IRM | institut du risk management

Mémoire présenté pour la validation de la Formation
« Certificat d'Expertise Actuarielle »
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le 27 mai 2021

Par : Antoine BONNANS

Titre : Cancer Insurance: Forcasting incidence and Impact on Profitability

Confidentialité : ☑ NON ☐ OUI (Durée : ☐ 1an ☐ 2 ans)
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de l'Institut des actuaires :

_____

_____

_____

Membres présents du jury de l'Institut du Risk Management :

_____

_____

_____

_____

_____

_____

_____

Secrétariat :

Bibliothèque :

Entreprise : BNP Paribas Cardif

Nom : _____

Signature et Cachet :

Directeur de mémoire en entreprise :

Nom : Fabien PERRUCOT

Signature :

Invité :

Nom : _____

Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels
(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Signature(s) du candidat(s)

Cancer Insurance: Forecasting Incidence Rates and Impact on Profitability

A dissertation submitted in partial fulfillment
of the requirements for the admission at the
French Institute of Actuaries

Antoine Bonnans

December 22, 2020

Abstract

Cancer denotes a large family of diseases, for which incidence & survival rates have generally increased in the past decades. Despite existing national coverages for cancer treatments as well as private health covers, out-of-pocket payments remain a burden for most patients. Then, multiple cancer insurance solutions have arisen. This dissertation deals with lump-sum cancer insurance and presents a pricing model as well as a study of French, Korean and Japanese markets. Then, historical data from Korean and Japanese incidences databases are used to perform incidence rates forecasting. VAR models, Neural Networks and Lee-Carter models are compared to select the optimal model using a nested cross-validation procedure. Models show dynamics in the incidence rates inciting an insurer to consider incidence rates forecasts to evaluate the risk of such products. Finally, a numerical illustration assesses the impact of the future drift of incidence rates on the profitability of a new product.

Keywords: Cancer, Cancer Insurance, Cancer Incidence, Forecasting, VAR, Sparse VAR, Neural Networks, Lee-Carter, Profitability, Korea, Japan.

Résumé

Le cancer désigne une large famille de maladies pour lesquelles l'incidence et les taux de survie ont généralement augmentés ces dernières décennies. Malgré l'existence de couvertures de santé publique pour le traitement du cancer, ainsi que les couvertures privées de soins médicaux, le reste à charge reste un fardeau important pour de nombreux patients. Ce mémoire traite de l'assurance-cancer en tant que garantie forfaitaire, et présente un modèle de tarification ainsi qu'une étude des marchés français, coréen et japonais. Des données historiques de taux d'incidence en Corée et au Japon sont ensuite utilisées pour réaliser des prédictions. Les modèles VAR, réseaux de neurones et Lee-Carter sont comparés pour sélectionner le modèle optimal par une validation croisée imbriquée. Les modèles montrent des dynamiques dans les taux d'incidence qui encouragent un assureur à considérer des prévisions pour évaluer le risque de ces produits. Enfin, une illustration numérique mesure l'impact de la dérive future des taux d'incidence sur la profitabilité attendue d'un nouveau produit.

Mots-clefs : Cancer, Assurance cancer, Incidence du cancer, Prédiction, VAR, Sparse VAR, Réseaux de neurones, Lee-Carter, Profitabilité, Corée, Japon.

Acknowledgments

I am extremely grateful to Fabien Perrucot, for accepting to be the supervisor of this work and the helpful comments and encouragements throughout this journey, as well as Silvina Monchietti and Tiphaine Richebourg for valuable advice and suggestions about cancer insurance products.

The topic of this work have arisen while working at BNP Paribas Cardif Life in Korea. I would like to warmly thank Jung Sup Yoon, Jung Hee Yoon, Min Hye Bae, Bo Ram Kim, Moon Hee Yoon, Ji Yeon Lee, Ji Young Kim, Jung Ha Kim and Myung Hun Kim from Pricing team I was very grateful to join, as well as Kyung Hee Kim, Han Na Yoon, Mi Young Hong, Jin Seok Park, Ji Yun Ha, Se Young Lee, Hyo Joo Lim and Hye Yeon Kim from Valuation team. You have all taught me so much both professionally and personally.

