

The Longevity Risk of Cancer Insurance

Hsin-Chung Wang¹ Jack C. Yue²

Abstract

In recent years, with the sustainable growth of the life expectancy, population aging becomes more apparent in Taiwan. Taiwan's population of ages 65 and over will exceed 20% within 10 years, before 2025. (Source: National Development Council - Population Projection on 2014). Ageing is a key risk factor for cancer in Taiwan, the number of deaths over 65 years increased from 61% to 70% at 1996 to 2015, and the proportion of deaths due to cancer (aged 65+) and death population (aged 65+) is from 21% (1996) rise to 25% (2015). According to World Cancer Report 2014 published by the World Health Organization shows that the incidence of cancer has increased from 12.7 million in 2008 to 14.1 million in 2012, an 11% increase. The total annual economic costs of cancer, not including long-term costs to families and caregivers, were approximate US\$ 1.16 trillion in 2010, which amounts to more than 2% of the total global gross domestic product. Cancer is the leading cause of death in Taiwan for more than 30 consecutive years. The loss ratio (i.e., the ratio of claim to premium) of long-term health products seems to increase with the policy year and many insurance companies have 100% or more loss ratios after 15 years. The mortality rate improvement of cancer patients complicates the pricing of cancer products, and the lack of availability of population data further limit the development of cancer mortality models. Many methods have been proposed to deal with longevity risk of annuity products, such as the longevity bond or natural hedging. Still, the longevity risk of health products is necessary to face in the future.

¹ Associate Professor, Department of Finance and Actuarial Science, Aletheia University, Taipei, Taiwan, Republic of China

² Professor, Department of Statistics, National Chengchi University, Taipei, Taiwan, Republic of China

In this study, we use the data from Taiwan National Health Insurance Research Database (NHIRD) to evaluate the challenge of designing cancer products. We use Gompertz's and Generalised Age-Period-Cohort (GAPC) stochastic mortality models to explore the trend of cancer incidence and mortality rates. We found the Lee-Carter and Age-Period-Cohort models have the smaller estimation errors, and the CBD and Gompertz models are also good alternatives for the elderly, based on the Taiwan cancer empirical analysis. Most of Taiwan's cancer insurance policies pay the insured when they have cancer and are still alive. The loss ratio of Taiwan's cancer products is expected to grow and this can be deemed as a form of longevity risk (reimbursement risk). Furthermore, we construct whole-life first diagnosis benefit cancer and annuity policies to demonstrate the pure premium differences of incidence/mortality increments /reduction influence and consider the natural hedging effects of the combination of insurance products for insurers.

Keywords: Longevity Risk, Stochastic Mortality Models, National Health Insurance Research Database, Loss Ratio, Natural Hedging

1. Preface

In recent years, with the sustainable growth of the life expectancy in our country, population aging becomes more apparent. Taiwan's population of ages 65 and over will exceed 20% within 10 years, before 2025. (Source: National Development Council - Population Projection on 2014). Rapid decrease in mortality is one of the causes of the population's aging in Taiwan. Figure 1 is the mortality logarithm by age bracket. Mortality rates are declining continuously in all age brackets and female mortality rates have been improved more obvious. However, the changes from all age brackets are very different. There are a significant improvement on both genders at juvenile ages (especially around 10-year-old) and around the age of 20. The upheaved mortality caused by accidents for males around the age of 20 is longer exist, but the mortality for male above the age of 30 is slightly reduced in the past decade and there is almost no difference from the age of 40 to the age of 50. The mortality improvement on the female above the age of 30 is quite a lot. However, the process of aging population will cause the change of population structure and the impact of health condition. Smoking, unhealthy diet lifestyle and the increase of obesity population will cause the change of the main cause of death in Taiwan.

Mortality improvement increased the life expectancy. The period life expectancy of Taiwan people increases about 0.2~0.3 years annually. Table 1 is the life expectancy excluded ten leading causes of death from 2009 to 2011. Cancer is the first specific cause of death and the greatest impact on life expectancy after remove it. Life expectancy will increase to 80.63 to male and 85.7 to female. Cancer is also the major cause of death in OECD advanced countries. It's also around 28% of the total annual death among the OECD countries in 2009. Whether our country or OECD countries, male standardized mortality ratios were all higher than female, while in Korea and Japan (also OECD countries), male mortality is twice that of women.

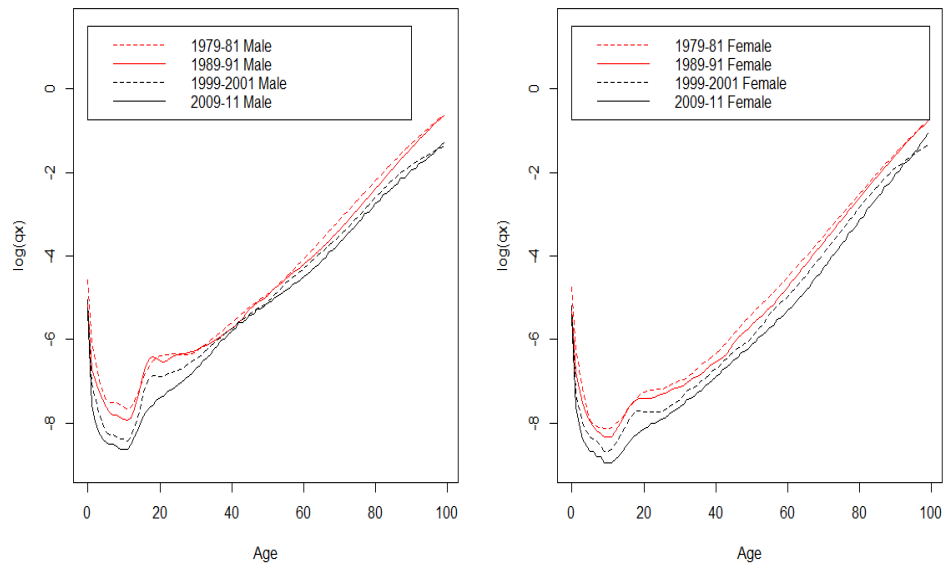


Figure 1. The Trend of Mortality Rates in Taiwan

Table 1. Life Expectancy Eliminated Specified Cause of Death
(Top 10 Causes in 1999-2001 and 2009-11)

	Male		Female	
	2009-11	1999-2001	2009-11	1999-2001
Cancer	4.17	3.93	2.88	2.68
Heart Disease	1.38	1.08	1.11	0.96
Accidents	1.08	1.69	0.73	0.80
Cardiovascular Disease	0.88	1.28	0.66	1.28
Pneumonia (Lung Disease)	0.71	0.34	0.63	0.24
Diabete	0.66	0.74	0.48	1.36
Chronic Liver Disease	0.64	0.65	0.34	0.35
Lower Respiratory Illness	0.53	—	0.33	—
Suicide	0.43	0.32	0.26	0.18
Chronic Kidney Disease	0.34	0.30	0.23	0.42
Life Expectancy	75.97	73.79	82.32	79.63

Note: Numbers in red are those with significant increases from 1999-2001 to 2009-11.

