breast cancer survival. The three models used are artificial neural network (ANN), supporting vector machine (SVM) and semi-supervised learning (SSL) model. Ahmed et al. (2013) used k-means clustering model to remove non-relevant data and applied AprioriTid algorithm to discover associate rules. The lung cancer is detected using decision tree models. Chen et al. (2014) used ANN to predict the survival of non-small cell lung cancer with genomic and clinical data. These are just a small subset of existing literatures on cancer risk prediction[3].

Advanced machine learning models such as deep learning have also been used in cancer detection. Deep learning is a complex machine learning model type based on the neural network. It uses multiple layers of non-linear relationships to extract features from the input data and then construct more complex features for prediction. Unlike traditional predictive models which have specified model format, deep learning allows a more flexible model structure and is able to reflect very complicated relationship between explanatory variables and explained variables by mimicking the neural network. On the other hand, it is difficult to implement deep learning because it is complex and difficult to optimize and make statistical inference. Most applications of deep learning in cancer research are related to image recognition. Estava et al. (2015) used convolutional neural network (CNN), a deep learning model, to classify skin cancer using images of skins.

Cancer mortality studies usually include statistics by age, gender, race, cancer type and social-economic status. Existing relevant literatures focus on the chance of survival.

[3] A compiled list of cancer risk prediction and assessment literatures can be found at http://epi.grants.cancer.gov/cancer_risk_prediction/

However, the term structure of mortality is also important for insurance product pricing. For example, for a term life product, the term structure of mortality affects not only the amount but also the timing of death benefit and premium income. In addition, studies on the volatility of cancer mortality are rare which makes the risk assessment difficult. With the proliferation of using predictive models in cancer prediction, cancer mortality prediction can be improved using these models as well.

This report explores cancer mortality term structure and its volatility using diagnostic, medical treatment and demographic information. The analysis is done based on the site of the tumor including breast cancer, digestive cancer, genital cancer, respiratory cancer and urinary cancer, and so on. Key factors for mortality prediction are sought for these cancer types separately. Several predictive models are applied and compared based on the accuracy of their mortality rate predictions. The report proceeds as follows:

- Section 2 (Data) explains the data used in the study, mortality rates on an aggregated level, and data processing.

- Section 3 (Predictive Models) discusses the details of predictive models used to study cancer mortality rates, methods of model training, result assessment and comparison among models.

- Section 4 (Individual Mortality Assessment) constructs the cancer mortality rate model for individual patients based on the predictive models. Both parametric and nonparametric estimation of prediction intervals is also explained.

- Section 5 (Actuarial Implication) discusses the application of cancer mortality analysis in the actuarial area, including insurance product design, pricing, underwriting and risk management.

- Section 6 (Future Development) touches on the potential improvements of supervised learning for cancer mortality prediction and suggests future research topics.

- Section 7 summarizes the key points of this research and concludes the main body of the report.

2. DATA

2.1 DATA SOURCE

The Surveillance, Epidemiology, and End Results (SEER) research data submitted in November 2014 is used in the study. The dataset includes U.S. cancer incidence and population data from 1972 to 2012 including age, gender, race, year of diagnosis, geographic location, tumor size, tumor location, tumor type (benign or malignant), histological information, diagnostic result, medical treatment, survival period, and so on. Each data record contains 146 variables, most of which are categorical.[4] The incidence dataset includes four subsets:

1. The data from 9 SEER registries including Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah for 1973-2012.

2. The data from 4 SEER registries including San Jose-Monterey, Los Angeles, Rural Georgia and Alaska Natives for 1992-2012.

3. The data from 5 SEER registries including Greater California, Kentucky, Louisiana, New Jersey, and Greater Georgia for 2000-2012. The Louisiana data for the second half of 2005 is not included in this subset because of hurricane Katrina's large impact on Louisiana's population during that period.

4. The data from Louisiana registry for the second half of 2005.

Due to its small data volume and undesired impact of short-term population volatility, the data from Louisiana registry for the second half of 2005 is excluded from the study in this report. Table 1 lists the data volume by tumor site in the SEER dataset.

[4] Detailed description of each data field can be found at http://seer.cancer.gov/data/seerstat/nov2014/TextData.FileDescription.pdf.

Table 1. SEER Data Volume (1972 – 2012)

Location of Tumor	No. of Records
Breast	1,384,575
Colon and Rectum	904,556
Other Digestive	660,518
Female Genital	666,973
Lymphoma of All Sites and Leukemia	681,151
Male Genital	1,172,566
Respiratory	1,141,701
Urinary	582,956
All Other Sites	1,484,223
Total	8,679,219

2.2 GENERIC MORTALITY RATE

Individual mortality rate assessment is more useful for insurance underwriting and pricing than aggregate mortality rate assessment. Aggregate mortality rates are still useful as a benchmark and help us understand the general trend. Tables of aggregate cancer mortality by tumor site, gender and age are constructed using the SEER incidence data. Table 2 lists the data fields used in developing the tables.

Table 2. Data Fields for Generic Cancer Mortality Table

Data Field	Description
Age at Diagnosis	The age of the patient at diagnosis for this cancer.
Month of Diagnosis	The month the tumor was first diagnosed by a recognized medical practitioner.
Year of Diagnosis	The month the tumor was first diagnosed by a recognized medical practitioner.
Sex	The gender of the patient at diagnosis
Survival Months – Presumed Alive	The number of months the patient survived by the end of the study period which is the end of 2012. If the patient was alive at last contact date which can be earlier than the end of the study period, it is assumed that the patient was alive at the end of 2012.

Cancer mortality tables are constructed following the steps given below.

1. The dataset is filtered by tumor size and gender. Data records with unknown diagnosis age and survival period are also removed from the analysis.

2. The number of months between diagnosis and the end of the study period is calculated as (2012 – Year of Diagnosis) × 12 + 12 – Month of Diagnosis. If it is equal to the variable "Survival Months – Presumed Alive", it is assumed that the patient was alive by the end of 2012.

3. Death count table is constructed with 131 rows (diagnosis ages 0 to 130) and 480 columns (months after diagnosis). Cell $\boldsymbol{D_{x,y}}$ represents the number of deaths happened at month $\boldsymbol{y}$ after diagnosis at age $\boldsymbol{x}$.

4. Initial exposure table is constructed with the same structure as the death count table. For the same diagnosis age, the diagnosis dates could be different. For example, the study period of a case diagnosed in 1973 is about 40 years while the study period of a case in 2007 is less than 6 years. The first case should affect the cancer mortality rate for 40 years and the second case should only affect the first 6 years. Initial exposure $\boldsymbol{E_{x,y}}$ is the number of patients diagnosed at age $\boldsymbol{x}$ and their study periods are longer than $\boldsymbol{y}$ months.

5. Mortality rate $\boldsymbol{I_{x,y}}$ is calculated as $\boldsymbol{D_{x,y}} / \boldsymbol{E_{x,y}}$. It is the rate that a cancer patient diagnosed at age x died in the yth month after diagnosis. It is different from the standard insurance mortality rate $\boldsymbol{q_x}$ which usually represents the death rate in the next period given that he/she is alive at age $\boldsymbol{x}$. $\boldsymbol{I_{x,y}}$ can be converted into the $q_{x,y} = \frac{I_{x,y}}{1 - \sum_{i=0}^{y-1} I_{x,i}}$ is the probability that a patient diagnosed with cancer at age x will die in the yth month after diagnosis given that he/she is alive at the end of the (y-1)th month.

6. Monthly mortality rates are converted into yearly mortality rates expect for the first year to smooth data and improve statistical credibility for each mortality rate in the table.

7. Survivorship $\boldsymbol{S_{x,y}}$ can be derived from mortality rates as $\prod_{i=1}^{y} 1 - q_{x,i}$.

Figure 2 compares the female cancer survivorship by tumor site and Figures 3 compares the male cancer survivorship for a diagnosis age of 45. A group of 1000 patients is assumed at initial with the number of expected survivors shown for the next 40 years. The survivorship of a standard insured person, either healthy or not, is also included in the comparison using mortality rates in the 2008 valuation basis table (VBT 2008).[5]

[5] https://www.soa.org/research/experience-study/ind-life/valuation/2008-vbt-report-tables.aspx

Figure 2. Cancer Survivorship by Tumor Site (Female, Age 45)

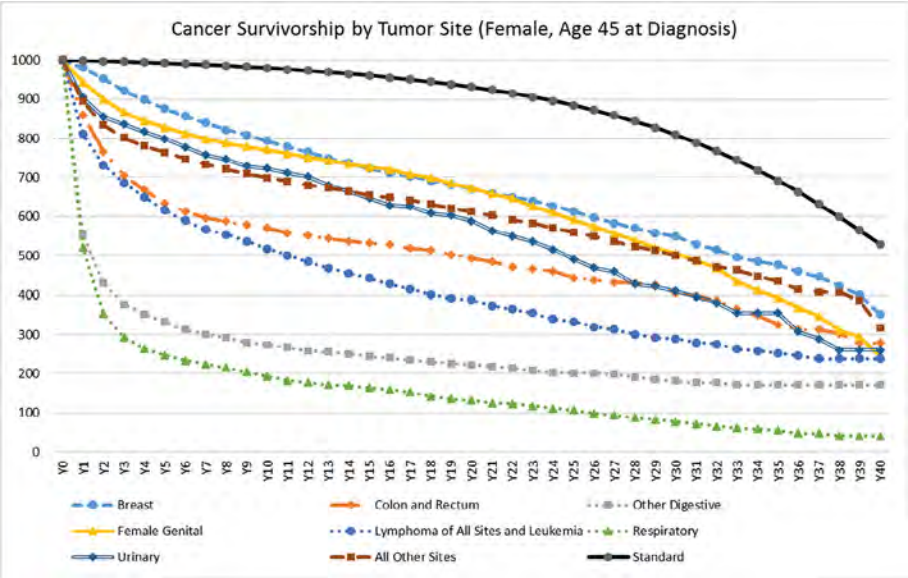
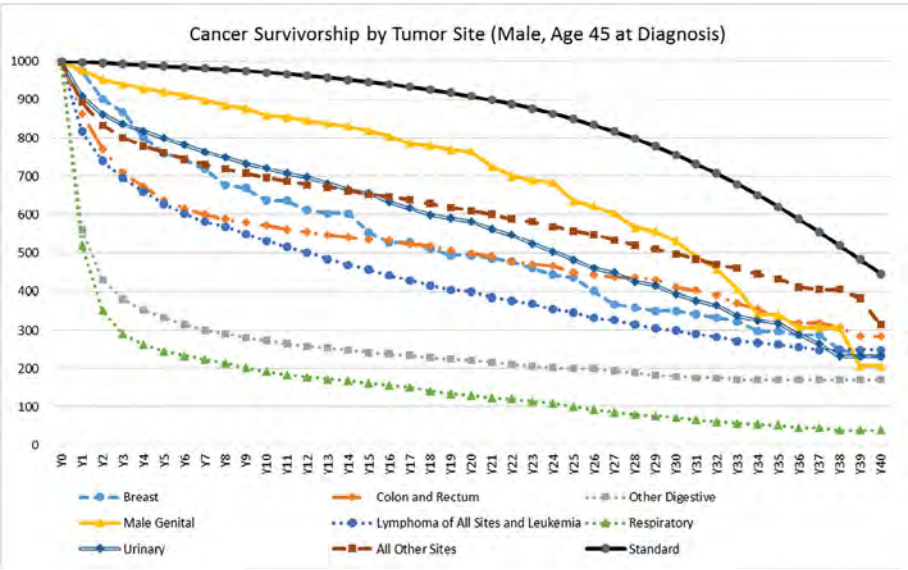


Figure 3. Cancer Survivorship by Tumor Site (Male, Age 45)



For cancer patients, the mortality rate is very high in the early years after diagnosis, especially for respiratory and other digestive sites. The mortality rate becomes lower thereafter. A simple indication is that for cancer patients who survived 2-3 years after diagnosis, the mortality risk is much smaller and more insurable.

For some tumor sites and ages at diagnosis, the number of incidences is too small to draw a statistically credible conclusion. In practice, it is necessary to assess the credibility, cross validate among ages and durations, and smooth the mortality rates appropriately before using the generic mortality tables and survivorship analysis. Table 3 shows the number of incidences used in the survivorship analysis for patients at age 45.

Table 3. Incidence Count for Age 45, Female and Male

Age 45, Female Tumor Site	Incidence Count	Age 45, Male Tumor Site	Incidence Count
Breast	27,650	Breast	135
Colon and Rectum	2,537	Colon and Rectum	2,659
Other Digestive	1,962	Other Digestive	2,820
Female Genital	9,554	Male Genital	9,554
Lymphoma of All Sites and Leukemia	2,635	Lymphoma of All Sites and Leukemia	3,398
Respiratory	2,861	Respiratory	4,114
Urinary	1,219	Urinary	2,922
All Other Sites	9,222	All Other Sites	10,587

Section A.1 lists the generic mortality tables by tumor site and gender. The incidence count is also included for assessment of credibility. Only ages with incidence count no less than 100 are listed to maintain some level of minimal credibility. Sometimes the mortality rate could be zero in a certain year for a certain age at diagnosis because there is no death reported in the dataset. It normally happens when the incidence count is small. In addition, the last contact date could be earlier than the end of the study period. In this analysis, the patient is assumed to be alive till the end of the study period if he/she was alive on the last contact date. Therefore, the mortality rate could be underestimated to some extent. However, the impact is small, if not negligible, considering the length of time assumed to be alive. Table 4 lists the average months between the last contact date and the study cut-off date.

Table 4. Average Duration of Assumed Alive for Age 45, Female and Male

Age 45, Female Tumor Site	Avg. Duration of Assumed Alive (Month)	Age 45, Male Tumor Site	Avg. Duration of Assumed Alive (Month)
Breast	4.2	Breast	2.8
Colon and Rectum	5.9	Colon and Rectum	4.9
Other Digestive	6.2	Other Digestive	6.7
Female Genital	11	Male Genital	2.9
Lymphoma of All Sites and Leukemia	3.4	Lymphoma of All Sites and Leukemia	3.7
Respiratory	5.5	Respiratory	5.2
Urinary	3.2	Urinary	3.8
All Other Sites	5.7	All Other Sites	5.3

Given the fast development of cancer research and medical treatment, cancer mortality rate changed over time. Generic mortality rate trend analysis is also meaningful for future insurance pricing and risk assessment although it may face the challenge of reduced credibility due to lower data volume. As the focus of this research in on individual assessment of cancer mortality, trend analysis is not covered here.

2.3 DATA PROCESSING

Before predicting the mortality rates of individual cancer patients, the raw data is analyzed and adjusted to prepare a clean input for predictive models.

Subjective Data Adjustment

In the SEER dataset, some data fields are only applicable for a small portion of the total dataset. They are either newly adopted by the data collection program or outdated coding systems that were not used anymore. Some of them are combined by aligning the coding system to cover the entire study period. For example, data variables “CS_EXT” and “EOD10_EXT” record the information on tumor extension from the primary site. Variable “CS_EXT” is used for cases after 2003 and variable “EOD10_EXT” is used for cases between 1988 and 2003. They are consolidated into one explanatory variable for

prediction. If a variable cannot be combined and the time coverage of that variable is short, the variable is removed from the set of explanatory variables because they can only provide limited insights for a portion of the entire study period and can only be used to predict the mortality rates for few years. Some variables have too many records with missing value and are removed from the analysis as well. Some variables are redundant because they contain the same type of information with different coding systems or statistical measures. Only one variable is kept for the same information. These subjective data adjustments lead to around 40% reduction of the number of explanatory variables.

Some variables have numerical values but are not likely to have a monotonic relationship with the mortality rate. They are converted to new variables to be used for predictive modeling. For example, data field "AGE_DX" is the age at diagnosis. For different cancer types, older patients do not necessarily have higher mortality rates. Newborn patients may have higher mortality rates. The variable is changed into about 20 new variables with each indicating whether the age at diagnosis is within the specified 5-year range.

Missing Value Treatment

Some variables have missing values for certain records. They may also contain values which have different meaning than the rest. For example, data field "EOD10_PN" is the number of regional lymph nodes examined by the pathologist that were found to contain metastases. This variable is supposed to contain the number of positive nodes. However, some values represent other meanings and need to be replaced, as shown in Table 5. Data records with missing values can be removed from the dataset if they are only a small portion of the total dataset. Alternatively, an unwanted or missing value can be replaced by a specific value, a random value based on the distribution of other records, the value of the most similar record, or a value predicted by some other variables. Table 5 shows the adjustments to some values of the variable "EOD10_PN" in the breast cancer dataset to make the variable comparable and consistent among records.

Table 5. Data Adjustment of Variable "EOD10_PN"

Value	Description	Percentage	Special Treatment
0-89	Number of positive nodes	64.7%	None
90	90 or more positive nodes	0.001%	None
95	Positive aspiration of lymph nodes was performed	0.3%	Random number (1-90)
97	Unknown number of positive nodes	0.3%	Random number (1-90)
98	Nodes not examined	21.5%	Change to 0
99	Unknown/NA/Not recorded	1.2%	Random number (0-90)
Blank	Missing value	12.0%	Random number (0-90)

*Random values are generated based on the distribution of records with values in the specific range.

Some data fields have a constant value for all data records. They are not useful for explaining the variance of the explained variable and are removed from the analysis.

Categorical Variable Treatment

Most variables in the SEER data are categorical variables. Some categorical variables have many possible categories. For example, the data field "HISTOLOGY (92-00) ICD-O-2" contains a 4-digit code that defines the histological type according to the International Classification of Diseases for Oncology, 2nd edition (ICD-O-2, 1992). 129 different codes exist in the breast cancer incidence dataset. The categorical variable is then converted into 129 dummy variables with the original categorical variable removed to avoid multicollinearity. Each dummy variable is associated with a specific code. It has a possible value of 0 and 1 indicating whether the code of the record is the one specified by the dummy variable. After the conversion, the number of variables increases by many times.

Collinearity

A variable can be predicted by its highly-correlated variables with high accuracy. Variables with high correlation may cause unstable estimates when using multivariable regression model. Even for models that are not vulnerable to collinearity, reducing variables with high correlation help reduce the data size and improve the model efficiency with little sacrifice of accuracy. Given the large amount of data used in this report, collinearity is checked and each pair of variables with a correlation coefficient higher than 0.85 are reduced to one variable. This reduces the number of variables by around 20%. Since

the dataset includes variables containing similar information with different coding systems, collinearity checking is a good way to reduce the redundancy of explanatory variables systematically without subjective judgement.

Data sampling

Given the large size of the dataset for each cancer type and the number of mortality rates to be predicted, subsets are generated for model calibration. Table 6 lists the subsets for predicting breast cancer mortality rates. Each subset contains all the patients that died in the specific year(s) after diagnosis. It also contains the same number of patients that are alive during that year, sampled from the entire dataset. The first 25 subsets contain data to train models predicting the mortality rates for each of the first 25 years. After that, 5-year data is grouped together because of the diminishing data volume and resulting model overfitting issues.

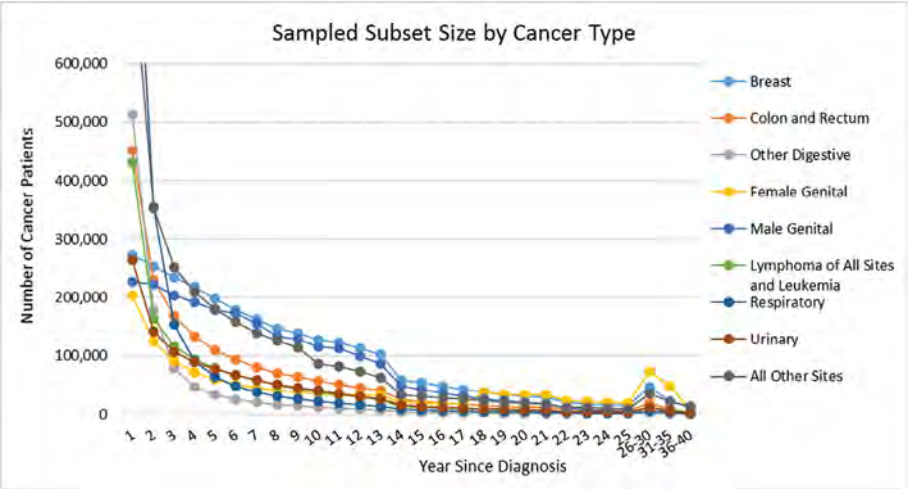
Table 6. Breast Cancer Subsets

Explained Variable: Mortality Rate	Number of Records			
	Death	Alive	Sampled Alive	Subset (Death + Sampled Alive)
1st Year	136,378	1,248,197	136,378	272,756
2nd Year	127,038	1,121,159	127,038	254,076
3rd Year	117,230	1,003,929	117,230	234,460
4th Year	108,944	894,985	108,944	217,888
5th Year	99,056	795,929	99,056	198,112
6th Year	89,467	706,462	89,467	178,934
7th Year	81,127	625,335	81,127	162,254
8th Year	73,136	552,199	73,136	146,272
9th Year	69,448	482,751	69,448	138,896
10th Year	63,829	418,922	63,829	127,658
11th Year	61,218	357,704	61,218	122,436
12th Year	56,608	301,096	56,608	113,216
13th Year	51,573	249,523	51,573	103,146
14th Year	29,590	219,933	29,590	59,180
15th Year	26,955	192,978	26,955	53,910
16th Year	23,928	169,050	23,928	47,856
17th Year	21,391	147,659	21,391	42,782
18th Year	18,943	128,716	18,943	37,886
19th Year	17,007	111,709	17,007	34,014
20th Year	15,281	96,428	15,281	30,562

Explained Variable: Mortality Rate	Number of Records			
	Death	Alive	Sampled Alive	Subset (Death + Sampled Alive)
21st Year	14,260	82,168	14,260	28,520
22nd Year	10,154	72,014	10,154	20,308
23rd Year	8,981	63,033	8,981	17,962
24th Year	8,002	55,031	8,002	16,004
25th Year	7,428	47,603	7,428	14,856
26th - 30th Year	23,537	24,066	23,537	47,074
31st - 35th Year	11,263	12,803	11,263	22,526
36th - 40th Year	6,033	6,770	6,033	12,066

Figure 4 shows the subset size of all cancer types studies in this report by time. A hike in the subset size happens for data period 26-30 years after diagnosis because of the grouping for all cancer types. In some rare cases, the death count is higher than the number of remaining survivors. All data are used without sampling in these cases.

Figure 4. Sampled Subset Size of Cancer Patients



Using the entire dataset is also a possible option and has been tested to measure the impact of data sampling on model accuracy. No material impact of sampling has been found, as discussed in Section 3.3. Because using the whole dataset requires a significant amount of runtime and high computing capability, using the sampling approach is more practical without a material loss of accuracy.

3. PREDICTIVE MODELS

Several predictive models are used to predict mortality rates of individual cancer patients.

3.1 MODEL SETUP

Unlike predicting the most likely survival period of a cancer patient, the probabilities of staying alive during each future year need to be predicted separately. As indicated in Table 6, 28 models are trained based on 28 data subsets for each cancer type. The first 25 models predict the annual mortality for the first 25 years after diagnosis. The remaining 3 models predict the 5-year mortality rates starting from the 26th, 31st and 36th year after diagnosis. The explained variable "LIVE" indicates whether the patient is alive or dead in the specific year after diagnosis. It has two possible values: 0 for death and 1 for alive.

In addition to cancer type and the year after diagnosis, other variables such as gender, race, and region can be used to further separate the data and set up predictive models on smaller subsets at a lower level. The methodology is the same expect that whether a data field is used as a data filter or an explanatory variable.

3.2 SUPERVISED TRAINING

Six predictive models are trained to predict the mortality rates by cancer type. They include K-nearest neighbors (KNN), linear regression, Logistic regression, classification and regression tree (CART), random forest (RF), artificial neural network (ANN).

1. KNN is a type of non-parametric models that predicts the explained variable based on the values of the k closest neighbors. In this report, the closeness is measured by Euclidean distance based on the explanatory variables after the data processing as described in Section 2.3. Different numbers of neighbors are tested and 5

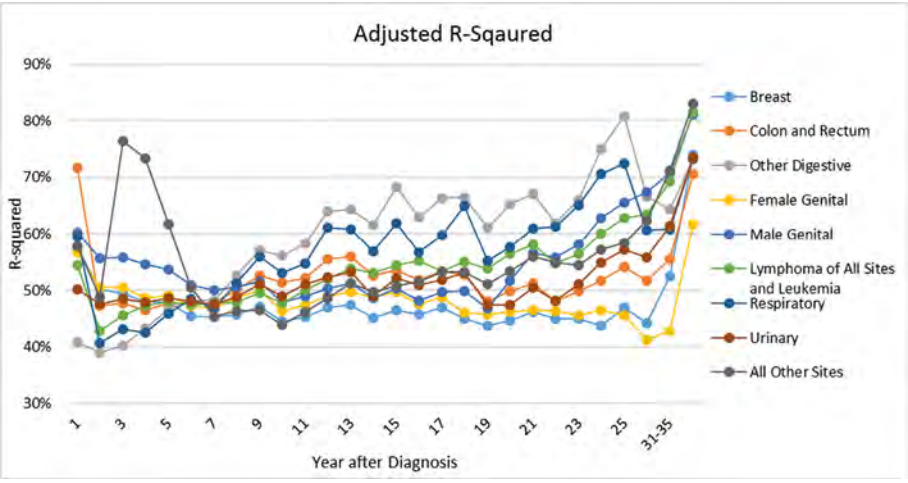
nearest neighbors are used for all models for its relatively better performance. The average survivorship of 5 nearest neighbors is then used to predict the mortality rate. For example, if 2 of the 5 neighbors were alive and 3 were dead, the probability of death in that year is predicted to be 60%. Accordingly, the probability of survival is 40%.