I would like to express my very great appreciation to Didier Féderlé, Martine Duclos, Puters Louis, Alexandre Pertriaux, Virginie Da Costa and Xavier Baradeau I met joining BNP Paribas Cardif during my first steps in the Actuarial Department in Head Office. I wish also to thank Adam Michalski and Jean-Christophe Darbès for their guidance in Seoul, and Patricia Lefèvre and Jean-Paul Félix, for having support me to enroll the French Institute of Actuaries continuing training.

I am very grateful to the excellent speakers of the CEA training, as well as the administrative staff from IRM and the University of Sorbonne.

This work has been the opportunity to share many views about actuarial science with my colleagues. Special thanks for Chobou Raphaël Traoré, for our discussions about the role of the actuary, and the polynomial actuary Marina Chou, for countless advice about Lee-Carter models.

I would also like to extend my deepest gratitude to my parents and my parents-in-law for everyday inspirations.

Finally, I would like to thank to my wife Nayeong for being the most important person in my life.

Contents

Cancer Insurance:
Forecasting Incidence Rates and Impact on Profitability

Executive Summary

Cancer

Cancer denotes a large family of diseases that causes abnormal growth of cells. The cancer incidence is traditionally measured through the Age-Standarized Rate (ASR), which allows to compare different populations, or the crude rate. A usual cancer classification is per site, as in the International Classification of Disease (ICD), while some include the tumor spread. The rates per site are compiled below using ASR rates at a world level in 2018:

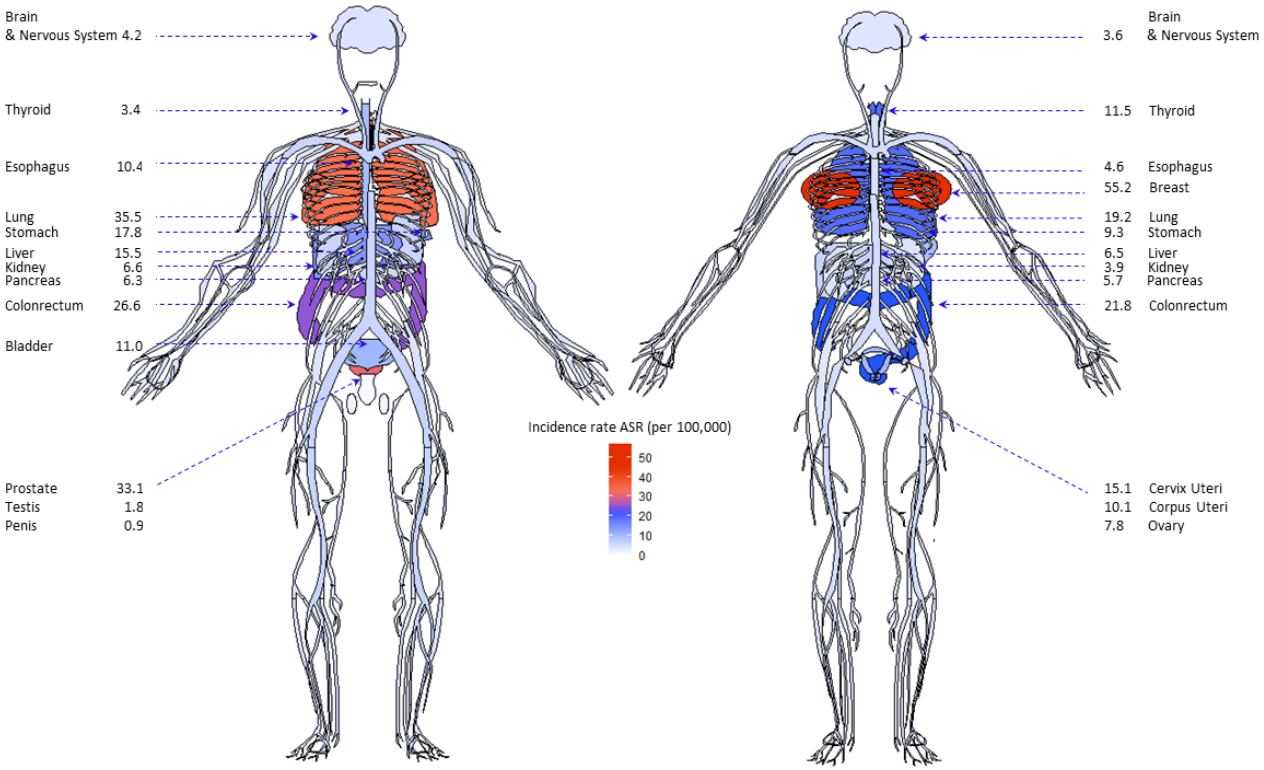


Figure 1: Estimated Cancer Incidence per gender and per site in 2018

Cancer detection techniques include blood test, biopsy and imaging. National detection programmes, named screening programmes encourage early detection. Different types of treatments exist and include surgery, chemotherapy, radiation therapy, immunotherapy or stem cell transplant. Survival cancer is usually measured at a 5-year time horizon and strongly differs by site and severity.

Cost of cancer is a concern at both macroeconomic and individual level. It has a major cost and represents for example 10% of the health expenditures in France. The cost of treatments is often under-estimated by individuals. Then, cancer insurance has emerged in some markets to protect against out-of-pocket payments in case of cancer diagnosis.

Cancer Insurance

Multiple forms of insurance against cancer risk exist including the health expenditures coverages. This dissertation discusses of the cancer insurance products as lump-sum benefits in case of cancer diagnosis. Two markets are studied: Korea, for which a study of public disclosure data shows that insured entry age can vary between 0 and 80 and almost all products are long-term guarantees, usually up to insured age 80 or 100; and Japan, where market penetration rate of cancer insurance reaches 42.6% in 2019.

A cancer insurance pricing model is then exposed for which the benefits are lump-sum for cancer diagnosis, the benefit varying per site and may or not terminate the policy and one maximum benefit per cancer site. Denoting by $q^{(i)}$ the annual incidence rate of cancer site i, the following product is used as an illustration:

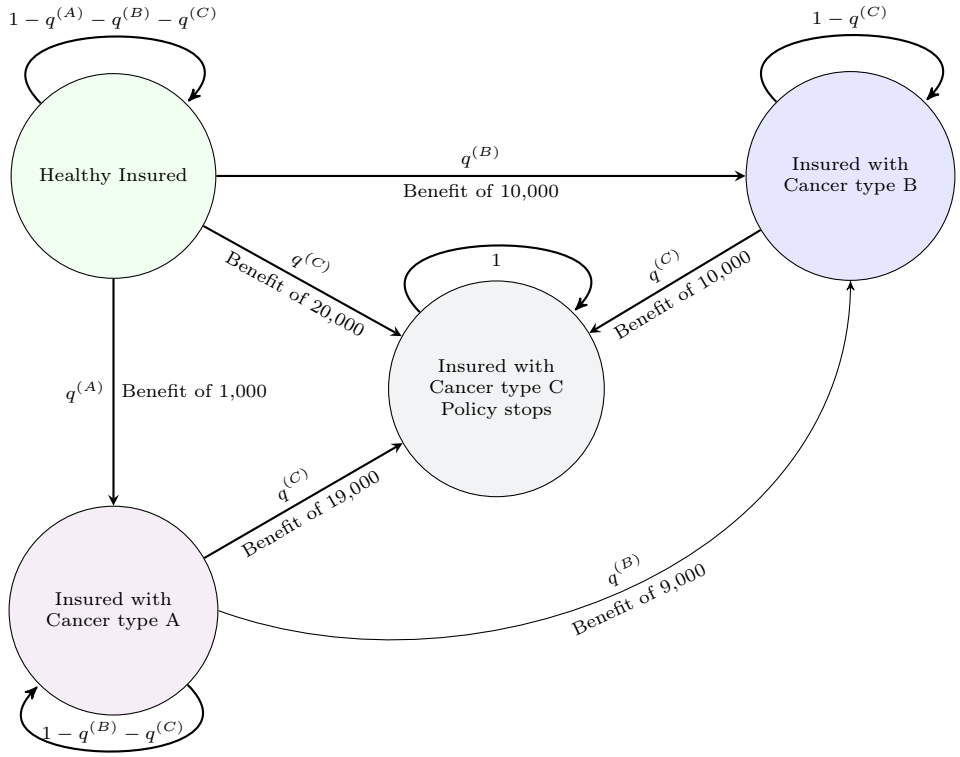


Figure 2: Cancer Insurance Product Transition Diagram with 3 coverages

A pricing illustration for the pure premium is collected in the following table. Cancer type A is skin cancer, type C esophagus, liver or pancreas cancer and type B other cancer types. Policy term is insured 80 years old. Incidences rates are from National Cancer Center in Japan in 2015. Results are as follows:

Age	Gender	Single Premium	Annual Premium	Monthly Premium
40	Male	4,662	131	11
60	Male	4,156	244	21
40	Female	2,921	81	7
60	Female	2,138	117	10

Table 1: Cancer Insurance Product: Pure Premiums Illustration (extract)

Cancer Prediction

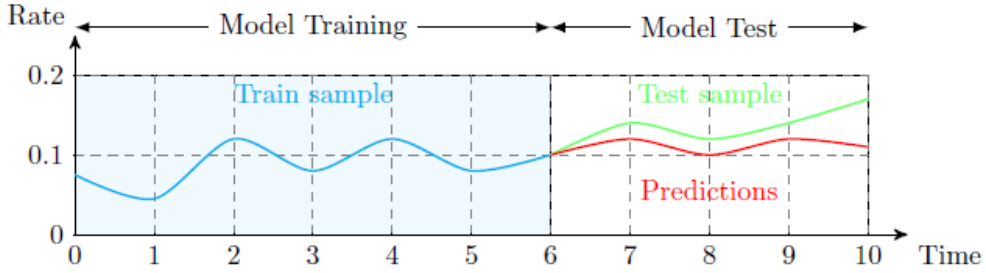
Two databases are used in the dissertation:

(a) Korea National data, from 1999 to 2017, which includes 24 cancer sites
(b) Japan National data, from 1975 to 2015, which includes 37 cancer sites

The following modelling are compared to predict cancer incidence:

1. Naïve modelling, in which forecasts are equal to the last known value
2. Sparse VAR modelling, compared to a VAR modelling
3. Neural Network without and with external regressors (NN I and NN II)
4. Lee-Carter model, traditionally used for mortality modelling

The estimation methods rely on a nested cross-validation procedure, in which the train set increases by 1 observation at each step. One step is illustrated below:



For evaluating the predictions on test sample, i.e. the gaps between test samples values and each model predictions, Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) are used. The decision criteria used to select model is the one with the smallest MSE. The MAPE is a more interpretable indicator and is used to control the quality of predictions. For cancer all causes, results are as follows:

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Cancer All causes	Naïve	6,760	5.5%	Naïve	4,523	8.5%
ICD C00-C96	Sparse VAR	3,985	5.0%	Sparse VAR	2,730	6.1%
	NN I	6,239	5.7%	NN I	5,279	8.8%
	NN II	9,282	5.6%	NN II	4,658	8.0%
	Lee-Carter	4,331	5.5%	Lee-Carter	4,290	7.8%

Table 2: Cross-Validation procedure results for Cancer all causes (extract)

Sparse VAR model is then selected for both male and female. Forecasting shows an anticipated increase of incidence rates, for example to 4.4% and 26.1% for male & female aged 40-44 at a 20-year time horizon:

Age	Male						Female					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
40-44	1.1%	2.2%	4.4%	6.6%	8.8%	11.0%	6.5%	13.0%	26.1%	39.1%	52.1%	65.1%
60-64	4.5%	9.0%	18.0%	27.0%	35.9%	44.9%	5.0%	10.1%	20.3%	30.4%	40.6%	50.8%

Table 3: Evolution between last historical value & forecasts for cancer all causes (extract)

The relative small number of data is a clear limitation for the prediction, especially for the long-term predictions. Then, such forecast underlying assumptions is that future will behave like the past and therefore the possible major advances in cancer detection cannot be included in the modelling. Sensibilities on incidence rates are then required.