According to the American Cancer Society (ACS) report displays about 1,658,370 new cancer cases are expected to be diagnosed in 2015. In 2015, about 589,430 Americans are expected to die of cancer, or about 1,620 people per day. Cancer is the second most common cause of death in the US, exceeded only by heart disease, and accounts for nearly 1 of every 4 deaths. Taiwan cancer registry annual report 2015 (Source: Health Promotion Administration, Ministry of Health and Welfare) also shows that the new cancer patients were 96,694 in 2012 which means one person was diagnosed with cancer in every 5 minutes and 26 seconds. From the crude incidence data, 415 out of one hundred thousands were diagnosed with cancer. We expect the number of cancer patient will continue to increase in every year. This shows the health damage caused by cancer.

Exploring the trend of cancer incidence and mortality will provide great help to decrease prevention, treatment improvement, personal financial burden reduction, and the development of government's public health policies. Many articles were focus on specific cancer to build models include: Shek and Godolphin (1988) used Cox proportional hazards model for Breast Cancer Survival. Rosner and Colditz(1996) proposed nonlinear regression methods developed log-incidence model of breast cancer incidence. Robertson and Boyle (1997) adopted the age-period-cohort model to interpret the time trends in breast cancer incidence and mortality rates in Scotland. Moger et al. (2004) applied frailty compound-Poisson distribution to modelling of testicular cancer incidence using Scandinavian data. Shibuya et al. (2005) modelled and projected lung cancer mortality in 4 industrialized countries by Age-period-cohort model. Di Cesare and Murphy (2009) analyzed trends and forecasted mortality rates for three major causes of death (lung cancer, influenza-pneumonia-bronchitis, and motor vehicle accidents) by the Bayesian, Lee-Carter, Booth-Maindonald-Smith, and Age-Period-Cohort models, to assess how far different causes of death need different

forecasting methods. They found that when a clear cohort pattern is detectable, such as with lung cancer, the Age-Period-Cohort model shows the best outcome. When complete and reliable historical trends are available the Bayesian model does not produce better results than the other models. Uddin et al. (2010) utilized logistic regression method to model of the Incidence of Breast Cancer in NWFP, Pakistan. Hoggart et al. (2012) presented a stratified survival model for lung cancer. Tharu et al. (2015) proposed Bayesian Age-Period-Cohort Model of Lung Cancer Mortality. Pokhrel and Tsokos (2015) applied functional data analysis techniques to model age-specific brain cancer mortality trend and forecast entire age-specific functions using exponential smoothing state-space models. The articles discuss all cancers were: Pompei and Wilson (2001) found a Beta distribution fits SEER (Surveillance, Epidemiology, and End Results (SEER) Program, based within the Cancer Surveillance Research Program at the National Cancer Institute) age-specific cancer incidence. Kruijshaar et al. (2002) explored the usefulness of incidence–prevalence–mortality (IPM) models in improving estimates of disease epidemiology. Arbeev et al. (2005) discussed several mathematical models (a revised Strehler and Mildvan Model, Age-period-cohort models, A gamma-frailty model, The Armitage-Doll (AD) model, and so on) that address questions pertaining to the decline in human cancer incidence rates at oldest old ages.

One of the advantages of using these models is their ease of use in the mortality projection. In the current study, similar to dealing with longevity risk in annuity, we have considered the Generalised Age-Period-Cohort (GAPC) stochastic mortality models (Villegas et al., 2016), which include most of the frequently used mortality models from the past studies, and applied them for projections. Our aim was to verify

which GAPC model is suitable for describing the cancer incidence and mortality rates.

The current paper has been arranged as follows. Section 2 gives a brief introduction of Taiwan's National Health Insurance Research Databases (NHIRD) and describes the idea of handling big data, including the exploratory data analysis for cancer incidence and mortality rates. The introduction of proposed models and the evaluation of modeling cancer incidence and mortality rates and their applications are given in Section 3. In the final section, we provided a discussion and suggestions on dealing with the potential risks in cancer insurance.

2. Exploratory Data Analysis of the Big Data

Taiwan launched the NHI program on March 1, 1995, and about 99.68% Taiwan residents were enrolled in the program at the end of 2015. The data from the NHI program, including registration files and original claim data for reimbursement, were collected by the NHI Bureau. Based on the principle of privacy protection, personal identification numbers were de-identified by using the scrambling program twice. After the scrambling, the data were sent to the National Health Research Institutes (NHRI) for storage. Scholars from research institutes and universities can apply to the NHRI for the NHI data, by paying a fee based on the size of the data requested.

Cancer is one of the Catastrophic Illnesses (CI) recorded in the NHI database. The CI is one of the key features in Taiwan's NHI, and to ease some financial burden, the government provides some medical privilege and a co-payment waiver to people diagnosed with CI. In 2014, the CI patients were about 4% of Taiwan population (about 0.9 million) but about 27% of total NHI expenditure was spent. The cancer patients (about 0.45 million) spent approximately 35% of the total CI expenditures, which was around 65 billion NT dollars (or 2 billion US dollars).

In this study, we used all the records of the CI patients to avoid the possibility of biased selection for cancer data. The size of the CI related databases is huge, around 228 GB, and the data size of cancer is approximately half of the CI data, about half million Taiwan people with cancer in 2016. Datasets considered in this study include the registry for beneficiaries (ID), registry for CI patients (HV), and CI patients' original claim data extracted from the CD (ambulatory care expenditures by visits) data file (HV_CD). We used database software (SQL) and applied big data techniques for handling data analysis, since it is impossible to use a regular statistical software to perform a data analysis.

Considering the data quality, we used the data only during the period 2002-2011. The size and data quality complicate the analysis of big data, and the NHIRD showed no differences. The data codebook was used to check whether there was any problem in the data content. However, we can still find problems by cross-checking different databases. For example, the numbers of HV patients are different in the databases of HV and HV_CD (Table 2). The numbers of patients in HV_CD are close to (but slightly larger than) the official records from Ministry of Health and Welfare (MHW). It is possible that patients can have two or more than two records in the HV_CD database if they have more than one CIs. On average, about 5% of CI patients have more than one CIs. On the other hand, the numbers of patients in the HV database are too large, and it is likely that the HV records might not have been updated (e.g. CI patients passed away or recovered but are still in the list). We need to remove this kind of content errors before conducting any data analysis.