2. Linear regression is also used to predict the probability of survival assuming a linear relationship between the explained variable "LIVE" and all explanatory variables.

$$Live = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

The number of explanatory variables varies by model predicting the mortality rates of different years. The impact of data processing such as collinearity checking is different on different sampled subsets. Most linear regression models can explain around 40% to 70% of the variance in the explained variable for most years. Figure 5 shows the adjusted coefficient of determination (Adjusted R^2) for each linear regression model used to predict cancer patient mortality by cancer type. The sharp increase of adjusted R2 in the last 10 years is likely to be caused by overfitting. The size of the subsets is not big enough considering the number of explanatory variables. The F-test of the overall significance of linear regression analysis has a p-value less than 2.2×10^{-16} for all models. Considering both the adjusted R^2 and F-test results, the explanatory variables are valuable for predicting the mortality rates of cancer patients. In addition to these two traditional statistical measures, prediction accuracy based on the validation datasets is used to assess the linear regression models and compare them with other models using the same criteria, as discussed in Section 3.3.

Figure 5. Adjusted R-Squared of Linear Regression Models for Cancer Mortality



3. Logistic regression estimates the probability of the explained variable based on the logistic function given below.

$$prob(Live) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}$$

The explanatory variables used are the same for linear regression. In a traditional classification framework, if the probability of the explained variable "LIVE" is estimated to be greater than 50%, the patient is predicted to survive in that year. Otherwise, the patient is expected to die. The probability of "LIVE" is used as the predicted mortality rate.

4. CART models build trees to split the data based on explanatory variables. At each split, a variable is used to separate the data into two subgroups. The variable is chosen to provide the best split that improves the purity of the data in the subgroups. Gini index is used to represent the data dispersion. It is calculated as

$$G(T) = \sum_{i=1}^{n} p_i(1 - p_i) = 1 - \sum_{i=1}^{n} p_i^2$$

Where

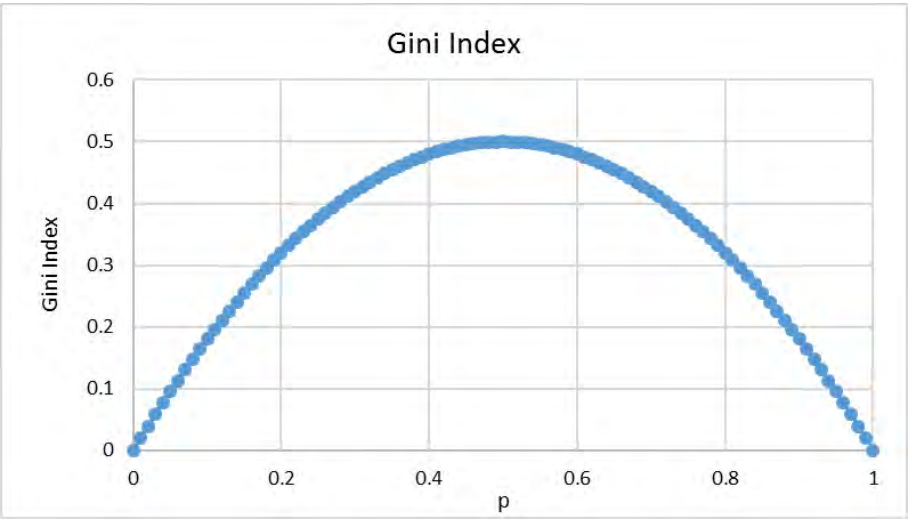
p_i is the probability that the data belongs to category i.

n is the number of categories in the data.

T is the dataset based on which Gini index is calculated.

If the data is pure, meaning that it only has one value, the Gini index is zero. If the data is evenly dispersed, such as 50% probability for each of two possible values, the Gini index is 0.5. Figure 6 shows the Gini index curve for data with only two categories. Gini index reaches the maximum when the probabilities are even between two categories.

Figure 6. Gini Index Curve



At each split, the increase in data purity in subsets is maximized when choosing the variable and the threshold for splitting.

$$\max_{x,threshold} G(T) - p(T_L)G(T_L) - p(T_R)G(T_R)$$

Where

T_L: The data subgroup of the split's left branch.

T_R: The data subgroup of the split's right branch.

p: The portion of the data subgroup in the dataset before splitting.

x: is the variable to be used for the splitting

$threshold$: the threshold used to set the split based on the value of x.

For example, a sampled cancer dataset has half of the data with variable "LIVE" equals 0 and the rest equals 1. If an explanatory variable x is found so that for all data with $x < 1$,

variable "LIVE" has a value of 0 and for all data with $x \geqslant 1$ variable "LIVE" has a value of 1, then the data purity after the splitting is perfect, as calculated below.

$$p(T_L)G(T_L) + p(T_R)G(T_R) = 0.5\times 0 + 0.5\times 0 = 0$$

In the calibration of CART models for cancer mortality prediction, three criteria are used to limit the size of the tree and avoid overfitting.

a. The minimum number of data points in a node for splitting is 20;

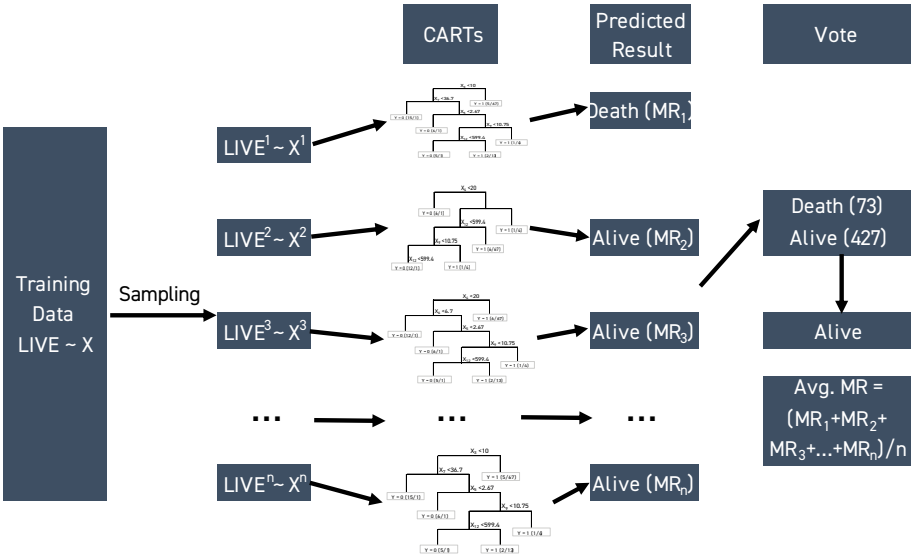
b. The minimum number of data points in a terminal node is 7;

c. The minimum improvement of data purity is 0.001 for each split.

CART models are useful for understanding the key variables that affect the mortality rates. Each terminal node has a mortality rate estimation based on the average value of the explained variable. For example, if the terminal node has 10 cancer patients with 3 deaths, the mortality rate is calculated as 3/10 = 30%.

5. Random Forest models are a random version of the CART models. Multiple subsets are sampled from the training dataset and each subset is used to build a CART model. Explanatory variables are sampled as well so that the relationship between the explained variable and the explanatory variables will not be dominated by the most important ones. Less important explanatory variables can contribute to the final prediction as well. Figure 7 illustrates the structure of the random forest models used in this report. The final mortality rate prediction is calculated as the average mortality rate predicted by the CART models.

Figure 7. Random Forest Model Structure



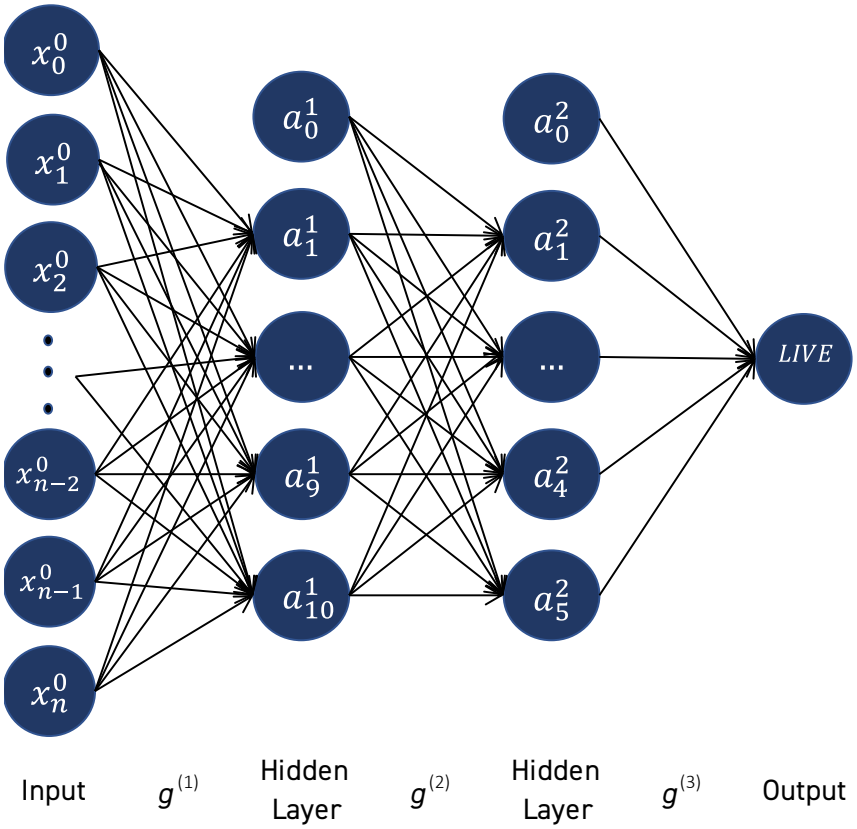
*MR: Mortality rate

500 trees are used in each RF model. The number of trees is at a medium level so that most of the data will be used in at least one of the CART models and overfitting is less likely. Given that the number of the explanatory variables is large, not all the variables are tested at each split in the CART models. At each split, 20 randomly chosen variables are tested to choose the best variable for splitting. The maximum depth of each tree is 20. Given the number of trees and the size of trees, each explanatory variable will get used in the RF model.

6. Artificial neural network models mimic biological neural networks to predict based on a large amount of data. Unlike traditional predictive models such as linear regression and logistic regression, the mathematical function that describe the relationship between the explained variable and explanatory variables is unknown. Figure 8 illustrates the ANN model structure used in this study. The ANN model has four layers: input layer that includes all the explanatory variables and the intercept, the first hidden layer with 10 neurons, the second hidden layer with 5 neurons and the output layer that includes the explained variable "LIVE". Since this ANN model contains more than one hidden layer, it is also a deep learning model to be able to

approximate complicated relationships. The number of neurons in each hidden layer is small to avoid overfitting.

Figure 8. Artificial Neural Network Model Structure



The layers are linked together with activation functions ***g***. Each neuron in the hidden layers and output layer is determined by the neurons in the previous layer.

First hidden layer: $a_0^1 = 0$ and $a_j^1 = g\left(\theta_j^0 \times x^0\right) for\ j > 0$

x^0: an (n+1) element column vector containing all the explanatory variables and the intercept x_0^0.

θ_j^0: an (n+1) element row vector containing the weights applied to all explanatory variables and the intercept to determine the value of neuron a_j^1.

Second hidden layer: $a_0^2 = 0$ and $a_j^2 = g(\theta_j^1 \times a^1)\ for\ j > 0$

a^1: an 11-element column vector containing all the neurons in the first layer.
θ_j^0: an 11-element row vector containing the weights applied to all the neurons in the first layer to determine the value of neuron a_j^1.

Output layer: $LIVE = g(\theta^2 \times a^2)$

a^2: a 6-element column vector containing all the neurons in the second layer.
θ^2: a 6-element row vector containing the weights applied to all the neurons in the second layer to determine the value of the output variable "LIVE".

Two activation functions are tested in model selection: logistic function and tanh function.

$$Logistic: g(x) = \frac{1}{1 + e^{-x}}$$

$$tanh: g(x) = \frac{2}{1 + e^{-2x}} - 1$$

The sum of squared errors is used as the error function to minimize for model calibration. L2 regularization is also added to mitigate the impact of overfitting. The neural network is a free forward network. Backpropagation algorithm is used for model training

$$\min_{\theta} \sum_{i=1}^{N} (LIVE_i - \overline{LIVE_i})^2 + \lambda \sum_{i=1}^{n} \theta_i^2$$

Where
N: total number training data records.
$LIVE_i$: the actual value of variable "LIVE".
$\overline{LIVE_i}$: the predicted value of variable "LIVE".
θ_i: The value of ith model parameter.
λ: Rate of L2 regularization. A λ of 0.0001 is used in the analysis.

Resilient backpropagation algorithm introduced by Reidmiller and Braun (1992) is used for fast ANN parameter optimization.

3.3 ASSESSMENT

To evaluate the model accuracy consistently, precision, recall and F-measure based on the confusion matrix are used. Precision measures the type I error and recall measures the type II error. F-measure is the harmonic average of precision and recall. From the perspective of an insurer, a "false alive" is a big risk for life insurance products. False alive means that a patient who died was predicted to be alive by the model. Therefore, a low precision means an underestimation of cancer patient mortality risk.

Table 7. Confusion Matrix

	Predicted: Alive	Predicted: Death
Actual: Alive	True Alive	False Death
Actual: Death	False Alive	True Death

$$Preision = \frac{True\ Alive}{True\ Alive + False\ Alive}$$

$$Recall = \frac{True\ Alive}{True\ Alive + False\ Death}$$

$$F - measure = \frac{2 \cdot Preision \cdot Recall}{Precision + Recall}$$

The model outputs are probabilities of death. The explanatory variable "LIVE" only has two values: 0 for death and 1 for being alive. It is necessary to translate the probabilities to a binary variable. Receiver operating characteristic (ROC) curve is used to understand the tradeoff between the true positive rate and the true negative rate by varying the threshold.

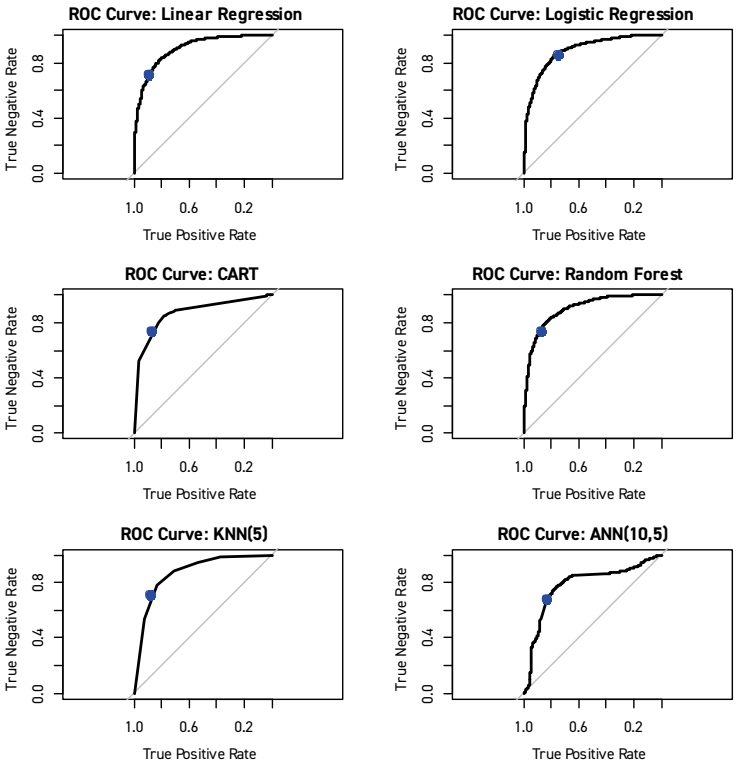
$$True\ Positive\ Rate\ (Recall) = \frac{True\ Alive}{True\ Alive + False\ Death}$$

$$True\ Negative\ Rate = \frac{True\ Death}{True\ Death + False\ Alive}$$

Figure 9 shows the ROC curve of outputs from the six models predicting the 25th year's mortality rate (probability of death) of breast cancer patients after diagnosis. In the

evaluation, 0.5 is used as a threshold to determine whether a patient is predicted to be alive ("LIVE" = 1) or dead ("LIVE" = 0), shown as the dots in the figures.

Figure 9. ROC Curves of Predictive Results for Breast Cancer 25th Year Mortality Rate



Notes:

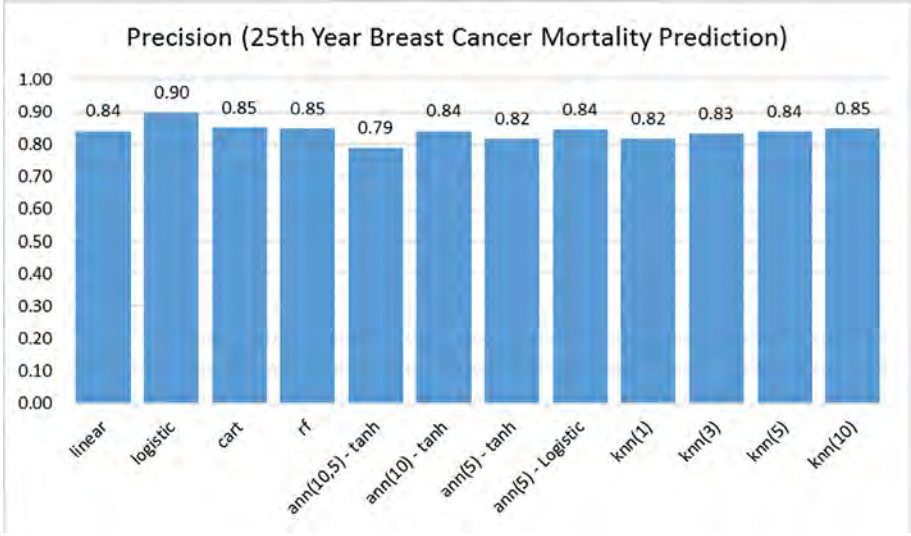
KNN(5): 5 nearest neighbors.

ANN(10,5): Artificial neural network model with two hidden layers having 10 and 5 neurons, respectively.

All the dots with a threshold of 0.5 on the ROC curves have a good balance of the true positive rate and the true negative rate. Other thresholds may be used to improve a preferred measure such as precision, recall or F-measure in practice. However, choosing the exact threshold based on the validation data could be misleading and sometimes arbitrary. Another set of validation data may give another optimal threshold. Therefore, a general threshold of 0.5 is used for all models.

The chosen criteria (precision, recall and F-measure) are used to compare and select the best models based on the validation result. Figure 10 to Figure 12 illustrate the model performance of predicting the breast cancer mortality rate of the 25th year after diagnosis. The Logistic regression model has a high precision but low recall rate. The KNN(5) model outperforms other KNN models. The RF model marginally outperforms other models. The ANN models generally underperform compared to linear regression model, CART, RF and KNN(5). However, the difference in performance is immaterial. A deep learning model with multiple hidden layers does not necessarily outperform a single layer ANN as well. ANN(10,5) has a lower precision rate than single layer ANN models but a higher recall rate, as indicated in the analysis. Changing the activation function also has limited impact.

Figure 10. Breast Cancer Mortality Prediction Sample Precision Results



Notes:

KNN(x): x nearest neighbors.

ANN(10,5) - tanh: Artificial neural network model with two hidden layers having 10 and 5 neurons, respectively. The activation function is tanh.

ANN(x) - tanh: Artificial neural network model with one hidden layers with x neurons. The activation function is tanh.

Figure 11. Breast Cancer Mortality Prediction Sample Recall Results

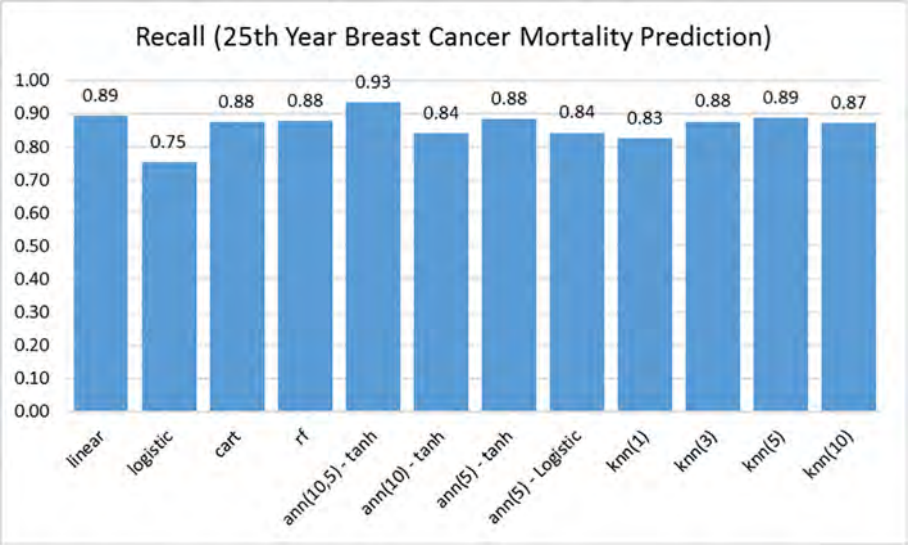
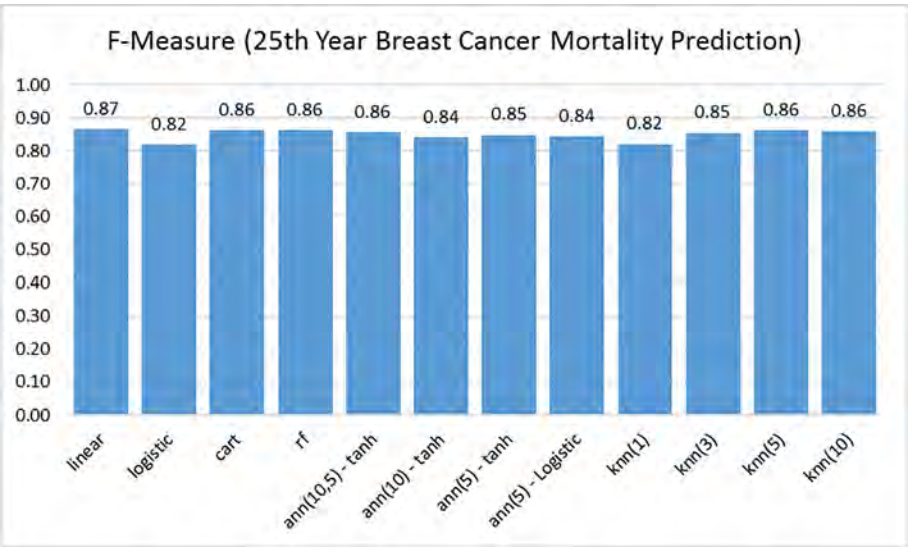


Figure 12. Breast Cancer Mortality Prediction Sample F-Measure Results



All models have a F-Measure value between 0.8 and 0.9. Based on the model performance assessment for the 25th year mortality rate prediction, linear regression, Logistic, CART, RF, ANN(10,5) and KNN(5) are selected for mortality rate prediction for all other years after diagnosis. Figure 13 to Figure 15 show the model performance for predicting breast cancer mortality rates for 40 years by the four models. The prediction performance is quite stable across the years with increased volatility at later years for all the models except the Logistic regression models. The RF, ANN(10,5) and linear regression models in general have a marginal outperformance than others.