Impact on Profitability

The impact of forecasted rates is analyzed in Japan through a product priced with 2015 constant rates, sold in 2020, and for which the profitability is assessed in three scenarios:

1. Scenario I: projection with constant incidence rates (used in pricing)
2. Scenario I*bis*: projection with 2020 forecasted incidence rates
3. Scenario II: projection with forecasted incidence rates from 2020

We assume the following characteristics for product & the sales:

(a) Product design as in pricing illustration: benefits are 10,000 for cancer all causes except those which follow (named general cancer), 1,000 for skin cancer and 20,000 in case of esophagus, pancreas or liver cancer (named special cancer). This latter event stops the policy.
(b) 10 sales segments of 10,000 insured each: (A) denotes male aged 30, (B) aged 40 to (E) male aged 70, (F) denotes female aged 30, (G) aged 40 to (J) female aged 70
(c) Policy and payment term are insured age 80.
(d) Risk margin rate 20%, loadings rate 10%, annual lapse 5%, annual discount rate 5%

The key main indicators are as follow:

Indicators	**Scenario I**	**Scenario I*bis***	**Scenario II**	**Gap I to II**
Number of Claims	12,756	13,282	14,612	+1,854
of which General; Skin; Special	10,894; 212; 1,650	11 330; 246; 1 706	12 366; 383; 1 863	+1,472;+171;+213
Customer Value	78.3%	86.6%	80.8%	+7.3pts
Pure Technical Loss Ratio	83.3%	86.8%	95.1%	+11.8pts
EPV Profit (PVFP)	22.9 million	19.3 million	13.6 million	- 9.3 million

Indicators show an increase of about 7% of claims and an increase of 11.8 points of the Pure Technical Loss Ratio. The profit decreases significantly but remains positive. The following figure presents the loss ratio (ratio between claims incurred and earned risk premium):

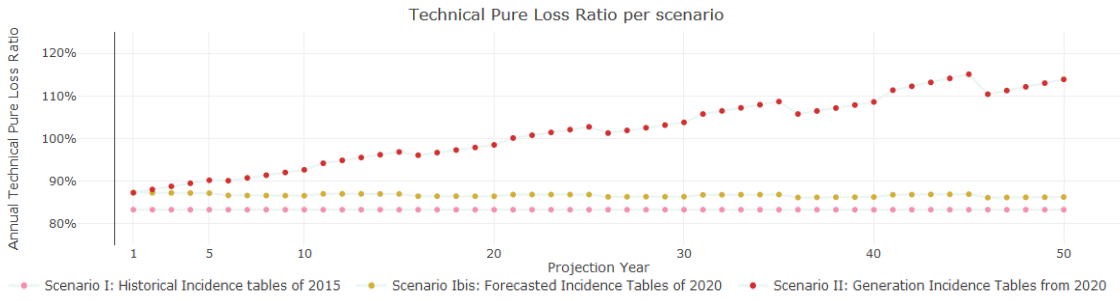


Figure 3: Loss Ratio in scenario I and II

While the loss ratio is scenario I is per definition 83.3% (due to 20% risk margin), the loss ratio rises in scenario II reaching 100% in 2042 and 115% at the end of projection. The analysis of change describes the evolution of the profit from scenario I to II.