Table 2. Discrepancy between Databases

Year	HV		HV_CD	
	#Patients	#Records	#Patients	#Visits
2007	1,103,431	1,453,483	649,106	17,946,211
2008	1,164,465	1,529,866	678,544	19,173,919
2009	1,276,315	1,733,251	712,828	20,357,173
2010	1,350,786	1,863,254	746,746	21,619,442
2011	1,401,449	1,933,455	779,179	22,861,178

The process of analyzing big data is tedious and time-consuming. Therefore, we used the example of judging whether a CI patient is alive as a demonstration. There is a data column (“death note”) in the HV_CD database showing the status (i.e. alive or dead) of each patient. However, the number of death recorded in this data column is too small. For example, the number of cancer deaths is more than 40,000 annually but only 7,000~8,000 deaths (e.g. 7,517 deaths in 2007) shown in this column, meaning under-reported deaths from the death note. Therefore, to judge whether a patient is dead or not, certain criteria were needed to be set.

The cancer patients recorded in the HV and HV_CD database are those with malignant tumors between stage 1 and stage 4 who require medical treatments. Thus, we can use the nature of cancer patients’ medical visits to judge if they are still alive. We found that, if cancer patients stop visiting doctors, especially after a series of regular treatments, it is likely that health conditions of the patients are worsening, is similar to Lee et al. (2011) research results.

After a few trial-and-errors, we determined to use the condition “no outpatient records for two consecutive years” (Condition 1) for cancer patients to judge if a patient is dead, since the percentage of misjudgment was smaller than 5%. We used

the International Classification of Diseases (ICD) codes in the medical records to identify whether cancer patients were receiving medical treatments. In addition, we also included the condition, whether the cancer patients stop visiting doctors suddenly, or more precisely, whether there are 3 or more outpatient visits for the month of last outpatient visit (Condition 2). We used the official records of cancer deaths to verify the above two conditions.

In this part, the cancer incidence rates and cancer mortality rates derived from the analysis of NHIRD have been shown. The age-specific incidence cancer rates are shown in Figure 2. Since there were a large volume of data, we only selected the results of 2002, 2005, 2008, and 2011 for demonstration. As expected, men showed a higher incidence rate than women for all the age groups. To note, the incidence rate did not show big changes from 2002 to 2011 but only showed small increments between two consecutive years. However, still the increments did indicate potential longevity risk. Next, we applied stochastic models to the incidence rates.

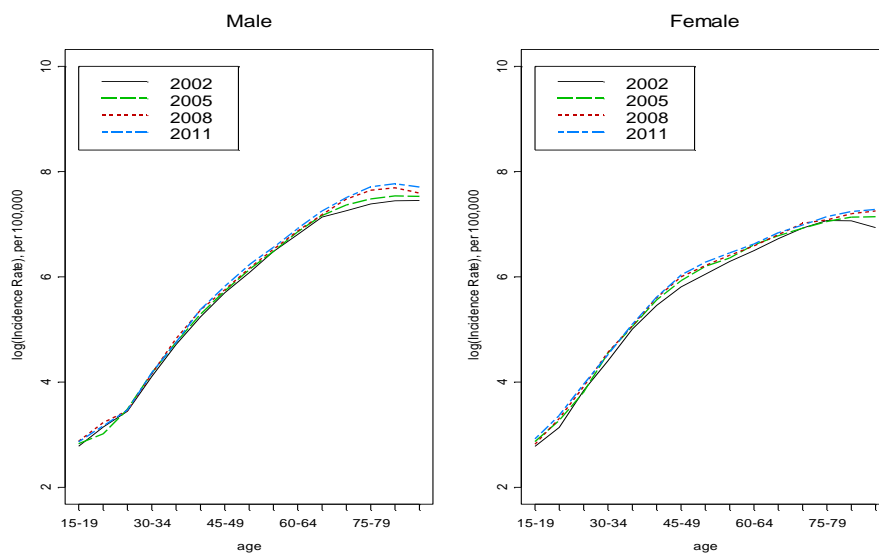


Figure 2. Taiwan age-specific cancer incidence rates (Estimated)

Since our goal is to design cancer insurance products, we computed the mortality rates for those who had cancer. These mortality rates were different from the official records where the cancer mortality rate is defined as the number of cancer deaths divided by all Taiwan residents. Moreover, since we used “no outpatient records for two consecutive years” (Condition 1) to determine if a patient is alive, we lost two years of data, i.e. could not determine the status of patients for 2010 and 2011. Thus, following the similar format as that in Figure 2, Figure 3 shows the age-specific mortality rates for cancer patients in 2002, 2005, 2008, and 2009. It seems that the mortality rate reduced slowly, and similar to those in Figure 2, the change in mortality rate between two consecutive years was not much.

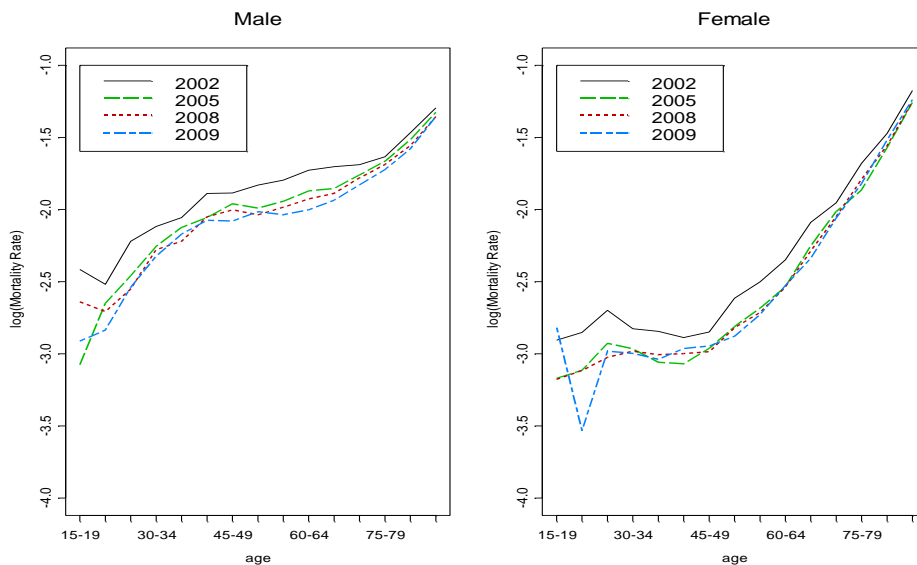


Figure 3. Taiwan age-specific cancer patients' mortality rates (Estimated)

Since the incidence rate increased slowly and the mortality rate reduce gradually, it is reasonable to expect that the number of Taiwan cancer patients (or the cancer prevalence rate) will increase. This is exactly the case in Taiwan and the number of Taiwan cancer patients increased about 30% from 2002 to 2011. Together with the population aging in Taiwan, the number of cancer patients will continue to increase in

the future. On the other hand, increasing incidence rate and decreasing mortality rate indicated a higher expenditure for cancer insurance, coinciding with the increasing loss ratio for cancer insurance in recent years. In the next section, we will detail the use stochastic models to explore the possibility of including longevity risk in cancer products.