Figure 13. Breast Cancer Mortality Prediction Precision Results

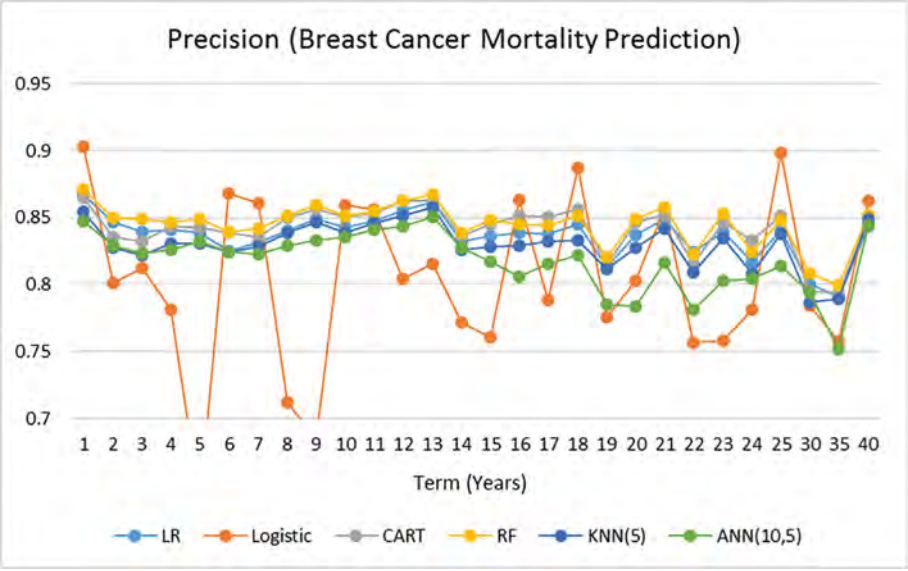


Figure 14. Breast Cancer Mortality Prediction Recall Results

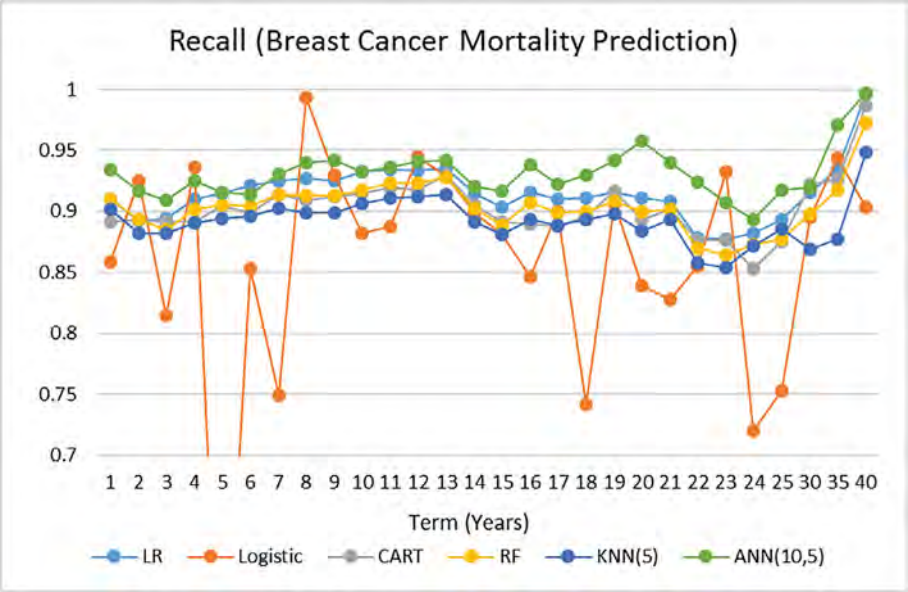
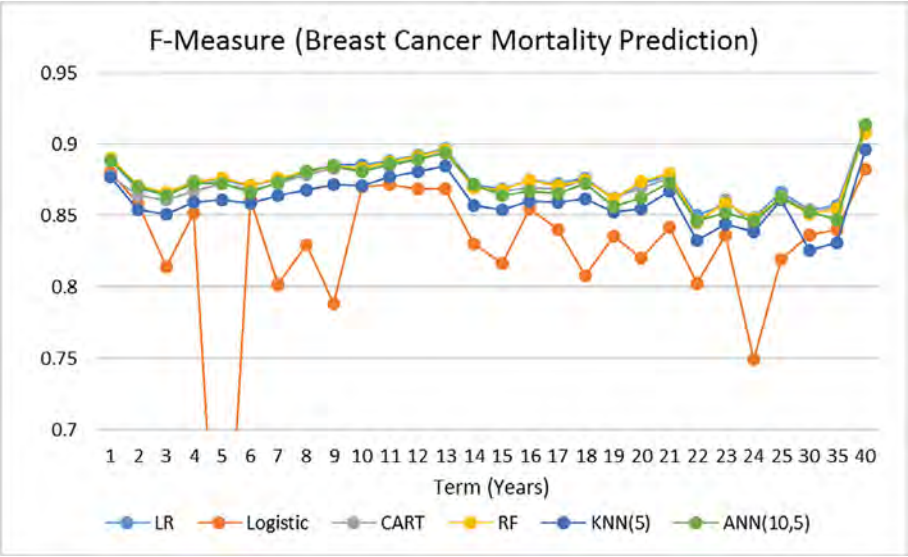
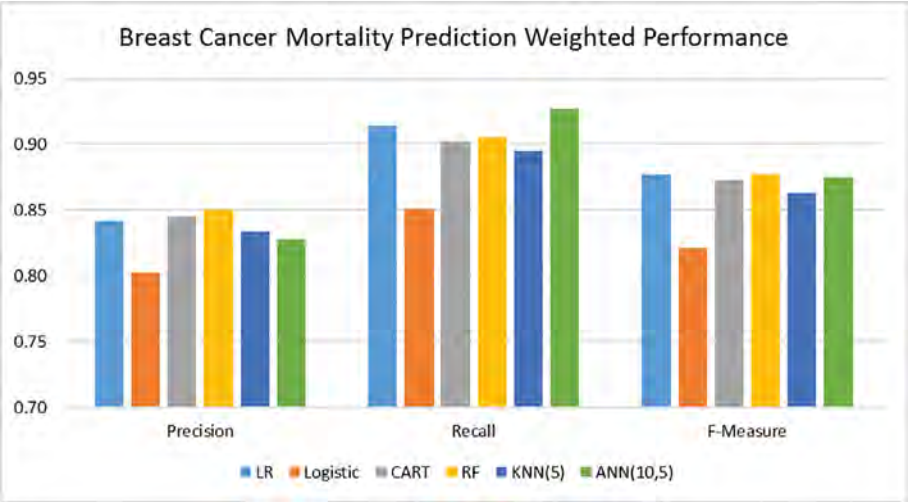


Figure 15. Breast Cancer Mortality Prediction F-Measure Results



To facilitate the comparison, weighted average measures can be used to represent the overall performance of predicting 40 years' mortality rates. The weight is the size of the subset used for each model. Figure 16 compares the weighted performance of the six models for breast cancer mortality prediction.

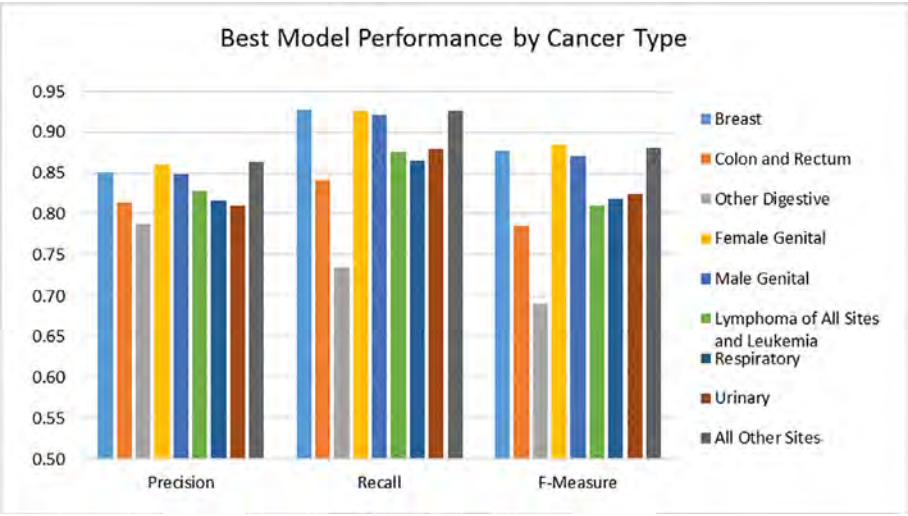
Figure 16. Breast Cancer Mortality Prediction Weighted Performance



All the assessment results discussed above are for breast cancer mortality prediction. Appendix A.2 summarizes the model performance for other cancer types as well. In general, RF models have slightly better performance than others.

The model accuracy varies by cancer type. For some cancer type, mortality rate prediction is more credible based on the validation results. Figure 17 compares the precision, recall and F-Measure among cancer types. The maximum precision, recall, and F-measure of the six models are shown for each cancer type. The selection is based on each measure, rather than three measures together. Therefore, the model with the highest precision could be different from the model with the highest recall or F-measure.

Figure 17. Best Model Performance by Cancer Type



Breast, female genital, male genital and all other sites cancer mortality prediction have the highest accuracy, with F-measure above 85%. Other digestive cancer has the lowest accuracy, with the F-measure below 70%, which may not be accurate enough for insurance application. Table 8 lists the model type with the best performance for each measure and cancer type. RF, ANN and Logistic regression models have a better performance than others considering the aggregate F-measure. However, the Logistic regression models have a very volatile performance by year despite their marginal advantage in some cases. Therefore, RF and ANN models are more preferred in cancer patient mortality prediction.

Table 8. Mode Type with Best Performance

Cancer Type	Precision	Recall	F-Measure
Breast	RF	ANN	RF
Colon and Rectum	LR	ANN	ANN
Other Digestive	RF	ANN	ANN
Female Genital	RF	ANN	RF
Male Genital	RF	ANN	RF
Lymphoma of All Sites and Leukemia	LR	RF	Logistic
Respiratory	LR	Logistic	Logistic
Urinary	LR	Logistic	Logistic
All Other Sites	LR	Logistic	Logistic

As discussed in Section 2.3, sampled subsets are generated for predicting each year's mortality rate to reduce the computing time. However, it is necessary test whether the sampling has a material impact on the model result. Figure 18 compares the model performance using both sampled datasets and the total dataset of breast cancer patients. The model performance based on the sampled datasets is a little better than the total dataset for all the tested predictive models except the ANN model. The impact of data sampling is mostly positive and immaterial for breast cancer.

Figure 18. Impact of Breast Cancer Data Sampling (25th Year's Mortality Rate)

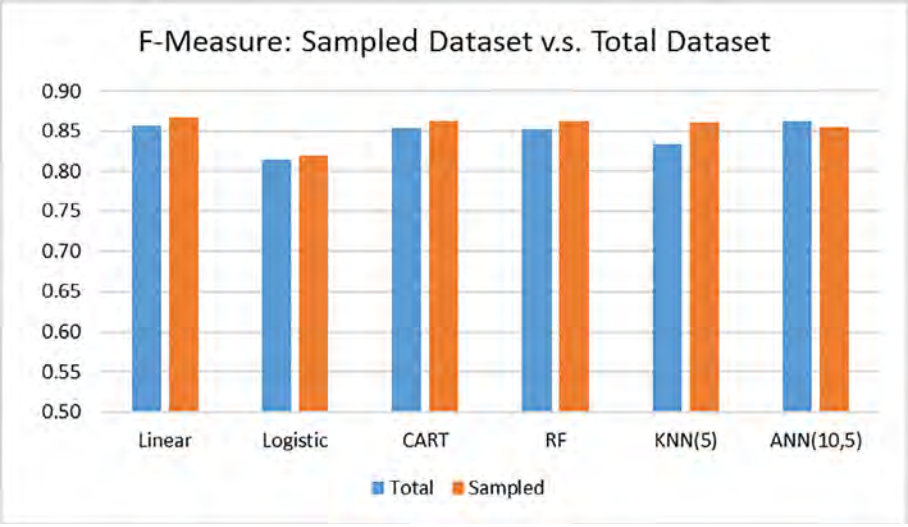


Table 9 lists the minimum, maximum and average impact of sampling on the F-measure among 6 models for predicting the 25th year's mortality rates after diagnosis for each cancer type. The average impact is positive for all cancer types except for all other sites cancer. The maximum positive impact of sampling exceeds 13% for three cancer types: Colon and Rectum, other Digestive, and Respiratory. This may indicate that the relationships are more easily to be found through a smaller dataset than the entire dataset. The negative impact can be as much as 3%, which is not significant considering that the average F-measure is greater than 75%. Appendix A.3 lists the detailed results for all cancer types.

Table 9. Sampling Impact on F-Measure

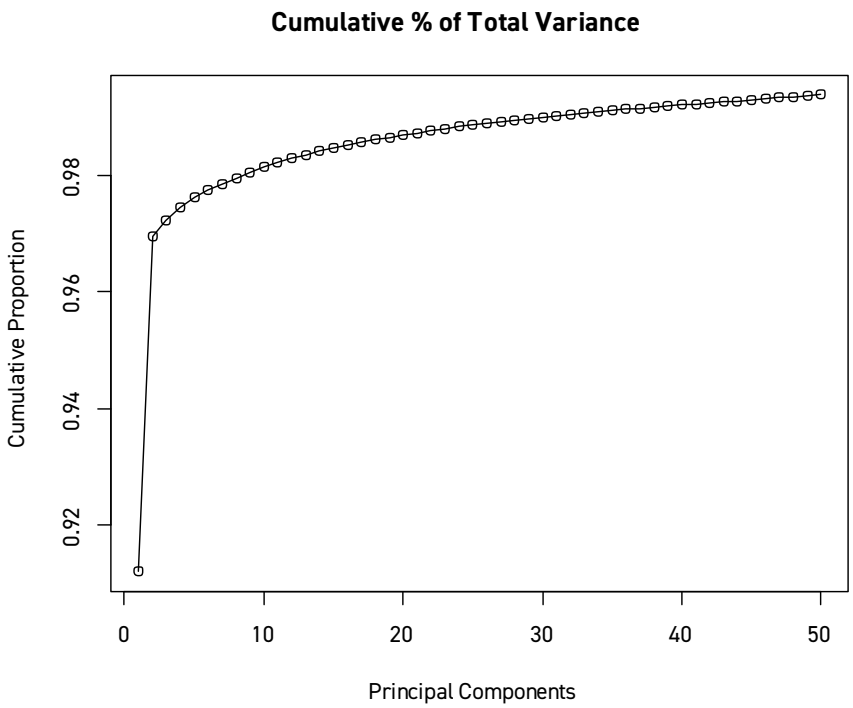
Cancer Type	Average F-measure without Sampling	Impact on F-measure		
		Minimum	Maximum	Average
Breast	85%	-1%	3%	1%
Colon and Rectum	76%	1%	13%	4%
Other Digestive	78%	1%	19%	11%
Female Genital	90%	-3%	1%	0%
Male Genital	86%	0%	5%	2%
Lymphoma of All Sites and Leukemia	88%	0%	1%	0%
Respiratory	75%	5%	18%	12%
Urinary	81%	-1%	1%	1%
All Other Sites	90%	-3%	0%	-1%

3.4 UNSUPERVISED PRE-TRAINING

In supervised training, random values are used as initial values of parameters for model calibration. Sometimes unsupervised training techniques are useful for pre-training that can improve the supervised training. In the presence of a large number of data inputs, unsupervised pre-training techniques such as principal component analysis (PCA) and restricted Boltzmann machine (RBM) can help reduce the data dimension and shorten the time needed for supervised training.

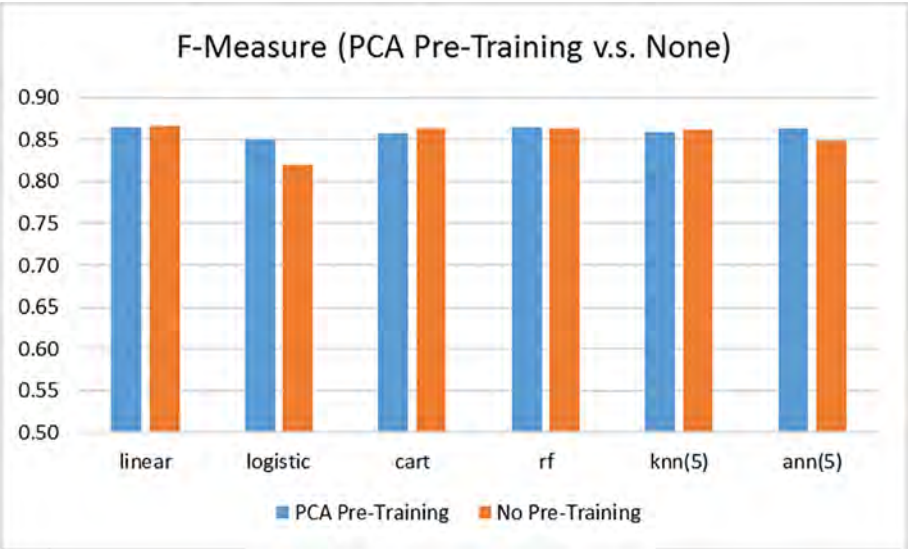
PCA transforms the explanatory variables into orthogonal variables with the first few principal components explaining the majority of the variance in the dataset. Using the dataset for studying the 25th year's breast cancer mortality rate, PCA is conducted to reduce 374 explanatory variables to 50 principal components, which explain 99.36% of the total variance. Figure 19 shows the cumulative variance explained by an increasing number of principal components.

Figure 19. Cumulative Proportion of Total Variance in PCA



The first 50 principal components are then used as explanatory variables to predict the mortality rate. The PCA pre-training improves the validation results of logistic regression and artificial neural network models slightly. Figure 20 compares the F-Measure results with and without the PCA. The RF model without pre-training still has the highest F-Measure.

Figure 20. Breast Cancer Mortality Rate Prediction F-Measure with PCA



Deep Belief Network (DBN) models are also tested using two layers of restricted Boltzmann machines. The two layers of RBMs are trained to extract the information of the explanatory variables in an unsupervised way. The output layer is trained in a supervised way to predict the explained variable "LIVE". However, the validation results are much worse than a pure supervised training.

Unsupervised pre-training is effective when lower data accuracy is acceptable. For example, in image recognition, the colors are not very important and can be transformed to a gray scale for object recognition. In cancer mortality rate projection, data accuracy is important as slight differences in certain key variables could have material impact on the mortality rate. The number of explanatory variables is only a few hundred which makes the computing power less of a concern. Therefore, using unsupervised pre-training is not necessary and is unlikely to improve the model accuracy by much in this case.

3.5 KEY EXPLANATORY VARIABLES

It is also useful to know the explanatory variables that have relatively big impact on the mortality rate prediction. Caner type and year after diagnosis are used to separate the

dataset for model calibration and are therefore key variables. Other key variables can be found by comparing their contributions to the prediction. In this report, linear regression, CART, and RF model are used to identify the key variables. For a linear regression model, the explanatory variables are normalized to the range of [0,1]. A variable's importance is measured by the absolute value of the coefficient of that variable. For a CART model, a variable's importance is measured by the increase of data purity because of a split based on that variable. If the variable is used in more than one split, the total impact of all relevant splits is used. For a RF model, a variable's importance is measured as the average reduction of squared error caused by the variable across all trees in the RF model.

$$Imp(x_i) = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{S}\frac{N_L \cdot N_R}{N_L + N_R}(\overline{Y_L} - \overline{Y_R})^2 \cdot Ind(split = x_i)\right)$$

Where

x_i: The ith input variable.

T: Total number of CART models in the RF model.

S: Total number of splits in a CART model.

$\overline{Y_L}$: The mean of Y in the left node after the split.

$\overline{Y_R}$: The mean of Y is the right node after the split.

N_L: The number of records in the left node after the split.

N_R: The number of records in the right node after the split.

$Ind(split = x_i)$: Indicator function with a value of 1 if the split is based on variable xi and a value of 0 otherwise.

Table 10 lists the top 10 variables based on the linear regression model, the CART model and the RF model used to predict the first-year breast cancer mortality rate after diagnosis. The weight is normalized so that the total weight for all the explanatory variables is 100. As category variables are converted to dummy variables before fed into the models, dummy variables are grouped into the original variables to determine the importance. Variables that contain similar information are grouped together as well. Age at diagnosis, age at 2012, cancer stage information and lymph nodes are important variables for both models. Age at 2012 reflects the trend of cancer mortality as medical treatments evolve over time. With the same age at diagnosis, an older age at 2012 indicates that the cancer incidence happened in an earlier time.

Table 10. Variable Importance in Breast Cancer Mortality Prediction

Linear		CART		RF	
Variable	Weight	Variable	Weight	Variable	Weight
Number of regional lymph nodes	32	**Age at 2012**	42	**Stage Information**	18
Age at Diagnosis	8	**Age at Diagnosis**	16	**Age at 2012**	16
Number of regional lymph nodes removed or examined	7	**Stage information**	16	Tumor type (positive/ negative)	8
Surgery procedure of primary site	7	Tumor type (positive/ negative)	8	Involvement of lymph nodes	6
Surgery type (site)	6	Insurance Status	4	Age at Diagnosis	5
Race	5	Primary site	2	Insurance Status	4
Age at 2012	4	Reason for no surgery	2	Surgery procedure of primary site	4
Number of regional lymph nodes that were found to contain metastases	4	Tumor extension	1	Tumor Size	3
Primary site and histology for children	4	Involvement of lymph nodes	1	Reason for no surgery	2
Stage information	3	Number of regional lymph nodes that were found to contain metastases	1	Estrogen Receptor (ER) Assay	2

Table 11 examines the changes in the key variables with the year after diagnosis using linear regression models. Comparing the key variables for predicting the first-year mortality rate and the 20th year mortality rate, 8 out of the top 10 variables are the same, although the degree of importance changes for each variable.

Table 11. Variable Importance in Breast Cancer Mortality Prediction by Year

1st Year Mortality Rate Variable	Weight	20th Year Mortality Rate Variable	Weight
Number of regional lymph nodes	32	**Number of regional lymph nodes removed or examined**	41
Age at Diagnosis	8	Histology information	8
Number of regional lymph nodes removed or examined	7	**Surgery type (site)**	7
Surgery procedure of primary site	7	**Primary site and histology for children**	6
Surgery type (site)	6	**Number of regional lymph nodes that were found to contain metastases**	6
Race	5	**Age at 2012**	5
Age at 2012	4	**Race**	4
Number of regional lymph nodes that were found to contain metastases	4	**Age at Diagnosis**	3
Primary site and histology for children	4	**Stage information**	2
Stage information	3	Primary site and morphology	1

Table 12 shows the importance of 12 main variables for each cancer type when predicting the first-year mortality rate after diagnosis based on the RF models. They can explain about 65%-80% of the model prediction power.

Table 12. Variable Importance by Cancer Type (%)

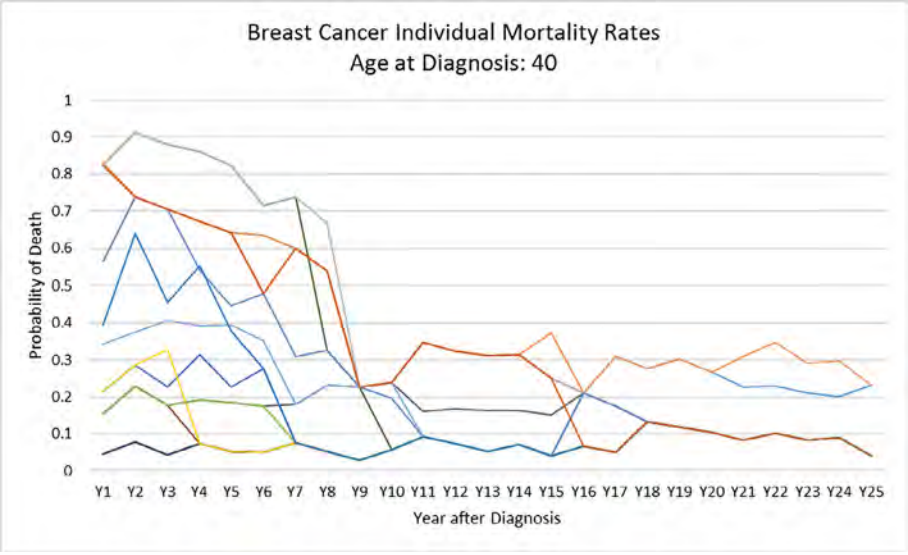
Variable	Cancer Type								
	Breast	Colon and Rectum	Other Digestive	Female Genital	Male Genital	Lymphoma of All Sites and Leukemia	Respiratory	Urinary	All Other Sites
Age at 2012	15.7	10.1	9.7	20.6	22.7	27.4	15.2	18.9	13.7
Age at diagnosis	5.3	3.1	4.4	7.1	4.9	4.7	5.3	4.9	1.9
Stage information	18.5	11.1	9.6	14.0	16.7	10.5	15.3	18.8	13.0
Insurance status	4.1	8.0	8.6	2.6	6.4	6.0	4.4	5.2	2.9

Variable	Cancer Type								
	Breast	Colon and Rectum	Other Digestive	Female Genital	Male Genital	Lymphoma of All Sites and Leukemia	Respiratory	Urinary	All Other Sites
Tumor type (positive/ negative)	9.6	<1%	<1%	4.5	<1%	<1%	9.1	<1%	3.4
Involvement of lymph nodes	5.8	6.5	9.0	4.4	12.1	7.8	8.4	10.0	6.5
Reason for no surgery	2.0	1.1	<1%	3.3	<1%	1.2	2.0	<1%	3.6
Primary site	3.5	3.7	8.6	4.5	5.8	9.0	3.6	3.7	16.9
Surgery procedure	1.9	2.6	3.2	2.2	4.0	4.9	1.5	3.5	3.6
Histological type	<1%	<1%	<1%	2.3	1.0	5.1	<1%	2.3	4.3
Tumor size and extension	4.4	11.3	11.7	6.7	3.4	3.0	4.7	4.5	7.6
Grade	1.0	<1%	1.6	1.5	<1%	5.0	1.4	1.3	<1%
Total	71.0	57.4	64.8	72.2	77.0	79.6	69.4	71.6	77.4

4. INDIVIDUAL MORTALITY ASSESSMENT

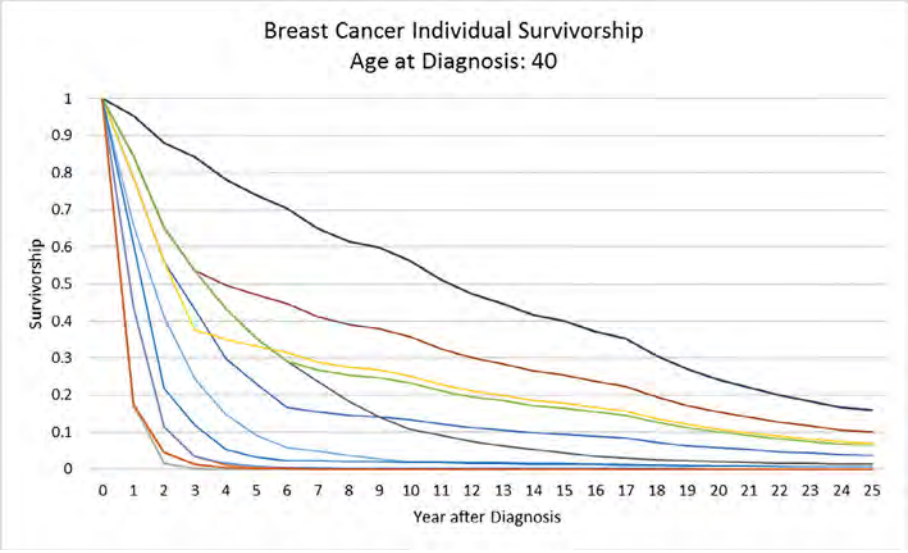
After model building and assessment, a predictive model can be chosen based the weighted performance to predict the mortality rates for an individual cancer patient. The best model may vary by cancer type and year after diagnosis, according to the model setting in this study. A cancer patient's information can be fed into the calibrated models to calculate the mortality rate (probability of death) for each year after diagnosis. If the patient's information is not complete compared to the data fields available in the SEER database, missing data can be treated using the methods discussed in Section 2.3. It may be replaced by the average value, the median value, the value of the nearest neighbors, or the value predicted based on other available information. Figure 21 illustrates the predicted mortality rates of a few female breast cancer patients diagnosed at age 40 using CART models.