3. Methodology and Application

Most of the stochastic models considered in the literature of mortality study can be categorized as the family of Generalised Age-Period-Cohort (GAPC) stochastic mortality models (Villegas et al., 2016). For instance, Age-Period-Cohort model (Cairns et al., 2009), the Lee-Carter (LC) model (Lee and Carter, 1992), the Renshaw-Haberman (RH) model (Renshaw and Haberman, 2003&2006), the Cairns-Blake-Dowd (CBD) model (Cairns et al., 2009), and Plat model (Plat, 2009) are some well-known examples.

First, an introduction to the Gompertz model and some of the GAPC stochastic mortality models will be provided. Originally, the Gompertz model was for modeling the mortality rates of higher ages. It is believed that the force of mortality μ_x at age x satisfies

$$\mu_x = BC^x, \quad (1)$$

where $B > 0$ and $C > 1$ are model parameters. Equation (1) can be converted to the form of central death rate m_x by $m_x = BC^x$. The Gompertz model has been applied to situations such as fertility and morbidity. Additionally, Strehler and Mildvan (1960) used the Gompertz mode to fit cancer mortality rate.

If m_{xt} denote the central death rate or incidence rate for a person aged x at time t . The LC model assumes that

$$\log(m_{xt}) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{x,t}, \quad (2)$$

with $\sum_x \beta_x^{(2)} = 1$ and $\sum_t \kappa_t^{(2)} = 0$, $\beta_x^{(i)}$ are age related parameters ($i = 1, 2$), and $\kappa_t^{(2)}$ represents the time related parameter. Note that $\beta_x^{(1)}$ is the general mortality level, $\beta_x^{(2)}$ is the decline in mortality at age x , and $\kappa_t^{(2)}$ is usually a linear function in time. The term $\varepsilon_{x,t}$ denotes the deviation of the model and is assumed to be white noise, with 0 mean and relatively small variance.

The residuals after fitting the LC model are often not random (Debón et al., 2008), and adding extra time or cohort component to the LC model is one of the possible modifications. The RH model can be treated as a version of LC model with an extra cohort component,

$$\ln(\mu_{xt}) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}, \quad (3)$$

where $\sum_x \beta_x^{(2)} = 1$, $\sum_t \kappa_t^{(2)} = 0$, $\sum_x \beta_x^{(3)} = 1$, $\sum_{x,t} \gamma_{t-x}^{(3)} = 0$, and the parameter $\beta_x^{(i)}$ denotes the average age-specific mortality, $\kappa_t^{(2)}$ represents the general mortality level, and $\gamma_{t-x}^{(3)}$ reflects the cohort-related effect.

The CBD model was designed to model mortality rates of higher ages and to deal with the longevity risk in pensions and annuities. For the CBD model, it assumes that the mortality rates satisfy

$$\text{logit}(m_{xt}) = \log \frac{m_{xt}}{1 - m_{xt}} = \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}, \quad (4)$$

where the parameters are $\beta_x^{(i)}$ and $\kappa_t^{(i)}$ ($i=1, 2$) denote the average age-specific mortality and the general mortality levels. If we assume $\beta_x^{(1)}=1$ and $\beta_x^{(2)}=x-\bar{x}$, then the model has a simple parametric form:

$$\text{logit}(m_{xt}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}). \quad (5)$$

The Age-Period-Cohort (APC) model is a popular tool for modelling disease incidence and mortality in epidemiology. Heuristically speaking, if we consider the notion of Analysis of Variance, the LC model considers the effects of Age and Age \times Period (Interaction), while the APC model considers three main effects, Age, Period, and Cohort:

$$\ln(m_{xt}) = \alpha_x + \kappa_t + \gamma_{t-x}, \quad (6)$$

where $\sum_{c=t-x} \gamma_c = 0$, $\sum_c c\gamma_c = 0$.

Three criteria are used to evaluate the models: mean absolute percentage error (MAPE), Akaike Information Criteria (AIC), and Bayesian Information Criterion (BIC). The MAPE is defined as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\varepsilon_i|}{Y_i} \times 100\%,$$

where Y_i and ε_i are the observed value and residual of observation i , $i=1, 2, \dots, n$.

The AIC and BIC are defined as

$$AIC = -2\log(L) + 2k,$$

$$BIC = -2\log(L) + k \log(n),$$

where L is the likelihood of the data, k is the number of parameters in the model, and n is the number of observations. A model with smaller AIC or BIC value is treated as a better model.

For the following discussion, we have chosen four GAPC mortality models (LC, APC, RH, and CBD) and used the R package StMoMo to explore these models. There were two reasons for choosing these four GAPC models. First, the LC model generally has the best model fit for all ages among the GAPC family models. The RH and CBD models provide fine modifications to the LC model under certain conditions. The RH model is more appropriate with obvious cohort effects, while the CBD model is a widespread alternative to LC model for the elderly people. On the other hand, the APC model is a popular choice in modeling disease incidence and mortality, and it is easy to provide interpretation.

Our previous studies showed that the other GAPC family models did not produce good fitting results. They either gave much larger fitting/prediction errors or showed no convergence during estimation. The goodness-of-fit of stochastic models can be verified by inspecting the residuals. If there are system errors in the residuals, then spatial patterns (such as hot spots) can be detected. Debón et al. (2008) found that there is spatial autocorrelation for using the LC model to fit mortality rates. However, in the present study, we did not find clusters or autocorrelation for using the LC model to fit the cancer data.

We evaluated the Gompertz model and the above-mentioned four mortality models from the GAPC models, using the R package StMoMo, based on the cancer incidence and mortality rates. The age-specific rates for male were in the format of 5-age group, except for the younger groups since their cancer incidence rate and the number of cancer patients for younger groups were very small. Therefore, the data

were divided into ages 0-14, 15-19, 20-24, ..., and 85-89. In addition, we also considered the model evaluation for the case of higher ages, viz. ages 50-54, 55-59, ..., and 85-89.

For the Gompertz model, we used the weighted least squares to obtain the parameter estimates of B and C . As for the GAPC family models, we have only shown the results under the log-Poisson assumption because the estimation results were about the same as those of the log-Poisson and logit-Binomial assumptions. We do not recommend the estimation under the normal assumption since it would produce unstable results when the population sizes are small (populations smaller than 50,000, according to our experience).

Next, we compared the model fits of Gompertz's model and the GAPC family models. Table 3 shows the MAPE's for fitting the incidence rates and mortality rates of cancer patients for males. The LC and APC models showed the smallest fitting errors. The MAPE's of LC and APC models were very small, indicating that the trend of incidence and mortality rates were well captured. For higher ages, i.e. ages 50-89, the Gompertz and CBD models turned to be good alternatives to the LC and APC models, especially for the mortality rates.

Table 3. MAPE's of Cancer Patients Mortality and Incidence(Male)

		Gompertz	LC	APC	RH	CBD
Incidence Rates	ages 0–89	47.8	2.4	2.8	45.4	50.1
	ages 50–89	16.1	1.6	2.0	55.8	15.6
Mortality Rates	ages 0–89	11.3	4.3	4.3	62.0	11.4
	ages 50–89	5.2	2.7	2.0	25.3	5.2

We also considered the criteria of AIC and BIC to avoid over-parameterization in the models. Since the LC, APC, and CBD had the smallest MAPE's, we have only shown the AIC and BIC values for these three models in Table 4, where smaller AIC

and BIC values are preferred. Since the number of parameters were about the same for these three models, the results of AIC and BIC were similar to those of MAPE. In general, the LC and APC models are preferred and the CBD model is also a possible choice for higher ages.

Table 4. AIC and BIC for LC, APC and CBD models(Male)

		LC		APC		CBD	
		AIC	BIC	AIC	BIC	AIC	BIC
Incidence Rates	ages 0–89	1705	1828	1778	1926	33071	33132
	ages 50–89	954	1011	1012	1088	8191	8239
Mortality Rates	ages 0–89	2146	2296	2157	2348	3219	3315
	ages 50–89	1246	1322	1219	1328	1747	1823

We also examined the differences between observed and estimated rates (or residuals) to double check the estimation results. Figure 4 shows the observed cancer incidence rates and the estimated rates for four GAPC family models, for the cases of Taiwan male in 2002, 2005, 2008, and 2011. Apparently, the fitting curves of LC and APC models almost overlapped with those of the observed values. The RH and CBD models seemed to produce overbiased estimates for the younger ages and sometimes overbiased estimates for the elderly as well. The results for mortality rates estimation were similar and thus the details are not shown.

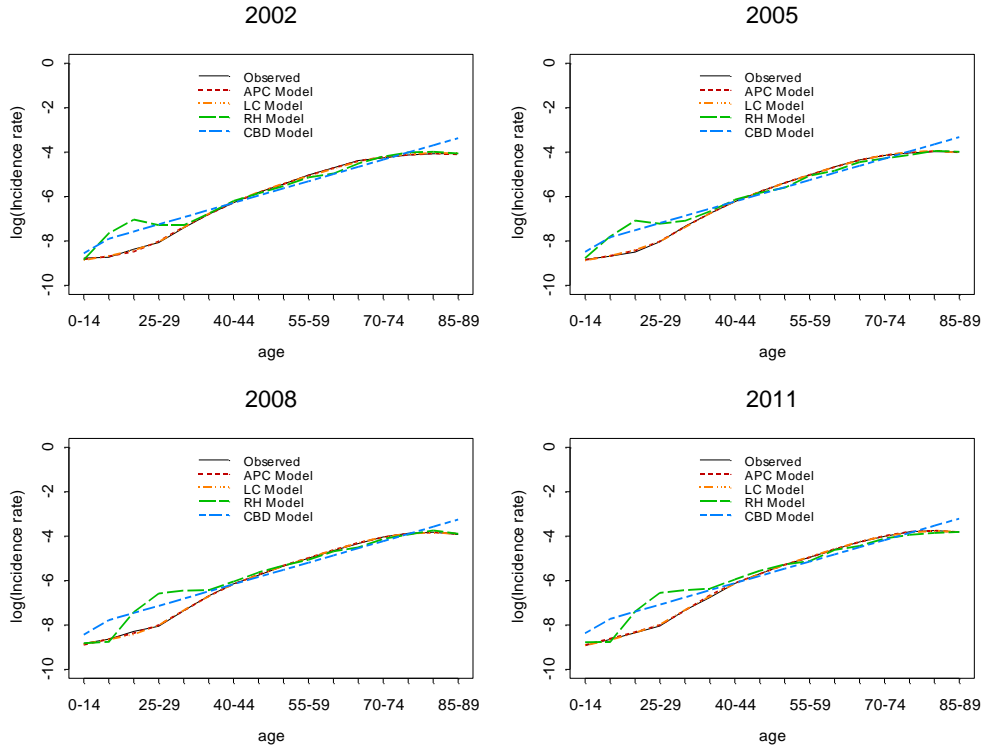


Figure 4. Observed vs. estimated cancer incidence rates (Male)

It seems that both the LC model and APC model are good candidates to model the cancer incidence and mortality rates. We should focus our discussions on the LC model, since we only need to consider the time-related parameter $\kappa_t^{(2)}$ of LC model for prediction. Note that $\kappa_t^{(2)}$ is often a linear function of time and we can derive rough estimates of annual increment/reduction in incidence/mortality rates. If $\kappa_t^{(2)} = a + bt$ then the annual increment/reduction rate at age x is $\beta_x^{(2)} \times b$. Figure 5 shows the annual increment of cancer incidence (left panel) and the annual reduction of cancer mortality (right panel). Older ages showed the largest increment in cancer incidence and younger ages displayed the largest reduction in cancer mortality.