Figure 21. Sample Individual Breast Cancer Patient Mortality Rates (Age at Diagnosis: 40)



The mortality rates are very high during the first few years after diagnosis and then decrease significantly after 8 years. The mortality rates are very volatile among the patients. The predictive models are able to identify less risky patients who may be denied for insurance coverage because of insufficient risk assessment. Figure 22 shows the wide range of survivorship of individual patients studied in Figure 21.

Figure 22. Sample Individual Breast Cancer Patient Survivorship (Age at Diagnosis: 40)

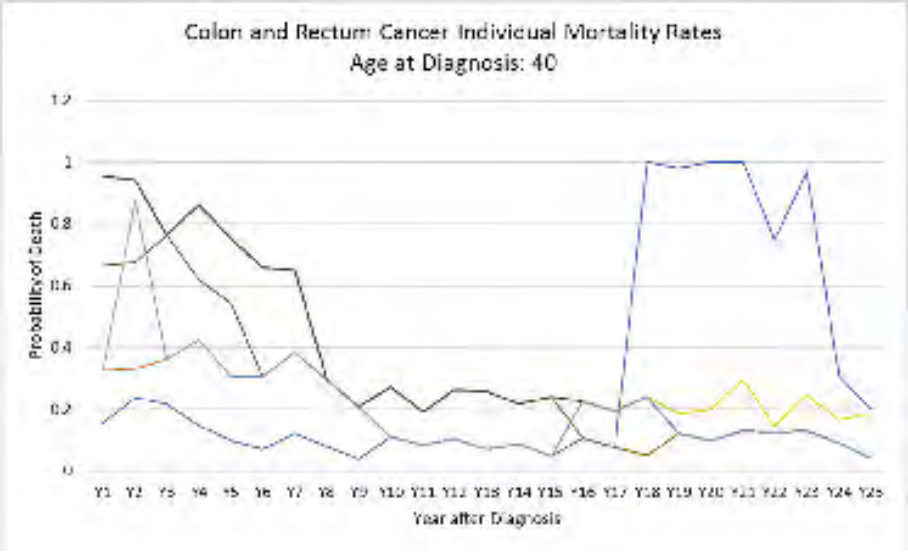


With the calibrated predictive models, it is also important that the predicted mortality rates of individual patients are consistent across time. The mortality rate is expected to be high for few years after diagnosis and then decreases to a relatively normal level increasing by age. If the mortality rate fluctuates significantly from year to year, even if the precision or recall rate of the model is high, the result is not meaningful for individual patients. In Figure 21, it is clear that the fluctuation of mortality rate prediction by years exists but with limited magnitude. This demonstrates the consistency of the CART models across the time.

Figure 23 illustrates some sample mortality rates of individual colon and rectum cancer female patients predicted by CART models. One patient (blue line) has nearly a mortality

rate of 100% at the 18th year from a mortality rate 7.7% in the previous year. The rate then stays at the very high level for about 6 years and drops sharply afterwards. The inconsistency of this individual patient's mortality rates invalidates the prediction results. For models with a small training dataset, the results could be driven by few data records and cause high uncertainty in the prediction results. In practice, a threshold needs be set to determine the usefulness of individual mortality rate prediction. When the fluctuation of predicted mortality rates is less than the threshold, the rates can be smoothed. Otherwise, the prediction result during the fluctuating period should not be used.

Figure 23. Sample Individual Colon and Rectum Cancer Patient Mortality Rates (Age at Diagnosis: 40)



Another approach to address the uncertainty of prediction is to provide a prediction interval of the mortality rates. However, for many predictive models, the prediction interval is not mathematically tractable. Nonparametric methods can be used to estimate the interval using the results of similar cancer patients. To estimate the prediction interval of a breast cancer patient's mortality rates, an extreme approach is to use the volatility of the mortality rate prediction of all patients diagnosed at the same age. However, many explanatory variables are used in the prediction and their power of explanation needs to be

considered. By finding the nearest neighbors, the variance of their estimates can be used to derive the prediction interval in a practical way. For example, if k nearest neighbors are found, the prediction interval can be constructed by the following approaches.

a. $(\hat{y} - t_{k-1}(\frac{1-CL}{2})s_k, \hat{y} + t_{k-1}(\frac{1+CL}{2})s_k)$

Where

$\hat{y}$: Predicted mortality rate.

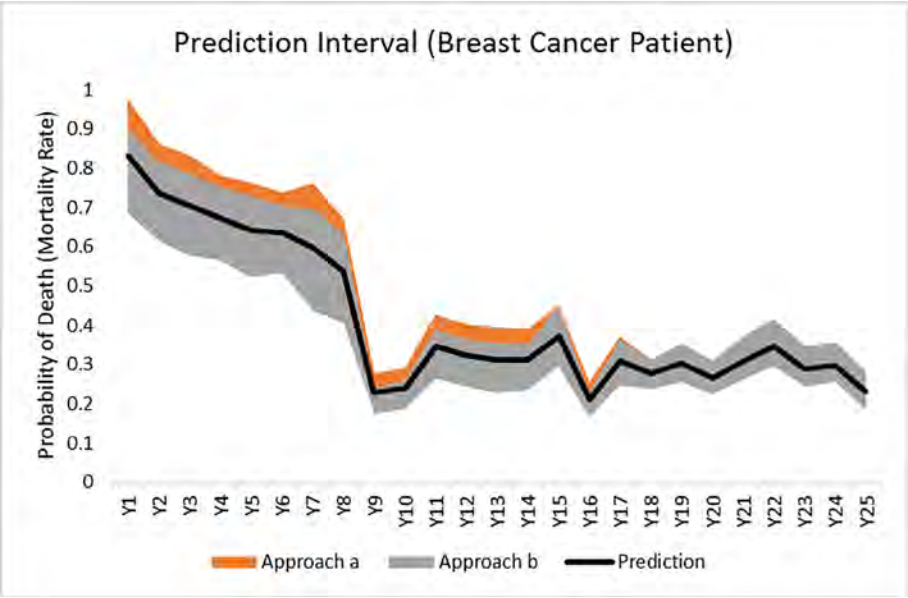
t_{k-1}:Student t value at the desired confidence level (CL) with a degree of freedom equals k-1.

s_k: Sample standard deviation $\sqrt{\frac{\sum_{i=1}^{k}(y_i-\bar{y})^2}{k-1}}$ with y_i as the prediction for neighbor i and $\bar{y}$ as the average prediction of the k neighbors.

b. $(\hat{y} - (\bar{y} - \left(\frac{1-CL}{2}\times 100\right) th\ percentile), \hat{y} + ((\left(\frac{1+CL}{2}\times 100\right) th\ percentile - \bar{y})$

The first approach assumes a Student t distribution and the second approach has no distribution assumption but uses empirical experience to determine the interval. The number of nearest neighbors to be sought can be set as an arbitrary number or determined by a maximum distance threshold. Setting k as a fixed value can guarantee the credibility of standard deviation estimate if k is large enough. However, it does not have any restriction on the degree of similarity between the data record and its k neighbors. Setting k by defining the maximum distance between the neighbors and the patient for prediction can guarantee a minimum level of similarity but may encounter insufficient neighbors if the maximum distance threshold is set too low. Figure 24 illustrates the mortality rate prediction interval of a breast cancer patient diagnosed at age 40. 100 nearest neighbors are chosen for the construction of confidence interval. The confidence level is 95%. The interval is quite large during the first 8 years.

Figure 24. Sample Prediction Interval of Breast Cancer Patient Mortality Rates (Age at Diagnosis: 40)



The prediction interval can be used to measure the robustness and credibility of individual mortality rate prediction. The bounds of the interval can also be used as stress scenarios to test the impact of prediction errors on insurance products.

5. ACTUARIAL IMPLICATION

With the mortality rate prediction for cancer patients, the cost of insurance can be calculated on an individual basis. For insurance companies, they will have a better understanding of the risk of the potential client and a better estimation of the fair insurance price. It will be easier for them to decide whether to accept the insurance application or not. Using the prediction interval, insurance companies can also calculate the amount of margin needed to be built in the insurance price. For cancer patients, the chance of getting an insurance coverage will be higher given the enhanced and personalized risk assessment.

Using 10-year term life (T10) and 20-year term life (T20) products, the single pay net premium is estimated for some sample breast cancer patients.

$$Net\ Premium = \sum_{i=1}^{T} \frac{m_i s_{i-1}}{(1+r)^{i-0.5}} \times Face\ Amount$$

Where

T: term of the life insurance product.

m_i: mortality rate in year i.

s_{i-1}: Survivorship at the beginning of year i.

r: Discount rate. A rate of 3% is used in this report. The deaths are assumed to happen in the middle of a year.

Figure 25 and Figure 26 illustrate the net premium per 10,000 face amount of some sample patients compared to the average insured using 2008 U.S. Valuation Basis Table (VBT 2008). The mortality rates are predicted by CART models. Both the average and minimum net premium of the sample cancer patients are much higher than the net premium for the average insured. For the least risky patients, the net premium is still less than 40% of the face amount. If the insurance product is used for estate planning, high net premium is less likely a major concern.

Figure 25. T10 Net Premium for Breast Cancer Patients

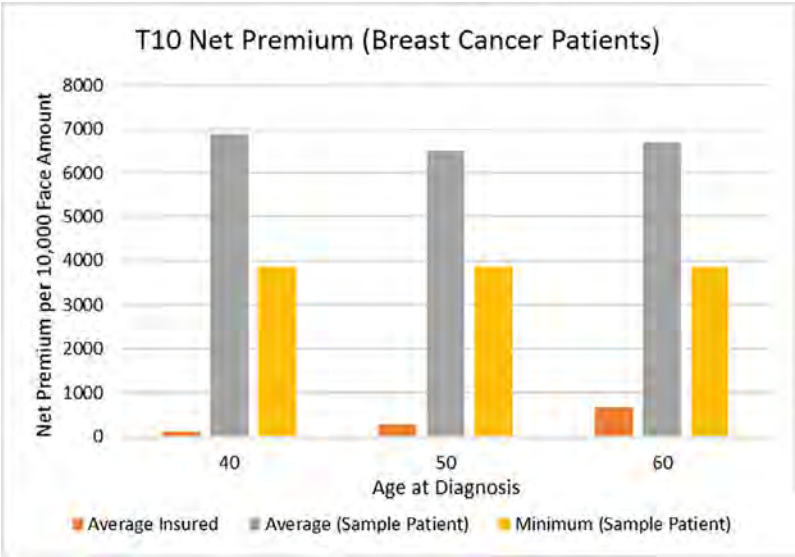
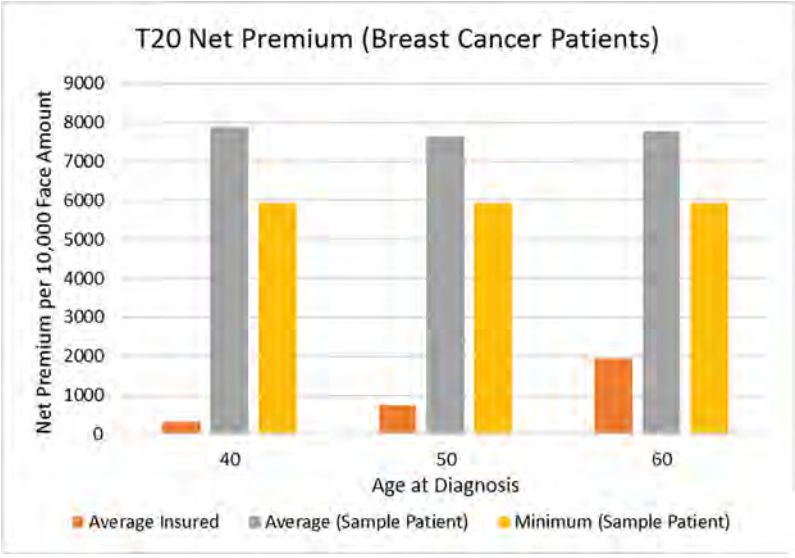


Figure 26. T20 Net Premium for Breast Cancer Patients



Since many deaths happened during the first few years after diagnosis, a possible solution to the high premium rate is to set a waiting period (WP) during which no death claims will be paid. The mortality risk can be significantly reduced and the net premium will decrease as well. It helps improve the affordability of insurance products for cancer patients. Figure 27 illustrates the impact of waiting period on T10 net premium. Using the sample patient data, both the average net premium and the minimum net premium are calculated using a waiting period of 1 year, 2 years, or 3 years. Applying a waiting period can significantly reduce the net premium. The impact of a waiting period on T20 net premium is similar, as shown in Figure 28.

Figure 27. T10 Net Premium for Breast Cancer Patients with Waiting Period

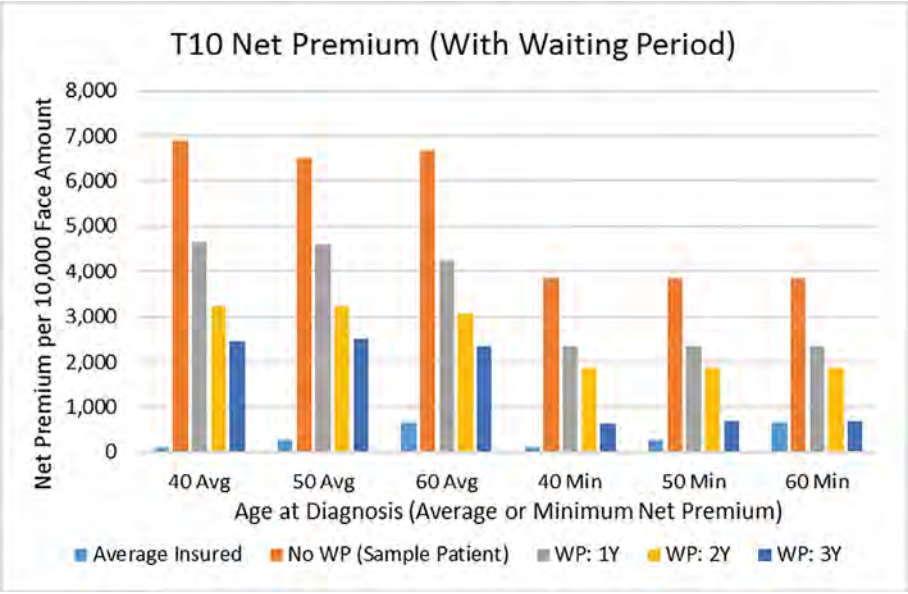
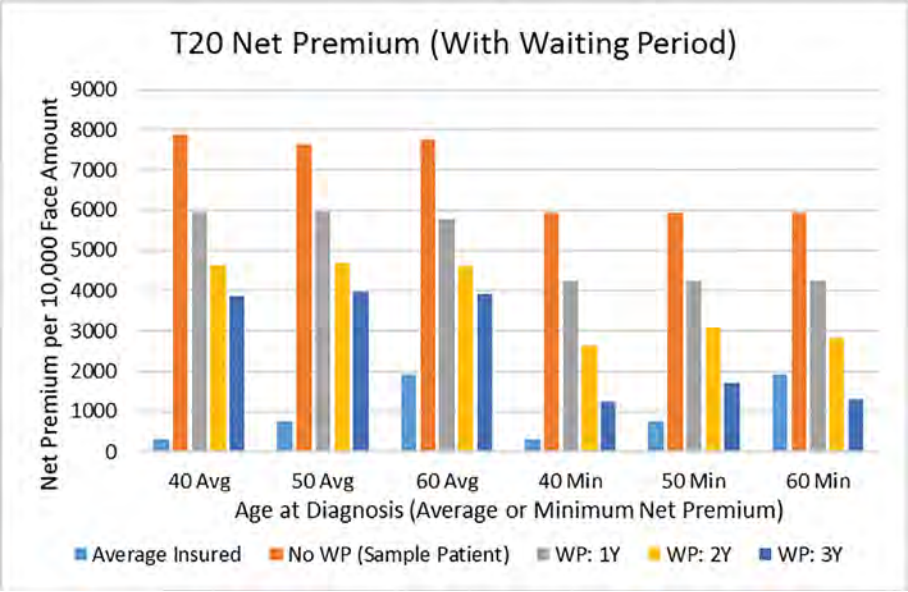


Figure 28. T20 Net Premium for Breast Cancer Patients with Waiting Period



Appendix A.4 lists the net premium of sample patients for female breast cancer. Similar analysis can be done for other cancer types as long as the model accuracy is acceptable and the risk is acceptable to insurance companies. Some cancer types such as other digestive cancer has a relatively low accuracy of model prediction. Further improvement of data and model may be needed before used for insurance pricing and risk assessment. Some cancer types such as respiratory cancer may have a very high mortality rate in the early years after diagnosis which leads to high insurance risk and a smaller number of survivors. Insurance companies may not be interested in underwriting patients having these cancer types.

With the waiting period mitigating the mortality risk taken by the insurance company, the prediction interval devised in Section 4 can be used to quantify the margin needed to be built in the net premium. Reserving and capital requirements can be calculated at different confidence levels as well. For example, the margin can be calculated using the upper bound of the prediction interval with a confidence level of 70%, deducted by the net premium using the predicted mortality rates. The reserve can be calculated using the upper

bound with a confidence level of 80%. The required capital can be calculated using the upper bound with a confidence level of 99.5%, deducted by the net premium. The examples of confidence level used here are arbitrarily but should be consistent with an insurer's business strategy and risk appetite in practice. With these methods, predictive modeling of individual cancer mortality rates can be fully integrated into an insurer's existing reserving and capital management framework.

Insurance companies can also design a combined product including both life insurance and annuity. Life insurance and annuity have a natural hedge of mortality risk. With the knowledge of individual mortality rates, pricing and risk management of annuity products are feasible as well. It could further improve the insurability of cancer patients and the affordability of insurance coverages.

6. FUTURE DEVELOPMENT

This study only uses part of the available data to predict the mortality rate of cancer patients. Other data such as genetic information and computed tomography scan of tumors can be used to improve the accuracy of prediction. Even though this type of information may not be accessible for all the patients, insurance applicants may voluntarily share it with an insurance company if it is necessary for underwriting and approving an insurance application. With the additional data, predictive models and model training methods can be enhanced to handle image data and a much larger data volume and image. Future studies on cancer mortality prediction and application may include the following areas:

1. Using image data to improve cancer mortality prediction. Techniques such deep learning and unsupervised pre-training will become more important in this area.

2. Using genetic information to improve cancer mortality prediction.

3. Incorporating financial information to assess the affordability of insurance products for cancer patients.

4. Product innovation that provides a more balanced insurance portfolio for cancer patients considering future medical cost, estate planning, mortality and longevity.

7. CONCLUSION

Using the SEER data, mortality rates of individual cancer patients can be predicted with a pretty high accuracy. Prediction interval of the mortality rates can be constructed as well to determine the appropriate margin in the insurance premium and calculate reserve and required capital. They are essential for assessing the mortality risk of cancer patients and setting appropriate insurance premium rates at the individual level. The feasibility of predicting individual cancer mortality rates helps improve the insurability of cancer patients and the affordability of insurance coverages.

In cancer mortality rate prediction, Random Forest models outperform other predictive models such as linear regression, logistic regression, CART, KNN and artificial neural network models in general. However, the improvement is immaterial in most cases. Unsupervised pre-training using principal component analysis or restricted Boltzmann machines does not help improve the model accuracy in this study. Information on age, surgery, lymph nodes, tumor primary site, tumor size and tumor extension explain most of the variation in the mortality rates.

8. REFERENCES

AHMED, KAWSAR, ABDULLAH-AL-EMRAN, TASNUBA JESMIN, ROUSHNEY FATIMA MUKTI, MD ZAMILUR RAHMAN, and FARZANA AHMED, "Early Detection of Lung Cancer Risk Using Data Mining." *Asian Pacific Journal of Cancer Prevention,* 14(1). (2013): 595-598.

CHEN, YEN-CHEN, WAN-CHI KE, AND HUNG-WEN CHIU, "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories." *Computers in Biology and Medicine,* V48, (2014): 1-7.

ESTEVA, ANDRE, BRETT KUPREL, and SEBASTIAN THRUN, "Deep Networks for Early Stage Skin Disease and Skin Cancer Classification." (2015): 1-8.
http://cs231n.stanford.edu/reports/esteva-paper.pdf

NATIONAL CANCER INSTITUTE, NIH, DHHS, "Cancer Trends Progress Report." (2015)
http://progressreport.cancer.gov

PARK, KANGHEE, AMNA ALI, DOKYOON KIM, YEOLWOO AN, MINKOO KIM, and HYUNJUNG SHIN, "Robust predictive model for evaluating breast cancer survivability." *Engineering Applications of Artificial Intelligence* (2013): 2194-2205.

RIEDMILLER, MARTIN and HEINRICH BRAUN, "Rprop - A Fast Adaptive Learning Algorithm." *Proceedings of the International Symposium on Computer and Information Science VII.* (1992)

WILLIAM E. BARLOW, EMILY WHITE, RACHEL BALLARD-BARBASH, PAMELA M. VACEK, LINDA TITUS-ERNSTOFF, PATRICIA A. CARNEY, JEFFREY A. TICE, DIANA S. M. BUIST, BERTA M. GELLER, ROBERT ROSENBERG, BONNIE C. YANKASKAS, and KARLA KERLIKOWSKE, "Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography." *Journal of the National Cancer Institute.* 98(17). (2006): 1204-1214.

APPENDICES

A.1 GENERIC MORTALITY TABLE

Generic mortality tables are constructed using the method described in Section 2.2. The table varies by tumor site and gender. Before using the tables, it is important to understand their limitation when applied to insurance pricing and risk assessment.

1. The data volume may not be big enough to draw a statistically credible conclusion. The incidence count for each age at diagnosis is listed in the second row of the tables. However, it is only for the first year after diagnosis. The incidence count decreases gradually for later years. For example, with a study period of 1973 to 2012, the cases diagnosed in 1973 can be used to analyze the mortality for 40 years, while the cases diagnosed in 2011 can only be used to analyze the mortality for less than 2 years. Therefore, when estimating the first year's mortality rate, 40 years' data can be used. When estimating the 40th year's mortality rate, only one year's data can be used.

2. Mortality rates for some diagnosis ages are not available due to data limitation. Ages with incidence count less than 100 are removed from the tables.

3. The Mortality rate could be zero in some cases due to the volatility brought by small data volume. The status of cancer survivors may also be outdated or cannot be confirmed at the last contact date. Smoothing and validation among ages are needed to avoid inconsistent mortality risk assessment. As the method of smoothing and validation varies by the usage of the mortality rates, the data presented here are not adjusted so that users can apply different adjustments as needed.