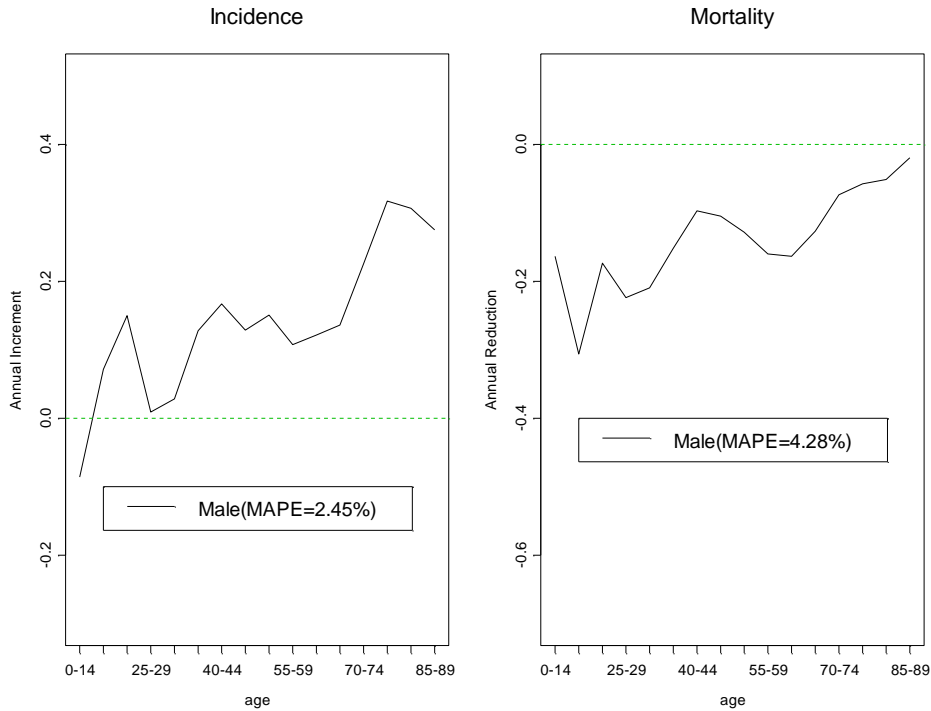


Figure 5. Annual increment/reduction of cancer data (LC Model)

The changes in the incidence rates were positive which indicated an increasing trend, while the changes in the mortality rates were negative, thereby indicating a decreasing trend. There are double risks of insurers in cancer product. The first is that the increasing cancer incidence rate should enhance the benefit paid when the insured is diagnosed with cancer. The second risk is that improvement in cancer patients' mortality rates would add medical benefits burden and annuity, which is similar to the longevity risk for annuity products. Note that, the annual reductions and increments are not a smooth function of age (due to smaller population sizes) and in practice, we can use graduation methods to obtain smoother values (e.g. Wang et al., 2016a).

We evaluated the impact of increasing cancer incidence rates and decreasing mortality rates of cancer patients via the pricing of cancer products. In Taiwan, if the insured is diagnosed with cancer, cancer insurance products usually include clinical

visit benefit and the death benefit. Thus, the increasing incidence rate and the decreasing mortality rate of cancer patients are expected to worsen the loss ratio of cancer products. For simplicity, we considered two types of whole-life cancer products: the first is that a (lump-sum) benefit is paid when the insured was diagnosed with cancer for the first time; The other product is that the annual annuity benefit will be paid after the insured is diagnosed with cancer and still alive.

The first type of cancer product is similar to the whole-life insurance products and the contract is terminated after the benefit is paid. We used the years 2002-2009 as the base-line years and compared the differences in pure premium whether the increasing cancer incidence rates were considered or not. The insured ages for the cancer insurance products were 30 to 60, with a premium payment period of 20 years. Also, the highest attained age was 110 years old and the interest rate was 2.25%. Since the results for different base-line years are similar, we only show the case of 2009 as a demonstration. Table 5 shows the pure premiums per \$1,000 for the whole-life cancer products for males, where “cohort” and “period” indicate the cases with/without the increment in cancer incidence respectively.

Table 5. Pure Premium of 20-Year Payment Whole-Life Cancer
(Per \$1,000, Interest Rate: 2.25%)

Age	Period	Cohort
30	10.58	17.66
35	11.88	19.04
40	13.27	20.47
45	14.74	21.92
50	16.31	23.44
55	18.03	25.08
60	19.98	27.03

Figure 6 shows the percentages of pure premium undercounts if the increment of cancer incidence rate is not considered. The differences in pure premium were very significant, especially for the younger ages, for all base-line years. This can be used as an evidence to explain why the loss ratios of cancer insurance products in Taiwan is increasing in recent years. Also, the differences in pure premium seem to be a decreasing function of time and the differences in 2002 are the largest.

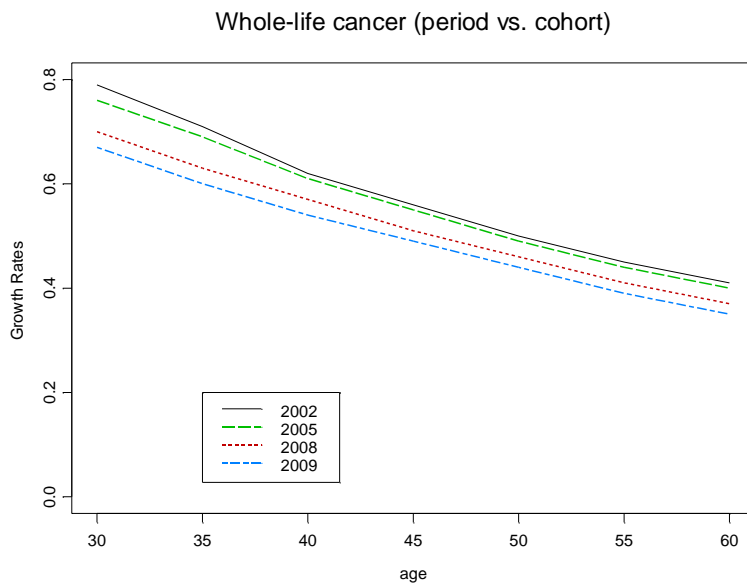


Figure 6. % Difference in pure premium whole-life cancer (LC Model)

Note: Cohort effect means cancer growth rate of incidence is considered.

The second type of cancer product is similar to the whole-life annuity product and the annuity is paid while the insured is alive, after diagnosed with cancer. Again, the year 2009 was used as the base-line year to evaluate the differences of pure premium if the reduction of cancer patient mortality rates is considered. The settings of age, interest rate, and highest attained age were the same as the first type. Table 6 shows the detailed numbers of pure premiums per \$1,000 for the whole-life annuity products. Figure 7 shows the percentage of pure premium undercount when the reduction of cancer mortality was not considered. Unlike in Figure 6, the differences

in pure premium with respect to mortality reduction were not very significant and the younger ages were the largest.