4. The data of the entire study period is used although mortality rate may change materially over period because of medical advancement. In practice, trend analysis is helpful for predicting future cancer mortality evolvement.

5. For cancer patients or survivors applying for a life insurance policy, the individual mortality rate could be very different from the aggregate rate here. Moral hazard and underwriting details can affect the aggregate mortality rate significantly as well and need to be taken into consideration before using the generic mortality tables for insurance pricing.

6. Quinquennial ages are used for easy presentation.

Table A.1 Cancer Mortality Table (Female, Site: Breast, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	609	2,459	6,871	16,536	25,816	32,579	32,117	33,314	33,981	28,856
Y1	39.2	30.6	22.1	22.1	18.8	21.8	26.3	30.5	37.0	45.1
Y2	63.6	64.8	54.6	38.8	29.8	30.3	34.9	34.0	38.7	40.8
Y3	64.3	72.9	48.8	35.9	31.5	30.3	34.2	35.8	41.1	46.1
Y4	47.5	54.5	46.6	37.3	25.5	25.9	30.4	30.6	36.9	44.8
Y5	38.0	37.7	38.2	30.7	24.5	21.7	28.7	34.5	35.5	43.3
Y6	34.1	39.9	30.9	31.0	22.2	25.0	28.6	30.4	36.5	45.0
Y7	43.2	33.0	28.7	25.1	20.0	21.7	25.4	29.8	38.7	47.3
Y8	32.3	31.7	22.5	21.9	21.2	19.7	24.1	28.1	34.1	45.6
Y9	35.7	34.4	19.1	21.3	16.9	17.6	22.5	27.0	38.1	50.9
Y10	14.3	38.0	23.1	20.2	17.4	19.6	25.0	29.5	42.2	52.9
Y11	23.2	20.4	18.4	18.9	19.4	19.2	26.0	32.9	37.5	59.6
Y12	17.7	8.1	19.9	18.0	17.1	15.2	24.1	28.8	38.5	57.4
Y13	29.1	12.3	26.6	19.7	21.5	18.3	24.4	30.8	44.8	56.2
Y14	22.0	5.9	15.2	15.1	16.6	15.3	22.4	31.6	44.5	68.3
Y15	11.6	18.6	10.9	14.0	18.3	18.0	26.3	34.3	50.0	77.2
Y16	12.1	15.4	13.2	13.8	14.8	17.6	26.0	35.2	50.0	76.9
Y17	13.3	16.6	18.6	8.8	12.4	17.8	25.2	34.7	54.1	84.4
Y18	14.2	5.0	17.5	14.3	15.8	20.4	22.6	34.4	57.2	91.2
Y19	48.9	24.2	12.0	14.0	14.9	22.7	24.9	39.5	60.1	89.5
Y20	18.5	2.9	14.2	12.9	15.9	23.9	27.1	35.7	56.3	110.5
Y21	38.9	12.8	15.7	14.3	17.4	18.8	29.9	36.9	63.0	101.6
Y22	0.0	41.2	18.5	18.7	14.1	23.5	29.8	38.6	76.0	110.6
Y23	22.5	15.1	13.3	22.2	15.2	24.7	35.0	41.4	73.7	104.6
Y24	24.5	0.0	8.1	13.0	21.3	25.4	35.4	56.8	70.2	102.9
Y25	0.0	4.5	12.3	23.6	21.2	23.2	35.6	50.4	81.9	107.3

Age	25	30	35	40	45	50	55	60	65	70
Y26	0.0	9.7	17.7	16.4	25.9	28.3	39.2	54.0	77.2	95.3
Y27	0.0	20.9	24.3	23.5	23.3	28.2	36.4	56.6	72.7	71.0
Y28	31.6	17.2	22.2	17.3	21.3	29.2	35.6	59.9	71.3	45.6
Y29	34.8	6.0	16.7	26.4	22.1	23.9	40.2	61.0	63.5	59.1
Y30	0.0	21.1	26.0	28.3	15.5	34.7	48.2	48.6	73.7	65.1
Y31	0.0	8.2	14.9	20.3	36.0	33.7	69.9	57.8	48.2	36.5
Y32	110.0	36.7	4.6	26.7	28.3	36.9	47.0	41.9	51.8	24.0
Y33	66.8	33.9	5.7	21.1	36.1	46.4	52.4	60.9	54.1	16.6
Y34	85.5	14.6	13.0	32.8	21.1	41.0	44.9	70.7	28.0	0.0
Y35	161.0	0.0	7.9	13.3	18.2	50.0	47.3	67.1	29.7	0.0
Y36	0.0	42.3	20.2	22.4	35.9	46.8	67.1	47.2	12.3	9.5
Y37	0.0	39.0	25.8	22.0	29.0	55.0	68.2	36.0	17.3	0.0
Y38	0.0	0.0	22.5	54.8	52.5	49.2	34.7	8.2	0.0	0.0
Y39	0.0	123.9	61.0	53.1	53.6	87.0	81.2	0.0	0.0	0.0
Y40	0.0	0.0	0.0	0.0	122.2	32.7	0.0	130.9	0.0	0.0

Table A.2 Cancer Mortality Table (Female, Site: Colon and Rectum, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	154	362	650	1,304	2,218	4,751	6,256	8,812	11,378	12,975
Y1	116.2	179.8	149.8	126.4	140.6	125.9	132.9	158.9	169.9	199.8
Y2	95.2	121.6	135.1	120.0	108.7	118.1	117.7	116.7	125.2	129.1
Y3	118.9	60.4	120.1	100.0	79.1	84.1	77.5	91.2	92.3	108.2
Y4	63.4	67.4	48.4	56.0	52.3	58.5	63.7	72.3	80.3	84.2
Y5	27.6	40.9	32.7	43.7	51.0	52.5	48.9	60.8	60.0	75.3
Y6	14.7	36.1	41.3	41.1	32.8	32.8	36.3	52.8	59.3	65.2
Y7	31.1	7.7	11.2	22.2	28.2	30.5	37.7	37.9	47.1	69.3
Y8	0.0	16.1	15.5	29.3	13.4	23.9	27.8	32.6	41.1	59.9
Y9	17.6	16.7	16.4	7.8	16.5	26.0	33.2	36.9	43.4	63.8
Y10	19.1	0.0	13.1	4.0	13.9	19.3	23.3	34.4	47.8	63.8
Y11	41.3	0.0	13.9	23.3	21.3	15.9	23.5	34.4	37.5	66.6
Y12	22.1	0.0	19.2	9.1	11.5	19.5	21.4	31.4	43.2	55.0
Y13	0.0	0.0	0.0	7.2	13.8	12.1	20.2	34.0	50.1	64.8
Y14	0.0	30.0	5.2	7.8	12.9	23.1	30.2	34.3	46.7	71.7
Y15	0.0	0.0	38.9	8.1	8.7	16.9	24.7	32.6	55.0	78.4
Y16	0.0	0.0	0.0	8.6	7.4	14.9	24.6	40.3	51.2	86.1
Y17	25.0	0.0	6.2	15.2	17.4	18.1	35.0	44.5	57.6	98.1
Y18	27.2	12.4	0.0	19.8	10.4	17.9	32.5	48.7	72.2	102.9

Age	25	30	35	40	45	50	55	60	65	70
Y19	0.0	13.6	6.8	7.4	22.2	15.3	23.3	39.5	69.0	106.5
Y20	0.0	0.0	14.7	27.5	19.3	18.8	34.6	38.3	74.6	92.1
Y21	0.0	44.3	7.9	12.8	15.7	18.9	30.2	52.3	80.9	127.3
Y22	0.0	0.0	0.0	27.4	25.8	35.9	35.5	59.1	73.4	110.3
Y23	0.0	0.0	0.0	5.2	15.4	37.6	29.2	57.3	92.1	116.4
Y24	37.1	37.2	20.3	28.4	10.3	20.8	40.8	63.8	105.3	130.0
Y25	0.0	0.0	11.6	6.6	36.6	38.9	47.1	75.6	87.0	91.5
Y26	0.0	0.0	0.0	13.9	12.3	34.3	49.3	72.7	113.1	120.3
Y27	0.0	25.1	13.9	16.2	13.3	33.1	57.0	45.1	99.8	78.9
Y28	0.0	0.0	0.0	0.0	5.1	36.8	53.8	94.8	97.6	57.5
Y29	0.0	0.0	35.7	9.8	10.7	19.2	53.8	54.4	97.3	60.7
Y30	0.0	0.0	0.0	21.1	46.9	24.7	57.7	53.1	88.7	52.1
Y31	57.2	73.7	43.7	0.0	19.3	19.3	62.1	82.3	105.7	45.2
Y32	0.0	0.0	24.9	13.6	30.6	58.1	66.8	48.7	87.6	27.0
Y33	0.0	0.0	29.6	48.7	53.7	48.2	77.3	102.5	66.0	23.1
Y34	0.0	0.0	35.8	39.5	42.4	32.1	53.9	40.6	66.3	0.0
Y35	0.0	74.5	0.0	79.6	69.7	61.3	99.8	64.3	21.8	21.8
Y36	0.0	0.0	0.0	63.1	37.0	28.8	49.6	67.0	12.1	0.0
Y37	0.0	0.0	0.0	58.7	0.0	83.7	107.2	38.4	15.9	0.0
Y38	0.0	0.0	0.0	0.0	33.5	40.3	149.8	70.8	0.0	0.0
Y39	0.0	0.0	0.0	0.0	78.4	94.0	73.4	29.2	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.3 Cancer Mortality Table (Female, Site: Other Digestive, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	109	240	491	978	1,788	3,179	4,648	5,912	7,076	7,869
Y1	337.0	322.6	385.0	416.6	444.4	514.0	542.3	572.5	603.5	630.5
Y2	229.9	199.7	194.7	212.7	227.3	282.8	299.1	314.7	357.5	358.0
Y3	61.6	109.6	66.0	117.4	121.4	143.4	171.5	196.7	193.6	216.2
Y4	140.3	58.3	55.9	76.7	67.1	84.1	115.3	115.8	159.5	134.7
Y5	28.0	25.7	41.6	50.4	56.1	65.4	80.4	79.6	106.3	95.0
Y6	59.2	26.9	15.3	36.1	58.3	48.7	53.4	69.5	89.8	83.4
Y7	0.0	14.5	48.1	39.4	39.0	32.6	48.4	53.5	72.8	66.0
Y8	0.0	0.0	17.5	38.4	28.6	33.3	50.3	51.8	60.1	84.7
Y9	34.4	31.4	9.0	15.9	43.3	29.7	22.9	26.3	62.9	72.5
Y10	0.0	33.2	9.6	45.5	20.2	35.2	22.7	38.7	45.7	72.5
Y11	0.0	17.7	20.5	18.8	25.0	21.9	33.5	51.8	33.6	56.2

Age	25	30	35	40	45	50	55	60	65	70
Y12	0.0	0.0	0.0	20.6	30.7	0.0	26.6	31.7	44.2	58.0
Y13	0.0	0.0	0.0	7.3	12.7	28.6	18.1	34.1	47.8	45.9
Y14	0.0	21.1	23.7	15.7	18.1	23.9	30.6	26.9	44.6	62.2
Y15	49.2	0.0	12.8	25.9	24.2	33.7	23.7	18.0	45.6	48.0
Y16	0.0	0.0	14.1	9.3	15.9	16.3	22.3	27.3	20.6	58.1
Y17	0.0	0.0	15.3	30.8	23.6	4.4	20.7	37.9	40.8	60.8
Y18	0.0	0.0	15.8	0.0	19.1	18.8	43.6	44.3	44.2	57.3
Y19	0.0	82.1	0.0	23.5	21.1	35.1	31.4	48.4	59.6	72.6
Y20	0.0	0.0	0.0	0.0	15.3	21.7	25.5	22.9	41.4	45.4
Y21	0.0	0.0	18.9	0.0	16.4	35.1	36.4	32.3	54.2	50.6
Y22	0.0	0.0	0.0	0.0	18.5	25.5	39.2	43.8	30.1	51.6
Y23	71.5	0.0	0.0	17.8	30.0	20.7	37.7	28.8	27.1	51.1
Y24	80.8	40.3	0.0	37.5	21.4	14.9	40.4	15.5	41.2	20.9
Y25	0.0	0.0	27.9	0.0	11.7	16.0	25.6	33.0	52.4	31.2
Y26	0.0	97.3	0.0	44.8	0.0	42.9	13.8	42.6	44.6	17.2
Y27	0.0	0.0	0.0	0.0	14.6	28.4	21.9	26.9	42.3	18.3
Y28	0.0	59.5	0.0	0.0	32.1	52.5	73.7	37.7	19.3	30.6
Y29	0.0	0.0	0.0	0.0	36.1	35.9	9.3	25.4	10.3	0.0
Y30	0.0	0.0	46.1	0.0	20.6	13.5	0.0	19.2	11.8	24.2
Y31	0.0	0.0	0.0	0.0	22.8	14.8	67.2	21.4	13.1	0.0
Y32	0.0	0.0	0.0	0.0	0.0	0.0	25.7	12.5	0.0	0.0
Y33	0.0	0.0	0.0	0.0	29.0	0.0	0.0	13.0	0.0	0.0
Y34	0.0	0.0	0.0	69.6	0.0	0.0	0.0	34.4	0.0	0.0
Y35	0.0	0.0	0.0	0.0	0.0	25.2	19.2	59.4	0.0	0.0
Y36	0.0	0.0	0.0	0.0	0.0	36.0	28.3	0.0	0.0	0.0
Y37	0.0	0.0	0.0	0.0	0.0	0.0	36.0	0.0	0.0	0.0
Y38	0.0	451.0	0.0	0.0	0.0	0.0	51.1	0.0	0.0	0.0
Y39	0.0	0.0	289.7	0.0	0.0	104.0	0.0	101.4	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.4 Cancer Mortality Table (Female, Site: Female Genital, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	7,743	9,414	8,622	8,525	9,292	11,055	13,154	14,074	13,393	11,023
Y1	11.0	12.6	22.3	41.2	56.2	71.6	83.0	100.0	127.4	160.9
Y2	5.2	9.7	19.7	40.3	47.0	64.9	69.0	78.9	95.5	121.3
Y3	3.8	6.5	12.3	24.4	37.0	46.3	49.8	61.3	71.4	90.1
Y4	2.2	3.9	9.2	14.6	25.1	33.6	37.5	44.3	54.6	69.7

Age	25	30	35	40	45	50	55	60	65	70
Y5	3.3	3.6	5.9	12.3	20.3	28.4	28.3	39.6	47.1	55.2
Y6	1.5	2.4	5.9	11.2	18.5	25.7	30.3	32.7	40.0	58.1
Y7	1.8	1.9	5.7	11.2	15.9	18.3	23.8	31.2	39.3	50.3
Y8	2.3	3.3	3.4	8.5	13.3	15.6	21.6	27.8	36.9	47.9
Y9	2.7	2.3	4.4	10.8	12.4	19.0	22.9	24.5	36.8	45.6
Y10	1.7	3.0	5.8	8.5	10.1	17.5	20.4	26.8	35.6	47.0
Y11	3.0	2.5	3.5	6.9	12.6	16.6	20.1	29.7	37.6	53.8
Y12	1.1	2.8	4.7	9.6	12.8	19.2	26.7	28.0	42.6	62.8
Y13	1.9	2.7	4.1	7.0	10.9	18.4	20.7	29.6	45.1	63.5
Y14	1.5	2.2	3.8	9.7	11.3	17.1	21.5	23.2	55.1	68.6
Y15	1.3	2.8	3.1	8.6	11.7	14.3	25.8	35.9	48.3	72.6
Y16	1.8	2.4	4.8	9.4	7.3	19.7	28.7	33.7	56.5	84.8
Y17	1.8	3.0	6.9	9.5	17.8	16.6	24.6	33.7	53.1	96.0
Y18	5.1	6.2	8.8	11.9	12.6	27.5	30.2	41.1	57.0	87.9
Y19	8.4	5.4	9.0	10.9	19.8	26.5	27.8	36.1	73.5	102.7
Y20	4.7	6.2	9.4	11.4	18.2	25.4	37.3	46.8	64.3	95.2
Y21	6.3	5.5	6.8	14.0	20.9	26.0	35.4	55.7	84.8	134.3
Y22	6.7	6.3	11.2	12.1	18.3	23.2	38.1	54.2	90.3	128.5
Y23	4.1	5.7	10.7	14.4	31.8	34.1	44.9	59.3	119.0	152.8
Y24	7.3	10.2	9.7	18.2	24.9	21.8	46.6	63.5	110.0	158.5
Y25	5.9	8.8	8.1	16.6	30.8	40.0	52.7	81.2	132.1	169.2
Y26	8.3	9.9	13.6	19.7	28.8	28.6	40.8	96.3	98.3	207.1
Y27	11.6	8.1	14.5	25.5	32.1	37.1	60.6	105.3	151.1	141.4
Y28	6.5	10.3	16.0	23.3	30.8	38.5	66.5	101.9	223.0	179.2
Y29	6.7	6.2	13.7	24.0	35.7	51.5	58.8	104.9	157.4	109.7
Y30	5.1	14.4	16.8	21.6	27.4	48.8	69.8	135.2	185.8	140.8
Y31	6.4	16.2	20.4	21.1	36.1	35.3	71.2	111.1	239.6	72.9
Y32	9.6	9.1	16.2	28.6	39.4	46.1	77.3	188.8	301.2	112.2
Y33	11.2	8.5	15.1	31.5	71.7	62.9	93.9	200.8	284.6	84.5
Y34	10.3	15.8	22.2	31.5	51.4	64.2	113.6	188.1	233.2	0.0
Y35	25.1	24.1	27.7	30.4	48.9	75.9	113.3	245.6	352.3	96.6
Y36	17.9	20.2	31.1	33.2	63.0	82.6	153.1	248.6	165.2	57.0
Y37	17.4	20.0	34.3	48.4	57.0	100.0	158.9	348.8	126.7	0.0
Y38	42.8	29.6	28.8	82.4	102.1	169.1	186.7	117.3	638.5	0.0
Y39	41.6	36.7	88.5	14.6	55.9	105.4	171.4	708.0	0.0	0.0
Y40	262.6	125.0	150.7	140.7	164.2	97.5	132.1	1000.0	0.0	0.0

Table A.5 Cancer Mortality Table (Female, Site: Lymphoma and Leukemia, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	1,298	1,447	1,716	2,038	2,562	3,376	4,355	5,436	6,533	7,226
Y1	144.0	194.2	194.3	221.1	189.0	181.4	201.3	215.4	263.1	296.0
Y2	76.1	79.8	100.1	78.1	97.6	105.2	102.2	116.9	138.1	147.5
Y3	37.9	57.0	49.5	54.5	61.9	73.7	75.8	99.0	106.1	120.4
Y4	33.4	34.5	47.2	34.4	53.9	58.3	61.5	72.4	98.7	111.3
Y5	23.1	33.3	32.3	34.1	51.4	58.0	62.3	87.1	89.3	107.4
Y6	20.4	25.9	27.5	36.3	42.8	49.8	59.4	66.9	92.9	107.4
Y7	12.1	27.2	20.5	30.5	39.1	50.3	63.7	65.4	93.1	101.2
Y8	11.4	13.6	21.6	22.7	21.2	52.2	51.4	73.9	90.2	98.9
Y9	13.4	14.2	18.9	27.6	31.3	39.0	46.8	63.7	65.2	102.6
Y10	9.3	8.8	28.3	25.8	34.9	39.7	53.7	59.4	79.7	82.3
Y11	11.4	12.2	16.5	21.0	34.0	38.0	60.5	64.4	69.2	101.3
Y12	11.9	9.5	14.2	19.5	32.9	42.9	49.7	54.4	64.9	83.4
Y13	8.8	5.0	13.3	23.5	32.7	43.4	42.1	56.8	69.1	101.5
Y14	16.7	8.7	20.9	20.5	29.0	32.0	55.0	59.3	66.2	94.3
Y15	11.5	12.5	14.7	22.2	27.8	48.2	41.8	51.8	64.2	77.2
Y16	7.9	15.0	13.8	18.2	33.8	25.3	50.1	53.1	68.4	85.4
Y17	6.3	9.9	10.4	13.7	29.1	33.1	45.1	47.3	55.8	94.2
Y18	8.7	8.3	18.2	18.6	32.0	32.2	41.1	54.9	69.6	82.7
Y19	13.9	15.6	17.2	22.8	25.8	49.1	28.3	47.1	46.2	66.5
Y20	14.8	11.8	15.8	12.4	11.2	22.7	34.5	49.9	66.6	64.9
Y21	10.5	12.8	23.5	10.9	37.0	35.6	31.2	54.8	68.8	64.9
Y22	22.3	8.6	29.4	36.9	21.3	41.0	50.7	47.8	75.6	67.4
Y23	27.3	31.0	15.0	30.9	28.1	39.4	48.9	48.5	79.9	49.2
Y24	6.7	21.3	12.5	24.1	41.0	40.0	42.3	51.7	34.8	23.5
Y25	21.8	15.4	18.9	26.9	20.6	22.8	44.0	51.8	46.8	62.1
Y26	20.5	22.4	5.4	5.0	40.9	36.1	34.4	29.2	44.5	24.5
Y27	9.1	44.4	36.9	18.3	20.9	23.0	56.4	48.8	45.2	7.0
Y28	4.9	16.7	13.6	6.7	39.8	25.5	24.5	28.1	40.4	0.0
Y29	10.4	24.4	23.3	23.5	26.5	27.8	28.1	41.8	12.7	0.0
Y30	18.3	20.9	58.5	36.6	10.2	8.0	37.8	48.1	15.2	9.1
Y31	20.9	32.3	24.2	11.7	34.4	27.8	13.9	51.1	25.4	10.8
Y32	16.3	37.0	28.9	39.4	12.6	51.8	16.5	44.9	0.0	0.0
Y33	36.8	32.3	67.6	46.6	45.9	52.8	27.8	20.4	11.0	0.0
Y34	67.4	27.6	19.7	37.7	17.8	61.6	11.7	11.6	14.9	0.0
Y35	40.4	34.1	25.5	24.5	20.1	98.9	26.4	16.2	0.0	0.0

Age	25	30	35	40	45	50	55	60	65	70
Y36	56.2	0.0	109.4	28.4	26.6	27.7	52.0	18.5	0.0	0.0
Y37	0.0	129.8	61.8	0.0	33.5	0.0	0.0	44.1	0.0	0.0
Y38	0.0	0.0	134.2	63.5	0.0	49.4	150.7	0.0	0.0	0.0
Y39	94.5	121.3	255.3	262.3	0.0	76.7	0.0	0.0	0.0	0.0
Y40	949.5	0.0	0.0	812.4	0.0	0.0	0.0	0.0	0.0	0.0

Table A.6 Cancer Mortality Table (Female, Site: Respiratory, Rate per Thousand)

Age	30	35	40	45	50	55	60	65	70
Incidence	128	362	1,058	2,404	5,067	8,280	11,901	14,953	15,633
Y1	301.5	447.6	463.6	479.4	472.5	498.7	514.2	533.2	562.0
Y2	107.7	283.6	316.0	321.3	329.2	329.1	329.8	330.5	343.9
Y3	64.7	106.3	163.1	169.8	175.0	183.6	178.1	185.2	221.9
Y4	89.0	60.4	93.3	98.2	96.7	109.3	122.6	138.8	150.9
Y5	20.0	22.0	43.9	67.2	68.6	70.6	95.7	104.8	136.5
Y6	20.4	23.0	46.6	52.1	61.9	74.0	88.4	98.8	115.5
Y7	0.0	23.9	39.4	39.8	46.9	60.6	81.0	99.7	110.1
Y8	21.5	12.7	15.7	37.7	50.7	64.6	73.2	90.1	104.3
Y9	23.3	52.5	54.1	56.4	49.6	60.2	67.8	89.1	115.8
Y10	0.0	28.2	17.5	49.7	26.9	47.4	71.3	90.1	104.3
Y11	0.0	30.7	49.3	57.4	36.8	61.8	64.3	79.6	101.9
Y12	0.0	16.1	13.4	28.1	36.0	78.7	82.2	96.2	112.2
Y13	26.1	33.5	35.6	29.8	38.5	43.7	61.7	62.0	103.7
Y14	53.5	0.0	22.6	12.0	35.7	75.5	82.4	89.1	111.5
Y15	28.6	0.0	24.1	33.7	63.9	52.5	65.1	86.1	109.1
Y16	0.0	0.0	42.6	31.6	54.2	66.2	64.8	93.2	91.0
Y17	0.0	58.8	47.2	43.6	57.0	58.0	78.4	101.5	96.2
Y18	0.0	0.0	10.7	63.5	37.6	66.2	85.1	102.7	104.4
Y19	0.0	0.0	22.5	47.1	65.5	80.9	63.9	58.9	97.6
Y20	0.0	0.0	47.0	32.1	59.6	99.4	82.6	87.9	66.8
Y21	34.5	26.7	63.9	48.3	43.0	59.2	83.7	67.6	42.5
Y22	36.8	29.8	44.4	15.5	55.5	62.6	78.9	74.2	67.4
Y23	0.0	0.0	16.2	51.0	38.1	62.6	89.7	130.4	71.3
Y24	38.9	103.7	34.5	46.5	41.7	92.3	72.4	68.8	76.5
Y25	0.0	0.0	0.0	40.1	51.9	72.4	79.2	64.2	29.7
Y26	0.0	0.0	41.3	79.2	102.3	58.8	40.7	74.1	0.0
Y27	0.0	0.0	0.0	54.8	105.0	65.9	69.4	58.9	45.6
Y28	64.3	48.8	25.7	61.0	53.4	61.2	54.3	75.5	10.2

Age	30	35	40	45	50	55	60	65	70
Y29	0.0	0.0	0.0	53.4	50.7	44.4	64.0	30.4	24.4
Y30	0.0	0.0	32.3	63.2	23.6	69.1	74.5	11.5	0.0
Y31	0.0	0.0	0.0	73.2	50.2	72.1	33.2	13.4	0.0
Y32	0.0	84.7	0.0	83.7	75.3	57.6	79.7	45.4	0.0
Y33	0.0	102.9	0.0	68.6	111.4	33.1	32.7	19.4	0.0
Y34	0.0	0.0	63.0	43.3	209.6	63.0	58.5	22.2	0.0
Y35	0.0	0.0	0.0	50.7	32.2	133.5	52.8	0.0	0.0
Y36	0.0	0.0	0.0	143.8	47.3	111.2	0.0	0.0	0.0
Y37	0.0	0.0	0.0	0.0	61.1	64.1	40.9	0.0	0.0
Y38	0.0	0.0	162.8	140.8	73.7	0.0	0.0	0.0	0.0
Y39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.7 Cancer Mortality Table (Female, Site: Urinary, Rate per Thousand)

Age	30	35	40	45	50	55	60	65	70
Incidence	143	312	635	1,090	1,825	2,727	3,796	4,802	5,292
Y1	58.6	81.1	67.7	95.6	106.6	117.2	127.4	160.5	156.1
Y2	19.8	44.1	47.5	54.8	64.9	78.2	66.7	83.4	103.5
Y3	30.9	19.3	37.1	22.2	37.6	54.1	51.7	63.3	76.3
Y4	32.2	10.1	19.1	23.3	32.7	34.2	44.2	46.0	75.5
Y5	47.1	15.6	8.6	21.8	28.1	33.8	47.7	55.0	74.2
Y6	24.9	5.4	9.0	26.6	19.3	29.4	43.9	52.0	62.7
Y7	13.1	32.9	18.4	25.1	30.1	33.9	45.7	53.4	74.5
Y8	13.6	28.9	6.5	15.2	33.5	30.0	42.3	44.6	72.4
Y9	13.8	0.0	13.9	22.5	21.0	38.6	40.7	50.4	64.2
Y10	0.0	6.6	7.3	8.7	16.9	36.9	44.1	58.5	71.7
Y11	30.3	6.8	19.3	16.2	18.0	31.8	48.9	59.8	84.2
Y12	0.0	20.8	16.4	12.3	17.7	31.1	49.8	54.2	96.5
Y13	0.0	7.2	13.2	32.3	25.3	30.1	34.5	69.1	87.0
Y14	33.1	7.7	14.2	23.7	15.2	33.8	62.3	66.7	99.2
Y15	0.0	16.1	19.9	25.8	25.3	43.1	45.1	89.3	92.0
Y16	0.0	0.0	10.5	28.1	27.5	25.7	39.9	89.5	108.0
Y17	0.0	9.4	23.7	3.8	28.3	34.8	55.2	65.0	110.6
Y18	19.5	10.2	6.5	28.3	28.7	41.5	65.3	82.8	103.5
Y19	19.9	10.6	6.9	8.8	36.7	30.0	79.3	116.6	149.8
Y20	0.0	22.1	15.2	23.6	25.8	60.9	71.0	75.5	132.2
Y21	24.1	0.0	23.6	42.3	38.2	61.6	73.4	110.8	117.1

Age	30	35	40	45	50	55	60	65	70
Y22	27.9	0.0	8.4	23.5	56.6	44.5	70.4	125.7	130.3
Y23	29.3	40.0	19.2	26.3	44.8	55.3	74.7	99.6	128.3
Y24	31.4	0.0	20.6	36.4	37.1	79.0	80.5	111.9	133.4
Y25	0.0	0.0	12.2	48.6	36.3	63.6	82.9	77.8	113.6
Y26	38.5	0.0	54.0	45.2	86.9	76.8	87.8	109.5	74.5
Y27	0.0	19.6	31.5	20.0	13.8	79.9	95.4	127.6	78.1
Y28	0.0	22.0	36.6	69.8	60.1	69.5	104.8	150.8	103.6
Y29	0.0	24.3	0.0	13.3	69.3	81.0	74.3	80.0	80.9
Y30	0.0	0.0	24.3	28.7	10.9	60.3	54.1	67.9	41.2
Y31	0.0	30.1	0.0	34.9	12.5	50.3	96.7	50.2	24.0
Y32	0.0	32.9	0.0	38.1	54.6	58.3	91.6	39.3	81.4
Y33	0.0	0.0	83.2	68.4	83.6	86.5	17.4	48.3	0.0
Y34	0.0	100.0	65.7	0.0	0.0	105.7	19.0	58.8	0.0
Y35	0.0	0.0	0.0	0.0	113.7	88.8	26.1	36.2	0.0
Y36	99.0	0.0	0.0	133.1	59.5	103.4	59.5	0.0	0.0
Y37	0.0	0.0	0.0	60.5	125.2	0.0	0.0	0.0	0.0
Y38	0.0	0.0	0.0	96.7	73.6	209.5	69.6	72.5	0.0
Y39	0.0	0.0	0.0	0.0	143.8	133.8	0.0	0.0	0.0
Y40	0.0	1000.0	0.0	0.0	0.0	301.0	0.0	0.0	0.0

Table A.8 Cancer Mortality Table (Female, Site: Others, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	3,325	4,986	6,282	7,517	8,964	10,765	12,560	13,606	14,453	13,706
Y1	58.6	78.3	86.7	95.0	104.4	127.8	168.5	215.7	250.7	302.6
Y2	53.0	54.9	58.2	61.9	68.2	79.8	90.4	106.0	104.7	118.2
Y3	23.7	34.9	33.4	36.8	39.2	41.9	48.3	61.0	68.1	81.7
Y4	17.0	22.2	23.7	24.3	26.4	31.6	32.7	48.1	56.4	66.2
Y5	18.0	18.4	17.8	19.2	21.4	27.3	31.2	38.7	44.4	57.9
Y6	18.5	15.6	15.5	19.0	22.0	21.0	30.9	38.9	44.3	54.5
Y7	10.9	12.6	12.4	16.4	18.3	22.7	27.2	34.0	43.2	53.3
Y8	6.6	9.6	12.5	8.1	16.3	20.6	29.3	31.2	46.4	62.2
Y9	8.2	11.7	10.5	13.4	15.4	18.7	26.8	27.7	38.1	61.2
Y10	12.3	7.0	10.1	10.6	15.3	21.5	22.8	31.9	44.7	58.0
Y11	11.1	11.1	10.9	10.3	13.1	21.6	25.4	34.1	40.6	50.3
Y12	12.9	9.0	6.1	8.7	14.4	18.3	21.0	31.3	44.0	58.8
Y13	2.2	8.5	7.5	9.9	9.3	18.8	24.5	25.1	39.8	60.6
Y14	8.3	8.4	9.8	10.9	13.6	17.9	28.3	29.3	40.0	59.4

Age	25	30	35	40	45	50	55	60	65	70
Y15	6.3	11.0	7.9	8.8	15.0	16.7	24.6	32.1	39.6	61.8
Y16	5.0	5.0	5.3	9.0	8.4	21.1	27.3	32.4	45.1	68.0
Y17	6.1	5.8	6.0	11.8	11.2	18.4	26.6	32.1	49.5	59.2
Y18	6.5	7.3	8.3	11.9	16.0	20.0	16.5	34.9	39.7	73.6
Y19	9.9	5.2	10.0	8.1	17.4	21.6	24.9	28.6	56.7	65.1
Y20	3.1	3.4	11.3	8.8	12.7	22.8	25.9	32.5	51.3	50.0
Y21	5.5	6.6	7.5	13.5	16.1	23.9	29.7	35.7	51.8	48.9
Y22	10.5	7.9	10.1	10.3	19.3	25.3	29.2	37.7	48.3	45.7
Y23	5.0	8.6	9.5	12.2	13.7	17.3	29.9	40.2	56.2	45.2
Y24	6.8	4.8	8.4	14.8	21.5	28.3	35.8	37.8	50.9	37.9
Y25	8.9	3.2	14.1	13.7	19.4	24.1	26.3	37.0	46.8	36.5
Y26	12.5	13.1	11.9	23.0	17.3	32.4	29.1	26.8	38.2	31.7
Y27	3.5	14.9	17.3	10.6	24.6	24.7	31.3	47.8	41.4	20.9
Y28	11.8	4.7	13.1	15.7	23.9	27.8	31.5	39.3	44.3	25.7
Y29	6.7	10.8	10.1	33.7	20.6	17.5	41.9	28.1	25.6	8.3
Y30	2.5	12.3	18.2	23.1	25.2	34.2	39.3	36.9	33.3	8.9
Y31	5.5	18.4	9.6	17.5	26.6	21.8	49.7	33.6	16.2	10.4
Y32	15.4	5.5	17.4	15.5	29.0	31.5	40.6	34.3	14.8	4.0
Y33	13.9	6.3	16.7	7.4	19.2	24.4	41.4	32.8	21.6	0.0
Y34	8.5	22.1	19.5	21.3	34.7	38.2	43.7	19.3	30.1	0.0
Y35	15.1	31.7	25.3	29.3	27.5	28.3	37.0	12.3	0.0	0.0
Y36	0.0	0.0	6.2	50.4	48.4	57.7	24.2	25.1	7.0	0.0
Y37	17.3	44.2	18.9	21.2	16.0	43.0	49.8	29.8	0.0	0.0
Y38	30.1	43.9	88.7	47.4	0.0	54.0	54.1	11.2	10.9	0.0
Y39	128.2	55.3	0.0	0.0	53.4	0.0	48.4	0.0	20.1	0.0
Y40	0.0	0.0	167.6	100.6	176.8	140.3	69.8	0.0	0.0	0.0

Table A.9 Cancer Mortality Table (Male, Site: Breast, Rate per Thousand)

Age	45	50	55	60	65	70
Incidence	165	188	198	211	209	218
Y1	24.8	31.1	26.5	77.5	53.4	55.7
Y2	76.3	42.8	43.5	100.0	90.0	85.0
Y3	36.7	95.0	62.5	87.7	88.5	79.8
Y4	76.2	55.6	78.8	32.2	80.2	86.9
Y5	51.5	65.4	46.1	86.5	43.7	68.8
Y6	21.7	55.9	41.4	80.5	52.6	68.6
Y7	33.3	37.0	50.4	71.6	48.8	49.3

Age	45	50	55	60	65	70
Y8	57.5	30.8	68.2	68.6	80.9	105.0
Y9	12.2	7.9	32.5	27.7	40.2	88.3
Y10	49.4	32.0	25.2	47.6	8.4	72.8
Y11	0.0	24.9	43.1	50.2	51.1	61.2
Y12	39.0	34.1	18.0	42.3	72.3	65.3
Y13	13.6	35.5	18.4	44.2	39.0	20.0
Y14	0.0	27.9	56.6	35.0	60.9	123.4
Y15	83.1	9.6	30.2	60.7	54.1	35.6
Y16	45.6	9.7	52.2	25.9	34.4	49.7
Y17	0.0	29.4	66.0	13.3	47.8	133.3
Y18	31.8	10.2	23.6	54.1	38.0	47.0
Y19	32.9	20.6	36.3	57.2	160.0	83.3
Y20	0.0	21.0	50.6	15.2	32.5	166.8
Y21	17.0	0.0	40.0	15.5	67.7	201.9
Y22	17.4	32.4	55.6	47.2	221.9	56.5
Y23	35.5	44.7	44.4	0.0	72.5	213.4
Y24	36.8	47.0	46.4	50.8	131.6	197.7
Y25	19.1	37.0	32.6	90.0	123.8	303.4
Y26	77.9	13.0	84.7	40.0	72.0	294.1
Y27	84.5	0.0	37.1	125.8	196.5	106.4
Y28	23.1	0.0	38.9	121.0	199.4	120.9
Y29	24.0	39.4	20.2	83.5	127.5	282.1
Y30	0.0	27.3	41.5	62.4	226.3	199.5
Y31	24.8	42.1	131.5	34.2	99.0	0.0
Y32	25.7	58.6	128.2	70.8	223.9	254.5
Y33	26.6	109.1	89.5	116.5	0.0	0.0
Y34	82.2	35.3	99.1	132.3	0.0	0.0
Y35	0.0	92.7	110.9	154.4	0.0	0.0
Y36	31.2	82.5	85.1	124.7	0.0	379.6
Y37	0.0	24.0	48.0	0.0	0.0	0.0
Y38	129.7	171.4	0.0	0.0	0.0	0.0
Y39	71.3	114.6	0.0	0.0	0.0	0.0
Y40	0.0	115.3	0.0	1000.0	0.0	0.0

Table A.10 Cancer Mortality Table (Male, Site: Colon and Rectum, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	159	386	671	1,372	2,343	4,925	6,492	9,112	11,804	13,504
Y1	114.2	177.2	150.8	124.9	135.8	125.4	132.3	157.9	169.3	198.1
Y2	93.5	120.8	131.1	120.3	107.1	117.6	117.7	115.5	125.9	129.3
Y3	116.4	65.8	116.5	101.7	80.8	83.2	77.6	89.9	91.9	107.2
Y4	61.9	62.1	48.5	59.0	51.6	58.0	63.8	72.3	80.1	83.3
Y5	26.9	37.3	30.8	42.2	52.2	52.0	48.9	60.5	60.1	75.1
Y6	14.3	32.9	38.8	39.6	32.7	32.3	35.6	51.9	60.0	65.7
Y7	30.2	7.0	13.9	21.3	26.4	29.0	36.2	37.8	47.2	69.2
Y8	0.0	14.5	14.5	28.0	18.8	22.6	27.7	32.6	41.1	59.7
Y9	17.0	15.0	15.3	7.4	16.6	26.5	31.8	37.4	43.5	64.1
Y10	18.5	0.0	16.4	5.7	13.0	20.9	25.5	33.6	46.8	64.4
Y11	39.8	0.0	13.0	22.2	19.8	15.0	22.9	34.3	38.2	66.8
Y12	21.2	0.0	17.9	8.7	10.7	18.3	20.9	32.2	42.5	55.7
Y13	0.0	8.6	0.0	9.1	12.8	11.4	20.8	34.5	49.7	65.9
Y14	0.0	27.2	9.9	7.5	12.0	22.7	29.0	35.1	46.0	70.9
Y15	0.0	0.0	36.9	7.8	8.1	16.1	24.4	32.1	54.3	79.2
Y16	0.0	0.0	0.0	8.3	6.9	15.3	24.9	39.8	50.2	85.4
Y17	24.0	0.0	5.9	14.6	16.4	19.6	33.9	45.0	58.7	97.2
Y18	26.4	11.6	0.0	19.2	9.9	18.5	32.3	48.4	71.2	105.4
Y19	0.0	12.7	6.5	7.3	21.3	16.1	23.5	38.6	68.5	107.5
Y20	0.0	0.0	14.3	27.2	18.8	18.4	34.0	38.2	74.5	91.6
Y21	0.0	43.0	7.7	12.6	15.4	18.6	29.8	52.2	80.0	128.8
Y22	0.0	0.0	0.0	27.2	25.5	35.6	35.2	58.7	72.8	111.3
Y23	0.0	0.0	0.0	5.2	15.3	37.4	29.0	56.9	92.4	116.7
Y24	36.5	36.3	20.1	28.3	10.1	20.6	40.5	63.4	104.6	130.4
Y25	0.0	0.0	11.4	6.6	36.2	38.7	46.8	75.0	86.5	91.9
Y26	0.0	0.0	0.0	13.9	12.2	34.1	48.9	72.1	112.3	120.8
Y27	0.0	24.5	13.8	16.1	13.1	32.9	56.5	44.7	98.9	79.3
Y28	0.0	0.0	0.0	0.0	5.0	36.6	53.3	93.9	96.7	57.8
Y29	0.0	0.0	35.3	9.8	10.6	19.1	53.3	53.8	96.3	61.0
Y30	0.0	35.1	0.0	21.0	46.4	24.5	57.1	52.5	87.7	52.5
Y31	56.2	75.3	43.2	0.0	19.1	19.2	61.4	81.4	104.4	45.5
Y32	0.0	0.0	24.6	13.6	30.2	57.7	66.0	48.1	86.4	27.2
Y33	0.0	0.0	29.3	48.5	53.0	47.8	76.4	101.2	65.0	23.3
Y34	0.0	0.0	35.4	39.4	41.8	31.8	53.2	40.0	65.3	0.0
Y35	0.0	76.1	0.0	79.3	68.7	60.7	98.6	70.7	21.5	21.9

Age	25	30	35	40	45	50	55	60	65	70
Y36	0.0	0.0	0.0	62.8	36.5	28.6	48.9	66.5	11.8	0.0
Y37	0.0	0.0	0.0	58.4	0.0	82.9	105.6	38.1	15.6	0.0
Y38	0.0	0.0	0.0	0.0	33.0	39.9	147.3	70.2	0.0	0.0
Y39	0.0	0.0	0.0	0.0	77.1	93.0	72.0	29.0	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.11. Cancer Mortality Table (Male, Site: Other Digestive, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	181	376	747	1,404	2,597	4,433	6,347	7,941	9,556	10,475
Y1	335.1	324.5	382.9	409.6	439.7	507.7	531.0	564.6	601.1	624.3
Y2	206.6	179.9	204.5	221.6	231.8	267.5	295.4	314.3	347.5	360.7
Y3	71.0	89.0	73.4	109.8	115.2	142.4	168.2	189.1	193.5	211.8
Y4	133.3	65.8	71.2	74.5	73.7	91.4	113.3	115.3	154.7	136.1
Y5	20.0	24.3	47.2	64.5	56.2	66.4	80.7	77.6	100.9	94.5
Y6	41.6	51.6	15.7	39.2	54.6	52.7	54.5	68.8	79.3	87.4
Y7	0.0	19.1	38.6	30.6	44.1	33.3	46.3	55.0	69.2	77.9
Y8	0.0	9.8	23.6	33.1	32.5	24.5	44.1	58.9	56.3	83.5
Y9	24.5	31.0	18.5	17.9	35.3	28.3	21.5	31.0	65.4	68.7
Y10	0.0	33.0	6.6	30.9	28.7	32.8	26.1	37.3	44.7	75.0
Y11	0.0	11.7	20.8	25.2	28.4	30.0	28.2	57.6	34.3	47.1
Y12	0.0	0.0	0.0	13.7	25.0	12.3	29.0	37.3	53.3	60.5
Y13	29.9	13.5	7.7	14.7	18.3	30.7	22.0	32.1	51.6	48.1
Y14	0.0	14.1	24.6	10.6	19.9	21.7	33.8	25.5	49.3	58.7
Y15	35.1	0.0	8.9	29.3	29.1	37.4	25.9	19.5	46.6	53.7
Y16	0.0	0.0	9.8	19.7	12.0	18.1	21.3	29.4	22.5	59.6
Y17	0.0	17.0	21.7	29.3	18.1	3.3	21.0	38.9	41.5	66.6
Y18	43.7	0.0	11.9	0.0	24.5	18.1	39.3	48.1	45.5	58.4
Y19	0.0	63.5	0.0	18.2	16.5	35.3	27.8	45.4	57.0	80.6
Y20	0.0	0.0	0.0	10.3	12.1	17.3	37.4	19.1	50.8	50.4
Y21	0.0	0.0	15.5	11.4	26.8	28.7	30.2	34.2	64.4	49.1
Y22	0.0	0.0	0.0	0.0	23.3	26.4	40.7	41.4	30.3	55.8
Y23	59.4	33.1	0.0	14.9	26.0	17.3	36.2	24.8	37.6	46.0
Y24	67.2	35.5	0.0	31.9	18.7	18.9	48.9	13.3	51.9	25.0
Y25	0.0	0.0	23.9	0.0	10.3	20.3	22.0	28.4	46.9	28.3
Y26	0.0	89.6	0.0	38.8	0.0	37.0	17.7	42.0	40.0	23.4
Y27	0.0	0.0	31.3	0.0	25.4	32.9	31.8	23.4	45.6	16.8
Y28	0.0	53.9	0.0	27.2	28.8	46.3	72.1	32.9	17.5	28.3

Age	25	30	35	40	45	50	55	60	65	70
Y29	0.0	0.0	0.0	0.0	32.6	31.5	16.8	22.4	18.9	0.0
Y30	0.0	0.0	44.1	0.0	18.7	11.9	0.0	17.1	11.0	22.7
Y31	0.0	0.0	0.0	0.0	21.2	13.2	61.6	19.2	12.4	0.0
Y32	0.0	0.0	0.0	0.0	0.0	0.0	23.6	11.4	0.0	0.0
Y33	0.0	0.0	0.0	0.0	27.0	0.0	0.0	12.0	0.0	0.0
Y34	0.0	0.0	0.0	69.3	0.0	0.0	0.0	31.8	0.0	0.0
Y35	0.0	0.0	0.0	0.0	0.0	23.1	17.8	55.0	0.0	0.0
Y36	0.0	0.0	0.0	0.0	0.0	33.0	26.4	0.0	0.0	0.0
Y37	0.0	0.0	0.0	0.0	0.0	0.0	34.2	0.0	0.0	0.0
Y38	0.0	448.5	0.0	0.0	0.0	0.0	48.7	0.0	0.0	0.0
Y39	0.0	0.0	310.1	0.0	0.0	100.6	0.0	96.6	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.12. Cancer Mortality Table (Male, Site: Male Genital, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	1,583	1,807	1,661	1,839	3,158	9,547	19,624	31,381	44,169	41,985
Y1	32.1	21.1	18.7	21.2	23.3	16.6	17.6	21.0	28.5	39.8
Y2	18.9	17.8	15.8	16.8	24.1	22.5	23.0	26.1	33.9	43.3
Y3	13.9	9.9	7.5	13.5	13.1	18.2	20.9	24.4	31.7	43.2
Y4	5.0	6.6	7.2	8.0	12.7	16.5	17.8	23.4	30.6	43.7
Y5	7.3	6.4	6.9	3.5	9.1	14.6	18.9	23.7	34.3	45.3
Y6	3.9	5.3	4.1	10.3	8.8	16.4	20.8	29.4	38.0	49.4
Y7	7.4	3.3	7.0	13.2	15.6	16.5	22.5	29.0	35.8	53.1
Y8	5.7	6.0	8.5	11.1	14.4	16.2	17.9	28.9	39.3	55.2
Y9	3.8	5.6	6.1	9.8	8.5	19.8	26.4	28.8	38.0	59.3
Y10	9.6	7.1	4.5	10.8	21.3	20.3	26.0	36.5	45.1	62.2
Y11	7.6	4.4	2.4	12.3	5.1	26.6	28.1	36.4	48.7	70.3
Y12	3.3	6.1	10.9	10.5	10.7	27.4	25.2	34.7	49.9	72.5
Y13	5.3	2.8	2.9	6.2	8.2	23.1	28.4	42.3	50.9	72.7
Y14	1.9	7.3	4.8	16.4	9.3	19.7	31.2	36.9	54.5	74.9
Y15	2.1	13.9	14.2	13.0	13.9	30.5	28.0	37.5	53.7	84.3
Y16	4.3	3.3	5.7	8.7	18.1	25.8	27.9	44.4	58.7	87.8
Y17	4.6	8.7	10.4	16.1	22.0	18.3	32.1	38.7	57.5	88.3
Y18	4.9	11.6	9.0	14.8	8.0	22.0	29.6	45.3	64.9	93.4
Y19	18.0	6.3	7.2	4.3	13.5	35.7	38.5	53.1	65.5	96.1
Y20	2.9	9.5	5.5	5.0	5.2	16.4	38.3	43.9	63.1	93.3
Y21	6.2	13.0	3.0	16.2	52.1	36.7	35.6	48.0	67.7	84.4