Table 6. Pure Premium of 20-Year Payment Whole-Life Annuity Cancer
(Per \$1,000, Interest Rate: 2.25%)

Age	Period	Cohort
30	1,048	1,074
35	1,042	1,065
40	1,038	1,059
45	1,034	1,052
50	1,029	1,041
55	1,022	1,030
60	1,016	1,020

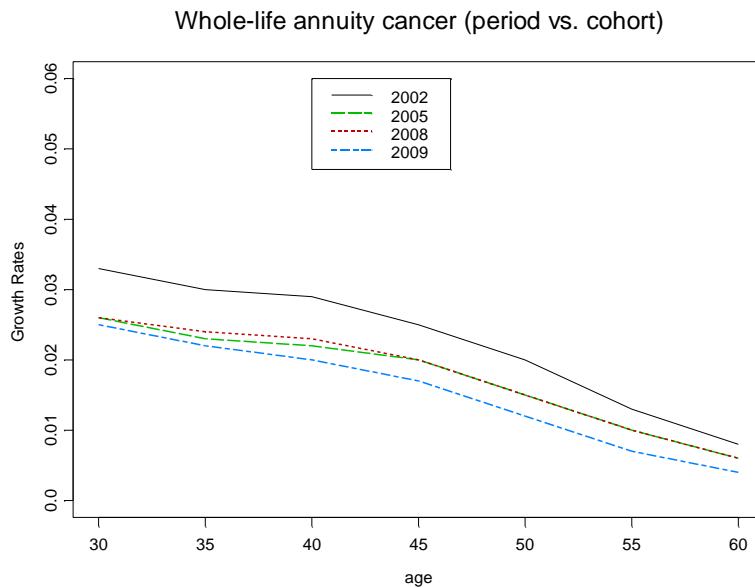


Figure 7. % Difference in pure premium whole-life annuity cancer (LC Model)

4. Discussion and Conclusions

The loss ratios of cancer insurance products in Taiwan worsened in recent years and the longevity risk seemed to exist in long-term health insurance products as well.

Using the databases from Taiwan's National Health Insurance (NHI), we found that the incidence rates slightly increased and the mortality rates gradually decreased over the years in Taiwan. This was in accord with the increasing trend of the loss ratios for the cancer insurance products. This suggested that the longevity risk should be included in long-term health products. We also evaluated whether the frequently used mortality models can be employed to cope with the longevity risk or not. The LC and APC models were the top choices for modeling the cancer incidence and mortality rates. The CBD and Gompertz models were possible alternatives for the higher ages.

Taiwan is not the only country with increasing loss ratios, many Asian countries have similar experiences for long-term health products. For example, the number of those diagnosed with cancer in South Korea increased annually by 2.9 percent between 1999 and 2007, catapulting the loss ratio to around 120 percent (Source: Korea Insurance Development Institute). Lack of relevant experience data and use of foreign experience rates are the main reasons of increasing loss ratio. Therefore, we used Taiwan's population data (i.e. NHIRD) to acquire the cancer incidence and mortality rates. However, this was still not enough to meet our goal to design long-term cancer insurance products. To this end, the notion of longevity risk should be considered. We think that the influence of mortality improvement is bigger than that of the interest in insurance products, and this became more obvious when we looked at long-term and whole-life products. In fact, the mortality improvement is not always a plus to the insurers. In Taiwan, "whole-life" and "return principal" are the two common attributes in most life insurance policies, and we found that the decreasing mortality rates would cause an increase in the insurance premium (Yue and Huang, 2011).

Using the stochastic models to capture the future trend is one of the possibilities for dealing the increasing loss ratios. However, there still remains potential risk, if the

changes of incidence and mortality rates are larger than expected. Natural hedge is one possible alternative, and we can bundle the long-term cancer insurance products with annuity products. Intuitively speaking, cancer patients often have shorter life expectancy and thus receive fewer payments in annuity. Another possibility is to transfer the financial burden of medical claim risk to the capital market, and the experience of the longevity bond can provide useful guidelines.

Furthermore, moral hazards also needs to be considered in cancer insurance. As mentioned in the first section, private clinics or health exam centers can help people to check if they have cancers without showing the exam results in the medical records. This increases the possibilities of adverse selection. There are quite a lot of possible remedies, such as extending the waiting period, reducing the coverage for the first policy year and coverage of the actual expenditures only. For example, a Taiwan's insurance company returns the premium paid to the insured persons if they are diagnosed with cancer in the first policy year, and returns 200% of the total premium paid if they are diagnosed with cancer in the second policy year.

The issue of big data also appeared in this study. We would expect more data to be available in future for the insurance companies, and not restricted to their own experience data. The insurance companies need to invest more on the manpower, such as organizing big data teams and training data scientists, in order to analyze the big data. It is neither cost effective nor feasible to hire outsiders to handle the data. For example, in order to handle the NHI databases (e.g. HV and HV_CD), our big data team has been operating for more than 10 years. Analysis of big data is a process of knowledge accumulation and trial-and-error method. In this regard, we keep updating the standard operating procedure for handling the NHI data over the years.

As for the stochastic models, we considered the GAPC family models in this study. Like in many previous studies, the LC and APC models showed better model

fits. It seemed that the RH model (i.e. cohort modification of the LC model) does not fit well. It can be a result of shorter data period (fewer than 15 years) which makes observation of the cohort effect difficult. For model development, we can also consider other models for cancer incidence and mortality rates. Possible candidates include the discount sequence method and its extension (Wang and Yue, 2011; Wang et al., 2016b) and spatial modification of LC model. For example, Debón et al. (2008) considered adding a term of autocorrelation when the residuals of LC model showed systematic errors. We can add clusters or other first degree moment (i.e. mean shift) to the LC model (Wang and Yue, 2013).

Reference

- Arbeev, K.G., Ukraintseva, S.V., Arbeeveva, L.S., and Yashin, A.I. (2005), “Mathematical Models for Human Cancer Incidence Rates.” *Demographic Research* 12 (10), 237–260.
- Brouhns, N., Denuit, M., Vermunt, J. (2002), “A Poisson log-bilinear regression approach to the construction of projected lifetables.” *Insurance: Mathematics and Economics* 31 (3), 373–393.
- Cairns, A. J. G., Blake, D., & Dowd, K. (2006), “A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration.” *Journal of Risk and Insurance*, 73 (4), 687–718.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009), “A quantitative comparison of stochastic mortality models using data from England and Wales and the United States.” *North American Actuarial Journal*, 13 (1), 1–35.
- Di Cesare, M. and Murphy, M. (2009), “Forecasting Mortality, Different Approaches for Different Causes of Death? The Case of Lung Cancer, Influenza, Pneumonia, and Bronchitis and Motor Vehicle Accidents.” *British Actuarial Journal* 15, Supplement, 185–211.
- Debón, A., Montes, F., Mateu, J., Porcu, E., Bevilacqua, M.. (2008), “Modelling Residuals Dependence in Dynamic Life Tables: A Geostatistical Approach.” *Computational Statistics and Data Analysis* 52, 3128–3147.
- Hoggart, C., Brennan, P., Tjonneland, A., et al., (2012), “A Risk Model for Lung Cancer Incidence.” *Cancer Prevention Research* 5, 834–846.
- Johsnon, Eileen.(2008), “Beyond the Sound Bite: November 2007 Review of Presidential Candidates' Proposals for Health Reform.” *PricewaterhouseCoopers'*

Health Research Institute, Source: JHQ Web Exclusives; Mar/Apr 2008, Vol. 30 Issue 2, pW2–7.