Y22	13.1	11.7	10.3	12.0	34.1	22.0	46.1	38.5	62.5	72.3
Y23	7.2	16.0	7.7	0.0	16.2	12.3	24.7	44.3	66.7	58.0
Y24	7.6	3.4	13.4	22.7	8.9	48.5	32.8	43.5	62.3	46.6
Y25	0.0	15.1	10.0	8.1	69.1	47.4	36.7	42.9	61.5	33.6
Y26	13.1	4.1	16.7	19.1	23.4	37.4	50.7	47.5	34.7	29.5
Y27	9.4	18.6	24.7	33.9	27.7	31.7	41.0	35.5	37.7	23.5
Y28	26.2	21.0	36.6	28.3	62.3	45.5	47.3	24.4	39.0	23.8
Y29	11.9	12.1	9.5	16.9	18.0	51.4	28.6	34.2	24.3	4.0
Y30	13.3	20.4	0.0	18.7	45.6	46.1	33.4	41.6	16.1	12.8
Y31	38.7	16.6	23.8	22.5	76.7	18.1	31.4	27.4	32.4	5.1
Y32	18.0	19.4	30.2	47.5	65.9	19.8	36.8	46.0	2.7	5.8
Y33	11.2	0.0	17.9	31.7	113.9	26.1	8.5	30.3	3.1	3.2
Y34	0.0	13.6	21.5	151.2	161.1	61.6	41.2	21.8	0.0	4.0
Y35	16.8	0.0	31.3	0.0	0.0	75.2	40.1	12.9	0.0	5.6
Y36	42.4	27.3	0.0	57.9	98.0	0.0	66.6	8.1	0.0	0.0
Y37	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y38	0.0	121.5	0.0	0.0	0.0	0.0	37.7	0.0	0.0	0.0
Y39	130.2	276.7	0.0	0.0	326.0	0.0	0.0	0.0	0.0	0.0
Y40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.13 Cancer Mortality Table (Male, Site: Lymphoma and Leukemia, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	1,616	1,826	2,194	2,653	3,279	4,346	5,574	6,705	7,988	8,783
Y1	139.6	187.9	201.3	214.6	183.2	178.5	192.0	210.9	257.6	288.1
Y2	72.1	77.3	101.4	81.7	93.5	104.9	95.5	112.4	133.1	142.9
Y3	38.3	54.6	44.7	54.7	59.7	66.6	72.7	92.0	106.0	116.1
Y4	31.7	34.4	42.5	31.9	53.2	55.3	57.6	68.7	93.7	110.4
Y5	19.7	32.2	29.7	31.5	47.9	56.0	60.4	83.1	84.3	103.9
Y6	19.8	25.7	23.5	35.8	40.9	45.3	57.1	61.0	91.3	108.1
Y7	13.2	28.1	21.9	29.5	34.0	49.2	59.5	61.1	86.7	96.7
Y8	9.3	12.0	21.2	23.0	22.3	49.2	51.9	75.3	85.5	99.3
Y9	13.5	11.4	16.0	26.4	33.1	35.2	48.6	61.0	62.0	95.7
Y10	10.3	10.6	24.6	26.4	32.4	35.5	53.2	57.1	74.5	77.7
Y11	9.5	13.6	14.3	19.9	29.3	40.4	53.9	61.2	65.4	99.2
Y12	12.8	7.8	11.2	17.9	28.8	37.4	46.6	52.7	63.4	76.3
Y13	7.5	5.5	11.9	23.7	34.6	38.1	40.6	48.8	68.3	97.1
Y14	15.7	10.1	19.5	21.9	30.7	33.0	51.2	54.7	63.2	83.6
Y15	13.2	12.1	13.4	20.0	26.8	47.4	46.8	47.4	57.9	72.6

Y16	8.6	13.0	11.4	16.9	34.5	20.9	43.4	50.5	65.6	81.1
Y17	7.4	8.7	10.6	11.8	26.6	30.7	47.7	44.3	53.6	86.1
Y18	7.7	7.3	15.6	18.1	33.3	30.7	38.6	48.6	63.9	79.0
Y19	14.4	13.9	25.7	22.4	26.8	47.8	33.1	45.0	48.3	60.6
Y20	15.4	12.6	16.4	11.2	9.8	26.2	32.5	51.5	58.8	64.7
Y21	9.4	11.5	21.0	9.9	38.0	33.4	33.6	60.3	62.4	58.2
Y22	20.0	7.8	26.2	39.3	18.8	38.6	51.5	48.1	68.2	57.4
Y23	27.0	28.0	13.5	31.4	24.8	34.7	43.7	47.2	74.4	45.1
Y24	6.0	22.6	11.2	22.1	36.1	38.4	37.6	45.5	35.8	19.8
Y25	19.5	14.0	16.9	24.7	22.6	19.9	41.9	48.2	40.4	52.1
Y26	21.8	20.6	4.9	4.6	41.0	39.9	30.4	28.8	42.0	20.4
Y27	12.3	45.2	33.3	16.8	18.4	24.8	53.5	46.0	38.5	17.0
Y28	8.8	15.4	18.6	6.1	35.3	27.5	21.6	28.6	34.2	6.4
Y29	9.5	34.0	21.0	21.7	31.1	24.2	24.6	36.5	16.3	0.0
Y30	32.8	26.4	61.5	42.0	18.5	6.9	33.2	41.7	12.8	7.7
Y31	19.2	30.5	21.8	10.9	30.9	23.8	18.6	44.1	21.5	9.2
Y32	30.2	35.1	25.8	37.0	22.5	44.2	14.5	45.8	0.0	0.0
Y33	34.9	30.4	60.0	59.0	41.3	55.5	24.4	17.5	9.3	0.0
Y34	64.9	26.0	17.4	35.8	15.9	52.4	10.3	10.0	12.6	0.0
Y35	38.9	32.1	22.4	23.2	18.3	83.6	23.4	13.9	0.0	0.0
Y36	54.4	0.0	97.5	26.9	24.6	23.4	46.3	16.0	0.0	0.0
Y37	0.0	122.2	101.7	0.0	30.8	0.0	0.0	38.3	0.0	0.0
Y38	0.0	0.0	132.0	61.5	0.0	43.5	136.6	0.0	0.0	0.0
Y39	94.9	117.8	250.5	258.3	0.0	68.8	0.0	0.0	0.0	0.0
Y40	954.4	0.0	0.0	795.8	0.0	0.0	0.0	0.0	0.0	0.0

Table A.14 Cancer Mortality Table (Male, Site: Respiratory, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	117	197	540	1,512	3,466	7,240	11,819	16,911	21,311	22,392
Y1	247.5	248.1	437.1	468.8	479.1	480.9	497.8	517.9	533.4	561.6
Y2	154.1	115.0	235.1	298.7	325.6	326.3	325.2	327.4	328.2	344.7
Y3	92.9	49.3	84.1	148.1	174.1	176.3	179.7	182.1	187.4	213.3
Y4	30.2	64.0	46.5	90.0	97.0	102.7	112.8	126.2	143.2	150.0
Y5	16.5	23.5	43.9	49.2	65.1	66.8	75.2	97.0	111.5	135.9
Y6	17.3	12.2	33.7	37.1	47.0	60.2	71.9	92.0	95.8	117.7
Y7	18.2	13.1	50.3	45.4	42.4	51.2	63.5	80.9	92.6	111.9
Y8	0.0	13.3	15.4	20.7	43.8	54.8	61.1	70.6	92.2	105.2
Y9	0.0	14.0	32.5	39.8	54.7	47.0	60.0	70.4	92.3	111.5

Age	25	30	35	40	45	50	55	60	65	70
Y10	58.5	14.3	17.1	15.3	47.2	35.2	49.2	70.6	86.7	108.2
Y11	21.0	0.0	18.4	32.3	49.1	37.8	61.0	61.2	80.7	98.2
Y12	0.0	0.0	9.6	21.6	26.8	35.7	67.4	80.8	95.0	100.4
Y13	0.0	31.5	19.9	32.0	36.4	36.9	50.2	68.9	62.7	104.0
Y14	0.0	32.8	31.8	19.4	19.7	40.0	70.7	83.3	81.4	107.3
Y15	26.0	18.0	23.1	21.1	39.3	56.1	57.9	75.2	81.6	104.1
Y16	27.4	0.0	0.0	39.6	29.8	48.5	65.8	72.1	88.9	104.9
Y17	0.0	0.0	37.5	31.7	39.4	53.4	52.5	87.8	105.4	93.5
Y18	30.2	0.0	14.2	21.4	59.5	47.3	72.6	86.2	92.2	96.9
Y19	0.0	0.0	0.0	15.5	53.6	57.0	77.2	69.8	66.8	101.1
Y20	0.0	0.0	0.0	40.8	30.0	61.8	87.9	85.3	76.6	71.4
Y21	38.5	23.9	17.9	44.5	54.9	49.4	61.9	84.0	65.0	54.7
Y22	0.0	25.4	20.4	40.6	12.5	48.1	58.9	75.9	72.5	56.7
Y23	0.0	0.0	21.6	22.3	54.5	37.9	54.1	94.8	121.4	68.1
Y24	50.1	26.8	95.0	24.3	52.8	54.4	90.6	74.2	53.0	69.8
Y25	59.5	32.0	0.0	0.0	66.8	51.8	67.0	93.7	85.4	28.9
Y26	0.0	0.0	0.0	57.9	86.8	94.0	72.1	44.1	82.5	6.5
Y27	0.0	0.0	31.2	0.0	71.1	91.5	77.2	70.8	52.6	50.2
Y28	0.0	42.6	35.2	36.6	67.2	50.4	66.4	71.0	72.8	16.6
Y29	0.0	0.0	0.0	0.0	48.2	49.1	42.4	70.4	24.1	19.6
Y30	0.0	0.0	0.0	68.8	57.6	47.3	70.8	84.0	9.1	0.0
Y31	0.0	0.0	0.0	0.0	87.2	52.4	77.1	40.7	10.6	0.0
Y32	0.0	64.3	62.0	62.7	77.3	76.9	46.9	73.8	35.8	0.0
Y33	0.0	0.0	72.8	0.0	62.4	95.1	26.7	45.5	15.2	0.0
Y34	0.0	0.0	0.0	51.8	39.2	196.7	68.9	72.8	17.4	0.0
Y35	0.0	0.0	0.0	0.0	45.8	26.7	108.6	51.0	0.0	0.0
Y36	0.0	0.0	0.0	0.0	133.6	40.6	88.3	29.3	0.0	0.0
Y37	0.0	0.0	0.0	0.0	0.0	52.7	52.3	41.6	0.0	0.0
Y38	0.0	0.0	0.0	143.4	134.7	63.7	0.0	0.0	0.0	0.0
Y39	1000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y40	1000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table A.15 Cancer Mortality Table (Male, Site: Urinary, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	184	336	765	1,487	2,591	4,263	6,422	8,679	10,856	12,008
Y1	101.9	62.1	70.8	75.3	91.2	103.7	116.1	121.1	152.2	152.4
Y2	38.0	39.0	38.9	44.1	51.7	59.7	70.4	70.0	82.7	99.0

Age	25	30	35	40	45	50	55	60	65	70
Y3	24.9	35.4	24.4	26.2	29.4	36.8	50.7	53.9	60.2	76.4
Y4	17.4	21.1	11.0	22.1	22.4	36.6	35.6	44.9	47.3	69.0
Y5	27.6	32.0	11.4	21.3	22.0	24.7	33.5	44.8	51.6	66.7
Y6	0.0	24.4	7.2	16.3	21.7	21.6	28.1	41.0	51.5	63.4
Y7	0.0	15.8	22.6	10.5	22.8	27.8	29.3	40.8	52.0	70.6
Y8	21.1	5.6	18.9	9.9	18.9	36.5	31.5	41.5	41.7	66.1
Y9	11.4	5.7	12.0	7.7	22.6	26.4	28.2	39.0	48.7	62.9
Y10	12.2	0.0	3.3	14.7	16.8	23.4	28.4	42.5	56.1	71.0
Y11	25.8	13.0	10.3	16.1	18.3	18.3	30.8	41.6	55.2	82.9
Y12	0.0	0.0	18.0	17.6	14.1	21.4	31.3	47.0	53.9	90.3
Y13	14.8	14.1	7.7	11.1	22.3	26.5	33.6	38.3	59.8	81.3
Y14	0.0	30.1	4.2	21.8	24.8	22.8	30.1	53.7	62.3	91.0
Y15	0.0	7.9	13.2	18.4	14.5	21.0	42.6	52.2	69.5	89.2
Y16	0.0	0.0	4.8	14.3	34.9	24.1	27.3	41.9	80.8	101.0
Y17	17.7	0.0	5.2	15.5	23.3	20.8	31.2	52.1	74.2	103.9
Y18	0.0	9.6	21.8	6.9	29.8	26.0	34.4	67.0	79.3	97.5
Y19	20.4	20.2	12.0	7.5	14.2	28.9	38.0	65.8	94.1	118.6
Y20	22.2	11.0	19.0	16.6	15.4	37.2	51.0	67.5	83.1	132.1
Y21	0.0	24.7	0.0	26.5	35.0	34.9	62.5	69.8	97.2	93.1
Y22	0.0	14.9	0.0	14.9	26.2	57.0	53.3	67.0	109.4	107.6
Y23	0.0	15.2	23.7	27.5	40.5	52.3	54.3	74.7	93.8	117.4
Y24	26.9	16.2	17.2	18.4	40.6	39.9	69.5	77.0	107.7	108.6
Y25	87.7	0.0	9.6	13.5	41.1	35.2	72.2	76.7	90.7	97.3
Y26	37.1	20.9	21.6	31.1	47.0	64.6	85.6	93.7	83.8	88.7
Y27	0.0	0.0	11.8	26.3	23.5	37.1	75.8	77.0	107.8	63.0
Y28	0.0	0.0	13.2	20.6	54.0	46.0	64.1	78.9	123.1	79.5
Y29	0.0	0.0	14.9	23.2	23.1	63.6	78.8	72.5	83.3	56.5
Y30	0.0	0.0	0.0	27.4	52.2	38.9	66.9	77.0	50.3	19.8
Y31	60.6	0.0	19.0	0.0	41.0	67.6	55.5	87.9	52.2	44.2
Y32	0.0	0.0	20.6	18.4	34.9	71.8	65.0	85.2	71.3	39.2
Y33	0.0	0.0	26.0	47.9	71.4	87.3	89.0	20.1	65.2	0.0
Y34	0.0	0.0	59.5	35.3	36.7	28.2	101.1	22.6	66.1	0.0
Y35	0.0	0.0	0.0	0.0	25.4	78.4	71.7	15.9	42.2	0.0
Y36	0.0	60.0	0.0	0.0	90.2	118.7	83.2	37.1	0.0	0.0
Y37	0.0	0.0	0.0	0.0	83.7	129.3	0.0	29.9	0.0	0.0
Y38	0.0	127.6	0.0	0.0	123.2	59.3	184.0	45.7	41.7	0.0
Y39	0.0	0.0	0.0	0.0	0.0	124.9	88.6	0.0	0.0	0.0
Y40	0.0	0.0	894.1	0.0	0.0	0.0	190.8	0.0	0.0	0.0

Table A.16 Cancer Mortality Table (Male, Site: Others, Rate per Thousand)

Age	25	30	35	40	45	50	55	60	65	70
Incidence	3,733	5,732	7,234	8,754	10,357	12,308	14,074	15,248	16,118	15,222
Y1	61.6	82.5	89.7	99.4	106.1	127.4	167.9	214.8	248.8	299.0
Y2	51.8	56.1	61.4	65.1	68.4	79.3	90.5	104.5	106.2	118.4
Y3	25.1	35.3	34.2	36.3	38.9	41.6	49.0	60.0	68.7	81.6
Y4	17.4	21.6	22.5	23.3	26.1	31.4	33.9	47.1	56.1	64.8
Y5	17.6	18.7	18.7	18.8	21.7	25.8	30.7	39.0	44.9	58.5
Y6	18.2	16.1	14.8	18.2	23.1	20.6	29.7	37.7	43.3	55.6
Y7	10.3	13.4	11.3	15.4	18.1	22.9	26.0	33.7	43.7	53.9
Y8	8.0	9.3	13.2	8.0	16.5	20.8	29.8	30.6	46.3	61.3
Y9	7.9	12.8	10.5	13.9	16.1	18.4	25.9	28.6	37.0	60.8
Y10	11.8	7.8	10.5	11.5	15.3	20.3	21.7	31.1	44.6	58.4
Y11	10.6	10.7	11.3	10.8	13.7	21.3	25.1	34.0	40.3	54.0
Y12	11.7	8.2	6.9	9.0	13.3	18.8	21.7	32.9	43.9	57.3
Y13	2.0	9.5	7.8	10.4	9.3	18.4	23.7	25.5	38.2	59.4
Y14	8.9	7.8	10.6	11.8	13.5	16.6	28.0	30.5	41.2	56.7
Y15	5.8	11.1	8.1	9.9	15.7	17.6	25.0	32.6	40.2	62.8
Y16	6.2	4.9	5.5	8.1	8.7	20.6	27.6	32.4	47.3	68.1
Y17	5.8	6.3	6.3	11.3	12.3	18.2	26.4	32.9	51.6	59.5
Y18	7.1	9.0	8.7	13.8	16.0	20.6	16.6	36.7	42.1	74.6
Y19	11.6	8.0	11.6	8.9	16.8	22.8	26.0	28.5	56.2	64.9
Y20	3.1	3.4	12.3	9.3	12.6	22.4	25.6	32.3	51.7	51.7
Y21	5.5	6.6	7.5	13.6	16.1	23.7	29.5	36.2	52.0	48.6
Y22	10.5	8.0	10.2	10.4	19.3	25.2	29.1	37.7	48.7	45.5
Y23	5.1	8.7	9.6	12.4	13.7	17.2	29.8	40.3	56.6	45.0
Y24	6.9	4.9	8.5	15.0	21.6	28.2	35.7	37.9	51.3	37.7
Y25	9.0	3.3	14.3	13.9	19.5	24.0	26.2	37.1	47.2	36.3
Y26	12.6	13.3	12.1	23.3	17.4	32.3	29.0	26.8	38.5	31.5
Y27	3.5	15.2	17.6	10.7	24.7	24.6	31.2	47.9	41.8	20.8
Y28	11.8	4.7	13.2	15.9	24.0	27.6	31.4	39.4	44.8	25.6
Y29	6.8	11.0	10.2	34.1	20.8	17.4	41.8	28.2	25.8	8.2
Y30	5.0	12.4	18.5	23.4	25.4	34.1	39.2	37.0	33.7	8.9
Y31	5.6	18.6	9.8	17.7	26.8	21.7	49.6	33.6	16.3	10.3
Y32	15.5	5.6	17.7	15.8	29.2	31.3	40.4	34.4	14.9	4.0
Y33	14.0	6.4	17.0	7.5	19.4	24.3	41.2	32.8	21.8	0.0
Y34	8.6	22.4	19.8	21.6	34.9	37.9	43.5	19.3	30.4	0.0
Y35	15.2	32.2	25.7	29.8	27.7	28.1	36.8	12.3	0.0	0.0

Age	25	30	35	40	45	50	55	60	65	70
Y36	0.0	0.0	6.3	51.1	48.8	57.4	24.1	25.2	7.1	0.0
Y37	17.5	44.9	19.2	21.5	16.1	42.7	49.5	29.8	0.0	0.0
Y38	30.4	44.6	90.2	48.1	0.0	53.7	53.9	11.2	11.1	0.0
Y39	129.4	56.2	0.0	0.0	53.8	0.0	48.2	0.0	20.3	0.0
Y40	0.0	0.0	170.7	102.2	178.2	139.3	69.5	0.0	0.0	0.0

A.2 MODEL PERFORMANCE SUMMARY

Table A.17 summarize the weighted model performance of predicting the cancer mortality rates of 40 years by linear regression, Logistic regression, CART, RF, ANN and KNN models.

Table A.17 Cancer Mortality Prediction Model Weighted Performance

Breast Cancer						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.842	0.802	0.845	0.850	0.834	0.828
Recall	0.915	0.851	0.903	0.906	0.895	0.928
F-Measure	0.877	0.821	0.873	0.877	0.863	0.875
Colon and Rectum						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.813	0.686	0.809	0.772	0.790	0.737
Recall	0.680	0.728	0.673	0.756	0.688	0.841
F-Measure	0.722	0.698	0.712	0.763	0.714	0.785
Other Digestive						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.776	0.417	0.718	0.787	0.759	0.653
Recall	0.463	0.525	0.523	0.506	0.490	0.734
F-Measure	0.540	0.456	0.589	0.584	0.560	0.690
Female Genital						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.856	0.785	0.853	0.861	0.840	0.838
Recall	0.910	0.769	0.904	0.909	0.906	0.926
F-Measure	0.882	0.773	0.878	0.884	0.871	0.880

Male Genital						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.842	0.822	0.840	0.848	0.826	0.824
Recall	0.900	0.832	0.891	0.897	0.893	0.921
F-Measure	0.870	0.824	0.865	0.872	0.858	0.869
Lymphoma of All Sites and Leukemia						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.828	0.755	0.802	0.743	0.792	0.743
Recall	0.678	0.875	0.694	0.876	0.700	0.868
F-Measure	0.706	0.810	0.720	0.804	0.724	0.801
Respiratory						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.816	0.777	0.804	0.780	0.760	0.767
Recall	0.788	0.865	0.801	0.854	0.771	0.852
F-Measure	0.799	0.818	0.801	0.814	0.764	0.806
Urinary						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.810	0.777	0.804	0.782	0.774	0.765
Recall	0.821	0.879	0.811	0.872	0.823	0.875
F-Measure	0.815	0.825	0.807	0.824	0.798	0.816
All Other Sites						
Model	LR	Logistic	CART	RF	KNN(5)	ANN(10,5)
Precision	0.864	0.838	0.858	0.826	0.844	0.831
Recall	0.827	0.927	0.819	0.921	0.850	0.921
F-Measure	0.837	0.880	0.831	0.871	0.845	0.874

A.3 DATA SAMPLING IMPACT ANALYSIS

Table A.18 summarize the sampled subset size used for predicting cancer patient mortality for each year after diagnosis by cancer type.

Table A.18 Sampled Subset Size for Cancer Mortality Prediction ('000)

Year	Breast	Colon and Rectum	Other Digestive	Female Genital	Male Genital	Lymphoma of All Sites and Leukemia	Respiratory	Urinary	All Other Sites
1	273	452	513	204	227	431	983	264	786
2	254	230	179	124	222	165	355	143	353
3	234	168	79	90	203	117	154	107	253
4	218	133	47	71	192	94	90	89	209
5	198	110	33	59	179	80	64	77	180
6	179	93	25	51	173	67	48	67	159
7	162	81	20	46	155	58	39	59	140
8	146	70	17	41	133	50	31	51	126
9	139	64	14	38	129	44	26	46	114
10	128	57	12	36	116	38	22	40	87
11	122	52	10	34	113	34	19	35	82
12	113	45	9	33	100	29	16	31	74
13	103	41	7	32	87	25	14	26	63
14	59	25	4	20	49	14	8	15	34
15	54	22	4	19	42	12	7	14	30
16	48	19	3	18	37	11	6	12	27
17	43	17	3	17	31	10	6	11	26
18	38	15	2	37	26	9	5	9	23
19	34	14	2	34	23	8	4	8	20
20	31	12	2	33	20	7	4	8	18
21	29	11	2	33	18	6	3	7	17
22	20	8	1	23	10	4	2	5	12
23	18	8	1	22	7	4	2	4	11
24	16	7	1	20	5	3	2	4	10
25	15	6	1	19	4	3	2	3	9
26-30	47	19	3	73	12	12	5	11	36
31-35	23	9	1	48	5	7	2	5	22
36-40	12	3	0.5	7	2	3	1	2	12

Table A.19 compares the impact of data sampling on the 25th year's cancer mortality rate prediction by cancer type.