James C. Robinson. (1997), “Use And Abuse Of The Medical Loss Ratio To Measure Health Plan Performance.” *Health Affairs* 16(4), 176–187.

Kruijshaar, M.E., Barendregt, J.J., and Hoeymans, N. (2002), “The Use of Models in the Estimation of Disease Epidemiology.” *Bull World Health Organ* 80, 622-628.

Lee, R.D. (2000), “The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications.” *North American Actuarial Journal* 4 (1), 80–93.

Lee, R.D., and Carter, L.R. (1992), “Modeling and Forecasting US Mortality.” *Journal of the American Statistical Association* 87 (419), 659–671.

Lee, T., Yang, C., and Wang, T. (2011), “Population Aging and NHI Expenditures in Taiwan.” *Population Studies* 43, 1–35.

Lewis, E.B. (1982), “Control of Body Segment Differentiation in *Drosophila* by the Bithorax Gene Complex.” *Embryonic Development, Part A: Genetics Aspects*, Edited by Burger, M. M. and R. Weber. Alan R. Liss, New York, 269–288.

Liu, S.Y. (2000), “Epidemiological transition in colonial Taiwan and its explanation.” Conference Paper for the Conference of Disease History, Academia Sinica, Taipei, Taiwan. (In Chinese)
<http://www.ihp.sinica.edu.tw/~medicine/conference/disease/shihyong.PDF>

National Vital Statistics Reports (2016), Vol. 64 No. 2.

OECD (2015), *Health at a Glance 2015: OECD Indicators*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/health_glance-2015-en

Moger, T.A., Aalen, O.O., Halvorsen, T.O., Storm, H.H., and Tretli, S. (2004), “Frailty Modelling of Testicular Cancer Incidence using Scandinavian Data.” *Biostatistics* 5,1–14.

- Plat, R. (2009), “ On stochastic mortality modeling.” *Insurance: Mathematics and Economics* 45 (3), 393–404.
- Pokhrel, K.P. and Tsokos, C.P. (2015), “Forecasting Age-Specific Brain Cancer Mortality Rates Using Functional Data Analysis Models.” *Advances in Epidemiology*, <http://dx.doi.org/10.1155/2015/721592>.
- Pompei, F. and Wilson, R. (2001), “The Age Distribution of Cancer: the Turnover at Old Age.” *Health Environment Risk Assess* 7, 1619-1650.
- Renshaw, A.E. and Haberman, S. (2006), “A Cohort-based Extension to the Lee–Carter Model for Mortality Reduction Factors.” *Insurance: Mathematics and Economics* 38(3), 556–570.
- Roberts, J.M. and Brewer, D. D. (2006), “Estimating the Prevalence of Male Clients of Prostitute Women in Vancouver with a Simple Capture–recapture Method.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169(4), 745–756.
- Robertson, C. and Boyle, P. (1997), “Statistical Modelling of Breast Cancer Incidence and Mortality Rates in Scotland.” *British Journal of Cancer* 76, 1248– 1252.
- Rosner, B. and Colditz, G. (1996), “Nurses' Health Study: Log-incidence Mathematical Model of Breast Cancer Incidence.” *J Natl Cancer Inst* 88, 359–364.
- Shek, L.L. and Godolphin W. (1988), “Model for Breast Cancer Survival: Relative Prognostic Roles of Axillary Nodal Status, TNM Stage, Estrogen Receptor Concentration, and Tumor Necrosis.” *Cancer Research* 48 (19), 5565–5569.
- Shibuya, K., Inoue, M., and Lopez, A. D. (2005), “Statistical Modeling and Projections of Lung Cancer Mortality in 4 Industrialized Countries.” *International Journal of Cancer* 117(3), 476–485.

- Statistical Abstract of Taiwan Province for the Past Fifty-One Years.(In Chinese)
<http://twstudy.iis.sinica.edu.tw/twstatistic50/>
- Tharu, B.P., Kafle, R.C., and Tsokos, C.P. (2015), “Bayesian Age-Period-Cohort Model of Lung Cancer Mortality.” *Epidemiology Biostatistics and Public Health* 12(3), DOI: 10.2427/11444.
- Uddin, S., Ullah, A., Najma, and Iqbal, M. (2010), “Statistical Modeling of the Incidence of Breast Cancer in NWFP.” *Pakistan. J App Quant Methods Med* 5, 159–65.
- Villegas,A.M., Millossovich, P. and Kaishev, V. K. (2016), StMoMo: An R Package for Stochastic Mortality Modelling. R package version 0.3.1.
 URL <http://CRAN.R-project.org/package=StMoMo>
- Wang, H. C.and Yue, C. J. (2011), “Using Regular Discount Sequence to Model Elderly Mortality.” *Journal of Population Studies* 43, 37-70. (In Chinese)
- Wang, H.C. and Yue, C.J. (2015), “Mortality, Health and Marriage: A Study Based on Taiwan’s Population Data.” *North American Actuarial Journal* 19(3), 187–199.
- Wang, H. C., Yue, C.J., and Tsai, Y. H. (2016a), “Marital Status as a Risk Factor in Life Insurance: An Empirical Study in Taiwan.” *ASTIN Bulletin* 46(2), 487-505.
- Wang, H. C.and Yue, C. J. Chen,Y. X. (2016b), “A Study of Elderly Mortality Models.” *Journal of Population Studies* 52, 1-42.(In Chinese)
- Wang, T. and Yue, C.J. (2013), “Spatial Clusters in a Global-Dependence Model.” *Spatial and Spatio-temporal Epidemiology* 5, 39-50.
- Yue, C.J. and Huang, H. C. (2011), “A Study of Incidence Experience for Taiwan Life Insurance.” *Geneva Papers on Risk and Insurance – Issues and Practice* 36, 718-733.