Table A.19 Data Sampling Impact on Cancer Mortality Prediction Models

Breast Cancer						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.856	0.866	0.856	0.866	0.856	0.866
Logistic	0.814	0.819	0.814	0.819	0.814	0.819
CART	0.854	0.863	0.854	0.863	0.854	0.863
RF	0.852	0.863	0.852	0.863	0.852	0.863
KNN(5)	0.833	0.861	0.833	0.861	0.833	0.861
ANN(10,5)	0.814	0.790	0.917	0.933	0.863	0.856
Colon and Rectum Cancer						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.777	0.828	0.820	0.787	0.798	0.807
Logistic	0.775	0.787	0.501	0.691	0.609	0.736
CART	0.768	0.815	0.819	0.787	0.792	0.801
RF	0.783	0.835	0.801	0.799	0.792	0.817
KNN(5)	0.743	0.785	0.793	0.827	0.767	0.805
ANN(10,5)	0.709	0.799	0.899	0.799	0.793	0.799
Other Digestive						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.781	0.925	0.809	0.831	0.795	0.875
Logistic	0.707	0.865	0.922	0.763	0.801	0.811
CART	0.778	0.945	0.819	0.881	0.798	0.912
RF	0.816	0.963	0.832	0.881	0.824	0.920
KNN(5)	0.718	0.881	0.667	0.881	0.691	0.881
ANN(10,5)	0.663	0.929	0.883	0.881	0.757	0.904

Female Genital						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.869	0.891	0.955	0.949	0.910	0.919
Logistic	0.851	0.882	0.928	0.843	0.888	0.862
CART	0.873	0.897	0.949	0.948	0.910	0.922
RF	0.873	0.900	0.949	0.942	0.910	0.921
KNN(5)	0.867	0.883	0.935	0.933	0.900	0.908
ANN(10,5)	0.859	0.891	0.964	0.946	0.908	0.918
Male Genital						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.838	0.876	0.888	0.891	0.863	0.883
Logistic	0.826	0.892	0.822	0.858	0.824	0.875
CART	0.856	0.856	0.886	0.893	0.871	0.874
RF	0.855	0.889	0.894	0.918	0.874	0.903
KNN(5)	0.830	0.798	0.865	0.893	0.847	0.843
ANN(10,5)	0.811	0.852	0.943	0.910	0.872	0.880
Lymphoma of All Sites and Leukemia						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.835	0.867	0.952	0.910	0.889	0.888
Logistic	0.833	0.836	0.953	0.959	0.889	0.893
CART	0.834	0.874	0.942	0.913	0.885	0.893
RF	0.829	0.840	0.958	0.962	0.889	0.897
KNN(5)	0.822	0.817	0.895	0.904	0.857	0.859
ANN(10,5)	0.833	0.854	0.931	0.923	0.880	0.887
Respiratory						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.775	0.905	0.700	0.792	0.736	0.844
Logistic	0.707	0.925	0.855	0.896	0.774	0.910
CART	0.752	0.926	0.729	0.906	0.741	0.916
RF	0.723	0.944	0.841	0.885	0.777	0.914
KNN(5)	0.702	0.825	0.702	0.833	0.702	0.829
ANN(10,5)	0.638	0.788	0.921	0.813	0.754	0.800

Urinary						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.802	0.824	0.830	0.827	0.816	0.826
Logistic	0.772	0.778	0.892	0.910	0.828	0.839
CART	0.804	0.820	0.824	0.801	0.814	0.810
RF	0.787	0.815	0.875	0.859	0.829	0.836
KNN(5)	0.762	0.761	0.809	0.795	0.785	0.777
ANN(10,5)	0.745	0.768	0.878	0.878	0.806	0.819
All Other Sites						
	Precision		Recall		F-Measure	
Models	Total	Sampled	Total	Sampled	Total	Sampled
Linear	0.853	0.874	0.968	0.925	0.907	0.899
Logistic	0.857	0.869	0.965	0.924	0.908	0.895
CART	0.857	0.885	0.955	0.897	0.903	0.891
RF	0.851	0.879	0.972	0.930	0.908	0.904
KNN(5)	0.852	0.831	0.932	0.886	0.890	0.857
ANN(10,5)	0.846	0.870	0.969	0.916	0.903	0.893

A.4 TERM LIFE NET PREMIUM ESTIMATION

Table A.20 lists the average and minimum net premium results of sampled breast cancer patients by age at diagnosis, waiting period, and the term of the life insurance.

Table A.20 Sample Net Premium Calculation

Breast Cancer						
T10	Average			Min		
Age at Diagnosis	40	50	60	40	50	60
Average Insured	99	274	668	99	274	668
No WP (Sample Patient)	6,900	6,506	6,690	3,855	3,855	3,855
WP: 1Y	4,659	4,582	4,235	2,358	2,358	2,358
WP: 2Y	3,220	3,240	3,059	1,865	1,865	1,865
WP: 3Y	2,464	2,520	2,359	650	682	682

Breast Cancer						
T20	Average			Min		
Age at Diagnosis	40	50	60	40	50	60
Average Insured	301	755	1,935	301	755	1,935
No WP (Sample Patient)	7,896	7,651	7,775	5,933	5,933	5,933
WP: 1Y	5,934	5,956	5,763	4,245	4,245	4,245
WP: 2Y	4,621	4,690	4,597	2,662	3,093	2,834
WP: 3Y	3,865	3,969	3,907	1,224	1,717	1,285

COLECCIÓN "CUADERNOS DE LA FUNDACIÓN"

Para cualquier información sobre nuestras publicaciones consulte:
www.fundacionmapfre.org

222. Individual Cancer Mortality Prediction. 2017

221. Defensa jurídica y dolo del asegurado en el seguro de responsabilidad civil. 2017

220. El proceso precontractual en el contrato de seguro: nuevo marco jurídico. 2017

219. Teoría de cópulas. Introducción y aplicaciones a Solvencia II. 2017

218. Cualificación profesional del actuario. Estudio internacional comparado. 2016

217. El seguro de responsabilidad civil derivada de la navegación de buques. 2016

216. El impacto de las últimas reformas en materia de jubilación: envejecimiento activo, sostenibilidad financiera y planes de pensiones. 2016

215. Previsión complementaria empresarial: estudio comparado internacional 2016

214. Normas sobre protección de los derechos de los consumidores en el contrato de seguro en Chile. 2016

213. *Gamificación*: un modelo de fomento y gestión de comportamientos deseados en las relaciones entre individuos y organizaciones. 2015

212. Modelo de gestión integral para el sector atunero. 2015

211. Opções embutidas em planos *unit-linkeds* brasileiros: avaliação sob a medida de probabilidade real. 2015

210. El enfoque de Solvencia II para las pensiones ocupacionales españolas. 2015

209. El seguro privado de obras de arte. 2015

208. Definición y medición de la cultura aseguradora. Aplicación al caso español. 2015

207. Tipos de interés para valorar las provisiones técnicas de seguros. 2015

206. Teledetección aplicada a la elaboración de mapas de peligrosidad de granizo en tiempo real y mapas de daños en cultivos e infraestructuras. 2015

205. Current Topics on Risk Analysis: ICRA6 and Risk 2015 Conference. 2015

204. Determinantes do Premio de Default de (Res)seguradores. 2014

203. Generación de escenarios económicos para la medición de riesgos de mercado en Solvencia II a través de modelos de series temporales. 2014

202. Valoración de los inmuebles del patrimonio histórico y los riesgos sísmicos en el contrato de seguro: el caso de Lorca. 2014

201. Inteligencia computacional en la gestión del riesgo asegurador: opera-dores de agregación OWA en proceso de tarificación. 2014

200. El componente transfronterizo de las relaciones aseguradoras. 2014

199. El seguro basado en el uso (Usage Based Insurance). 2014

198. El seguro de decesos en la normativa aseguradora. Su encaje en Solvencia II. 2014

197. El seguro de responsabilidad civil en el arbitraje. 2014

196. La reputación corporativa en empresas aseguradoras: análisis y evaluación de factores explicativos. 2014

195. La acción directa del perjudicado en el ordenamiento jurídico comunitario. 2013

194. Investigaciones en Seguros y Gestión del Riesgo: RIESGO 2013

193. Viability of Patent Insurance in Spain. 2013

192. Viabilidad del seguro de patentes en España. 2013

191. Determinación de zonas homogéneas de riesgo para los rendimientos de distintos cultivos de la región pampeana en Argentina. 2013

190. Género y promoción en los sectores financiero y asegurador. 2013

189. An Introduction to Reinsurance. 2013

188. El control interno y la responsabilidad penal en la mediación de seguros privados. 2013

187. Una introducción al gobierno corporativo en la industria aseguradora en América Latina. 2013

186. Mortalidad de jóvenes en accidentes de tráfico. 2012

185. Las reclamaciones derivadas de accidentes de circulación por carretera transfronterizos. 2012

184. Efecto disuasorio del tipo de contrato sobre el fraude. 2012

183. Claves del Seguro Español: una aproximación a la Historia del Seguro en España. 2012

182. La responsabilidad civil del asegurador de asistencia sanitaria. 2012

181. Colaboración en el contrato de Reaseguro. 2012

180. Origen, situación actual y futuro del seguro de Protección Jurídica. 2012

179. Experiencias de microseguros en Colombia, Perú y Brasil. Modelo socio agente. 2012

178. El agente de seguros y su Responsabilidad Civil. 2012

177. Riesgo operacional en el marco de Solvencia II. 2012

176. Un siglo de seguros marítimos barceloneses en el comercio con América. (1770-1870). 2012

175. El seguro de Caución. 2012

174. La contabilidad de los corredores de seguros y los planes y fondos de pensiones. 2012

173. El seguro de Vida en América Latina. 2011

172. Gerencia de riesgos sostenibles y Responsabilidad Social Empresarial en la entidad aseguradora. 2011

171. Investigaciones en Seguros y Gestión del Riesgo. RIESGO 2011

170. Introdução ao Resseguro. 2011

169. La salud y su aseguramiento en Argentina, Chile, Colombia y España. 2011

168. Diferencias de sexo en conductas de riesgo y tasa de mortalidad diferencial entre hombres y mujeres. 2011

167. Movilización y rescate de los compromisos por pensiones garantizados mediante contrato de seguros. 2011

166. Embedded Value aplicado al ramo No Vida. 2011

165. Las sociedades cautivas de Reaseguro. 2011

164. Daños del amianto: litigación, aseguramiento de riesgos,y fondos de compensación. 2011

163. El riesgo de tipo de interés: experiencia española y Solvencia II. 2011

162. I Congreso sobre las Nuevas Tecnologías y sus repercusiones en el Seguro: Internet, Biotecnología y Nanotecnología. 2011

161. La incertidumbre bioactuarial en el riesgo de la longevidad. Reflexiones bioéticas. 2011

160. Actividad aseguradora y defensa de la competencia. La exención antitrust del sector asegurador. 2011

159. Estudio empírico sobre la tributación de los seguros de vida. 2010

158. Métodos estocásticos de estimación de las provisiones técnicas en el marco de Solvencia II. 2010

157. Introducción al Reaseguro. 2010

156. Encuentro Internacional sobre la Historia del Seguro. 2010

155. Los sistemas de salud en Latinoamérica y el papel del seguro privado. 2010

154. El Seguro de Crédito en Chile. 2010

153. El análisis financiero dinámico como herramienta para el desarrollo de modelos internos en el marco de Solvencia II. 2010

152. Características sociodemográficas de las personas con doble cobertura sanitaria. Un estudio empírico. 2010

151. Solidaridad impropia y seguro de Responsabilidad Civil. 2010

150. La prevención del blanqueo de capitales en las entidades aseguradoras, las gestoras y los corredores de seguros 2010

149. Fondos de aseguramiento agropecuario y rural: la experiencia mexicana en el mutualismo agropecuario y sus organizaciones superiores. 2010

148. Avaliaçao das Provisoes de Sinistro sob o Enfoque das Novas Regras de Solvência do Brasil. 2010

147. El principio de igualdad sexual en el Seguro de Salud: análisis actuarial de su impacto y alcance. 2010

146. Investigaciones históricas sobre el Seguro español. 2010

145. Perspectivas y análisis económico de la futura reforma del sistema español de valoración del daño corporal. 2009

144. Contabilidad y Análisis de Cuentas Anuales de Entidades Aseguradoras (Plan contable 24 de julio de 2008). 2009

143. Mudanças Climáticas e Análise de Risco da Indústria de Petróleo no Litoral Brasileiro. 2009

142. Bases técnicas dinámicas del Seguro de Dependencia en España. Una aproximación en campo discreto. 2009

141. Transferencia Alternativa de Riesgos en el Seguro de Vida: Titulización de Riesgos Aseguradores. 2009

140. Riesgo de negocio ante asegurados con múltiples contratos. 2009

139. Optimización económica del Reaseguro cedido: modelos de decisión. 2009

138. Inversiones en el Seguro de Vida en la actualidad y perspectivas de futuro. 2009

137. El Seguro de Vida en España. Factores que influyen en su progreso. 2009

136. Investigaciones en Seguros y Gestión de Riesgos. RIESGO 2009.

135. Análisis e interpretación de la gestión del fondo de maniobra en entidades aseguradoras de incendio y lucro cesante en grandes riesgos industriales. 2009

134. Gestión integral de Riesgos Corporativos como fuente de ventaja competitiva: cultura positiva del riesgo y reorganización estructural. 2009

133. La designación de la pareja de hecho como beneficiaria en los seguros de vida. 2009

132. Aproximación a la Responsabilidad Social de la empresa: reflexiones y propuesta de un modelo. 2009

131. La cobertura pública en el seguro de crédito a la exportación en España: cuestiones prácticas-jurídicas. 2009

130. La mediación en seguros privados: análisis de un complejo proceso de cambio legislativo. 2009

129. Temas relevantes del Derecho de Seguros contemporáneo. 2009

128. Cuestiones sobre la cláusula cut through. Transferencia y reconstrucción. 2008

127. La responsabilidad derivada de la utilización de organismos genéticamente modificados y la redistribución del riesgo a través del seguro. 2008

126. Ponencias de las Jornadas Internacionales sobre Catástrofes Naturales. 2008

125. La seguridad jurídica de las tecnologías de la información en el sector asegurador. 2008

124. Predicción de tablas de mortalidad dinámicas mediante un procedimiento bootstrap. 2008

123. Las compañías aseguradoras en los procesos penal y contencioso-administrativo. 2008

122. Factores de riesgo y cálculo de primas mediante técnicas de aprendizaje. 2008

121. La solicitud de seguro en la Ley 50/1980, de 8 de octubre, de Contrato de Seguro. 2008

120. Propuestas para un sistema de cobertura de enfermedades catastróficas en Argentina. 2008

119. Análisis del riesgo en seguros en el marco de Solvencia II: Técnicas estadísticas avanzadas Monte Carlo y Bootstrapping. 2008

118. Los planes de pensiones y los planes de previsión asegurados: su inclusión en el caudal hereditario. 2007

117. Evolução de resultados tecnicos e financieros no mercado segurador iberoamericano. 2007

116. Análisis de la Ley 26/2006 de Mediación de Seguros y Reaseguros Privados. 2007

115. Sistemas de cofinanciación de la dependencia: seguro privado frente a hipoteca inversa. 2007

114. El sector asegurador ante el cambio climático: riesgos y oportunidades. 2007

113. Responsabilidade social empresarial no mercado de seguros brasileiro influências culturais e implicações relacionais. 2007

112. Contabilidad y análisis de cuentas anuales de entidades aseguradoras. 2007

111. Fundamentos actuariales de primas y reservas de fianzas. 2007

110. El Fair Value de las provisiones técnicas de los seguros de Vida. 2007

109. El Seguro como instrumento de gestión de los M.E.R. (Materiales Específicados de Riesgo). 2006

108. Mercados de absorción de riesgos. 2006

107. La exteriorización de los compromisos por pensiones en la negociación colectiva. 2006

106. La utilización de datos médicos y genéticos en el ámbito de las compañías aseguradoras. 2006

105. Los seguros contra incendios forestales y su aplicación en Galicia. 2006

104. Fiscalidad del seguro en América Latina. 2006

103. Las NIC y su relación con el Plan Contable de Entidades Asegura-doras. 2006

102. Naturaleza jurídica del Seguro de Asistencia en Viaje. 2006

101. El Seguro de Automóviles en Iberoamérica. 2006

100. El nuevo perfil productivo y los seguros agropecuarios en Argentina. 2006

99. Modelos alternativos de transferencia y financiación de riesgos "ART": situación actual y perspectivas futuras. 2005

98. Disciplina de mercado en la industria de seguros en América Latina. 2005

97. Aplicación de métodos de inteligencia artificial para el análisis de la solvencia en entidades aseguradoras. 2005

96. El Sistema ABC-ABM: su aplicación en las entidades aseguradoras. 2005

95. Papel del docente universitario: ¿enseñar o ayudar a aprender? 2005

94. La renovación del Pacto de Toledo y la reforma del sistema de pensiones: ¿es suficiente el pacto político? 2005

92. Medición de la esperanza de vida residual según niveles de dependencia en España y costes de cuidados de larga duración. 2005

91. Problemática de la reforma de la Ley de Contrato de Seguro. 2005

90. Centros de atención telefónica del sector asegurador. 2005

89. Mercados aseguradores en el área mediterránea y cooperación para su desarrollo. 2005

88. Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación. 2004

87. Dependencia en el modelo individual, aplicación al riesgo de crédito. 2004

86. El margen de solvencia de las entidades aseguradoras en Iberoamérica. 2004

85. La matriz valor-fidelidad en el análisis de los asegurados en el ramo del automóvil. 2004

84. Estudio de la estructura de una cartera de pólizas y de la eficacia de un Bonus-Malus. 2004

83. La teoría del valor extremo: fundamentos y aplicación al seguro, ramo de responsabilidad civil autos. 2004

81. El Seguro de Dependencia: una visión general. 2004

80. Los planes y fondos de pensiones en el contexto europeo: la necesidad de una armonización. 2004

79. La actividad de las compañías aseguradoras de vida en el marco de la gestión integral de activos y pasivos. 2003

78. Nuevas perspectivas de la educación universitaria a distancia. 2003

77. El coste de los riesgos en la empresa española: 2001.

76. La incorporación de los sistemas privados de pensiones en las pequeñas y medianas empresas. 2003

75. Incidencia de la nueva Ley de Enjuiciamiento Civil en los procesos de responsabilidad civil derivada del uso de vehículos a motor. 2002

74. Estructuras de propiedad, organización y canales de distribución de las empresas aseguradoras en el mercado español. 2002

73. Financiación del capital-riesgo mediante el seguro. 2002

72. Análisis del proceso de exteriorización de los compromisos por pensiones. 2002

71. Gestión de activos y pasivos en la cartera de un fondo de pensiones. 2002

70. El cuadro de mando integral para las entidades aseguradoras. 2002

69. Provisiones para prestaciones a la luz del Reglamento de Ordenación y Supervisión de los Seguros Privados; métodos estadísticos de cálculo. 2002

68. Los seguros de crédito y de caución en Iberoamérica. 2001

67. Gestión directiva en la internacionalización de la empresa. 2001

65. Ética empresarial y globalización. 2001

64. Fundamentos técnicos de la regulación del margen de solvencia. 2001

63. Análisis de la repercusión fiscal del seguro de vida y los planes de pensiones. Instrumentos de previsión social individual y empresarial. 2001

62. Seguridad Social: temas generales y régimen de clases pasivas del Estado. 2001

61. Sistemas Bonus-Malus generalizados con inclusión de los costes de los siniestros. 2001

60. Análisis técnico y económico del conjunto de las empresas aseguradoras de la Unión Europea. 2001

59. Estudio sobre el euro y el seguro. 2000

58. Problemática contable de las operaciones de reaseguro. 2000

56. Análisis económico y estadístico de los factores determinantes de la demanda de los seguros privados en España. 2000

54. El corredor de reaseguros y su legislación específica en América y Europa. 2000

53. Habilidades directivas: estudio de sesgo de género en instrumentos de evaluación. 2000

52. La estructura financiera de las entidades de seguros, S.A. 2000

51. Seguridades y riesgos del joven en los grupos de edad. 2000

50. Mixturas de distribuciones: aplicación a las variables más relevantes que modelan la siniestralidad en la empresa aseguradora. 1999

49. Solvencia y estabilidad financiera en la empresa de seguros: metodología y evaluación empírica mediante análisis multivariante. 1999

48. Matemática Actuarial no vida con MapleV. 1999

47. El fraude en el Seguro de Automóvil: cómo detectarlo. 1999

46. Evolución y predicción de las tablas de mortalidad dinámicas para la población española. 1999

45. Los Impuestos en una economía global. 1999

42. La Responsabilidad Civil por contaminación del entorno y su aseguramiento. 1998

41. De Maastricht a Amsterdam: un paso más en la integración europea. 1998

39. Perspectiva histórica de los documentos estadístico-contables del órgano de control: aspectos jurídicos, formalización y explotación. 1997

38. Legislación y estadísticas del mercado de seguros en la comunidad iberoamericana. 1997

37. La responsabilidad civil por accidente de circulación. Puntual comparación de los derechos francés y español. 1997

36. Cláusulas limitativas de los derechos de los asegurados y cláusulas delimitadoras del riesgo cubierto: las cláusulas de limitación temporal de la cobertura en el Seguro de Responsabilidad Civil. 1997

35. El control de riesgos en fraudes informáticos. 1997

34. El coste de los riesgos en la empresa española: 1995

33. La función del derecho en la economía. 1997

32. Decisiones racionales en reaseguro. 1996

31. Tipos estratégicos, orientación al mercado y resultados económicos: análisis empírico del sector asegurador español. 1996

30. El tiempo del directivo. 1996

29. Ruina y Seguro de Responsabilidad Civil Decenal. 1996

28. La naturaleza jurídica del Seguro de Responsabilidad Civil. 1995

27. La calidad total como factor para elevar la cuota de mercado en empresas de seguros. 1995

26. El coste de los riesgos en la empresa española: 1993

25. El reaseguro financiero. 1995

24. El seguro: expresión de solidaridad desde la perspectiva del derecho. 1995

23. Análisis de la demanda del seguro sanitario privado. 1993

22. Rentabilidad y productividad de entidades aseguradoras. 1994

21. La nueva regulación de las provisiones técnicas en la Directiva de Cuentas de la C.E.E. 1994

20. El Reaseguro en los procesos de integración económica. 1994

19. Una teoría de la educación. 1994

18. El Seguro de Crédito a la exportación en los países de la OCDE (evaluación de los resultados de los aseguradores públicos). 1994

16. La legislación española de seguros y su adaptación a la normativa comunitaria. 1993

15. El coste de los riesgos en la empresa española: 1991

14. El Reaseguro de exceso de pérdidas 1993

12. Los seguros de salud y la sanidad privada. 1993

10. Desarrollo directivo: una inversión estratégica. 1992

9. Técnicas de trabajo intelectual. 1992

8. La implantación de un sistema de controlling estratégico en la empresa. 1992

7. Los seguros de responsabilidad civil y su obligatoriedad de aseguramiento. 1992

6. Elementos de dirección estratégica de la empresa. 1992

5. La distribución comercial del seguro: sus estrategias y riesgos. 1991

4. Los seguros en una Europa cambiante: 1990-95. 1991

2. Resultados de la encuesta sobre la formación superior para los profesionales de entidades aseguradoras (A.P.S.). 1991

1. Filosofía empresarial: selección de artículos y ejemplos prácticos. 1991

LIBROS

La Responsabilidad Civil en el ámbito de los ciberriesgos. 2017

Longevidad y envejecimiento en el tercer milenio. 2017

El Ahorro en perspectiva histórica. 2016

Lo bueno, si breve... Microrrelatos de Seguros. 2016

The risk of longevity and its practical application to Solvency II. 2015

Historia de FIDES –Federación Interamericana de Empresas de Seguros. 2015

El riesgo de longevidad y su aplicación práctica a Solvencia II. 2014

Historia del Seguro en España. 2014

Actas del III Congreso Internacional de Nuevas Tecnologías: sus repercusiones en el seguro: internet, biotecnología y nanotecnología: 12 y 13 de noviembre de 2012, Santiago de Chile. 2013

Emergencia y reconstrucción: el antes y después del terremoto y tsunami del 27F en Chile. 2012

Riesgo sistémico y actividad aseguradora. 2012

La historia del seguro en Chile (1810-2010). 2012

Modelo de proyección de carteras de seguros para el ramo de decesos. 2011

Desarrollo comercial del seguro colectivo de dependencia en España. 2010

La mediación de seguros en España: análisis de la Ley 26/2006, de Mediación de Seguros y Reaseguros Privados. 2010

Museo del Seguro. Catálogo. 2010

Diccionario MAPFRE de Seguros. 2008

Teoría de la credibilidad: desarrollo y aplicaciones en primas de seguros y riesgos operacionales. 2008

El seguro de caución: una aproximación práctica. 2007

El seguro de pensiones. 2007

Las cargas del acreedor en el seguro de responsabilidad civil. 2006

Diccionario bilingüe de expresiones y términos de seguros: inglés-español, español-inglés. 2006

El seguro de riesgos catastróficos: reaseguro tradicional y transferencia alternativa de riesgos. 2005

La liquidación administrativa de entidades aseguradoras.2005

INFORMES Y RANKINGS

Desde 1994 se publican anualmente estudios que presentan una panorámica concreta de los mercados aseguradores europeos, de España e Iberoamérica y que pueden consultarse en formato electrónico desde la página Web:
www.fundacionmapfre.org

Mercado español de seguros

Mercado asegurador latinoamericano

Ranking de grupos aseguradores europeos

Ranking de grupos aseguradores iberoamericanos

Los millennials y el seguro en España. 2016

Tendencias de crecimiento de los mercados aseguradores de América Latina para 2016

Estudio social sobre la jubilación: expectativas y experiencias. 2015

La percepción social del seguro en España 2014

Informe de predicción de la actividad aseguradora en España. 2014

La internacionalización de la empresa española: riesgos y oportunidades. 2014

El seguro en la sociedad y la economía españolas. 2013

Papel del seguro en el desarrollo sostenible. ICEA, 2013

Emprender en momentos de crisis: riesgos y factores de éxito. 2012

La percepción social del seguro en España. 2012

CENTRO DE DOCUMENTACIÓN

Todas nuestras publicaciones a tu alcance

Además del acceso gratuito a nuestro fondo documental especializado en:

- Seguros
- Gerencia de riesgos
- Prevención

Fundación MAPFRE

Centro de Documentación

www.fundacionmapfre.org/documentacion

Fundación **MAPFRE**

2022

Paseo de Recoletos, 23
28004 Madrid (España)
www.fundacionmapfre.org

P.V.P.: 15 €
ISBN 978-84-9844-665-4