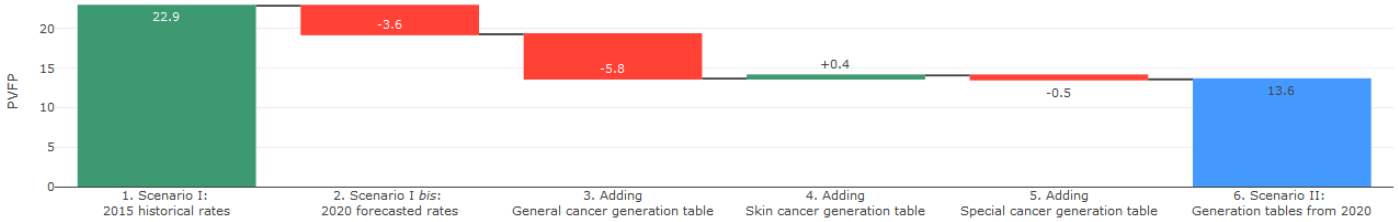
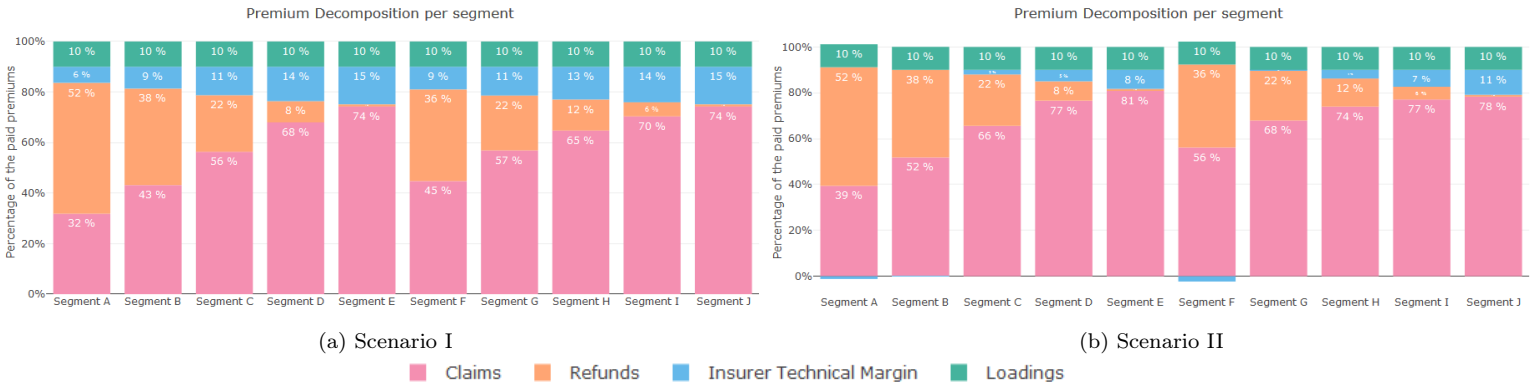


Figure 4: Analysis of movements of the PVFP depending cancer incidence rates assumptions

The impact of the change to 2020 forecasted tables represents 40% of the total variation, while the move to generation tables represents 60%. Results per segment are analyzed below through the premium decomposition, which summarizes the destination of premium received:



(a) Scenario I

(b) Scenario II

Figure 5: Premium decomposition in scenario I & II

While the loadings component is always constant at 10% of the premiums, the technical margins strongly decrease in scenario II compared to scenario I, especially for young ages for which the increase of cancer incidence rates will be the largest throughout their lifetime. The customer value, defined as the premium decomposition going back to the insured (claims and refunds) rise in the scenario II.

Then, some sensibilities on claim levels have been performed with levels of 80% and 120% of expected claims and the presence of anti-selection modelled by +150% claims in year 1 and +125% in year 2:

Scenario I - Sensibilities	No Anti-Selection			Anti-Selection		
Claims level	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
Scenario I	36.2	22.9	6.5	35.1	16.9	-0.7
Scenario II	32.2	13.6	-4.2	27.0	7.3	-11.7

Table 4: Profit in alternative scenarios (extract)

The numerical illustrations show a significant impact on profitability of the incidence rates forecast, with two main effects: the incidence rates change between technical basis and policy inception, and then the incidence evolution until policy term. A decision-making without such information may overly optimistic about product profitability. The discussion evaluates the uncertainty on claim level, due to adverse selection and the general level of claims. Such parameters have a also strong impacts on profitability and variations can turn expected profits into losses.

Overall, cancer insurance aims to reduce financial burden after a cancer diagnosis. For the insurer, it requires a precise knowledge of the risk and anticipations of future incidence. It is also a tool to raise cancer awareness and then accompany the insureds throughout prevention and in case of cancer diagnosis. The development of cancer insurance products is then an useful additional tool in fighting cancer.

Note de synthèse

Le cancer

Le cancer désigne une large famille de maladies qui cause une croissance anormale des cellules. L'incidence du cancer est traditionnellement mesuré à travers le taux standardisé par âge (ASR), qui permet de comparer différentes populations, ou par le taux brut. Les cancers sont classiquement classés selon leur localisation, comme dans la classification internationale des maladies (ICD), tandis que d'autres considèrent la propagation de la maladie. Les taux selon leur localisation sont présentés ci-dessous, avec les taux ASR mondiaux en 2018 :

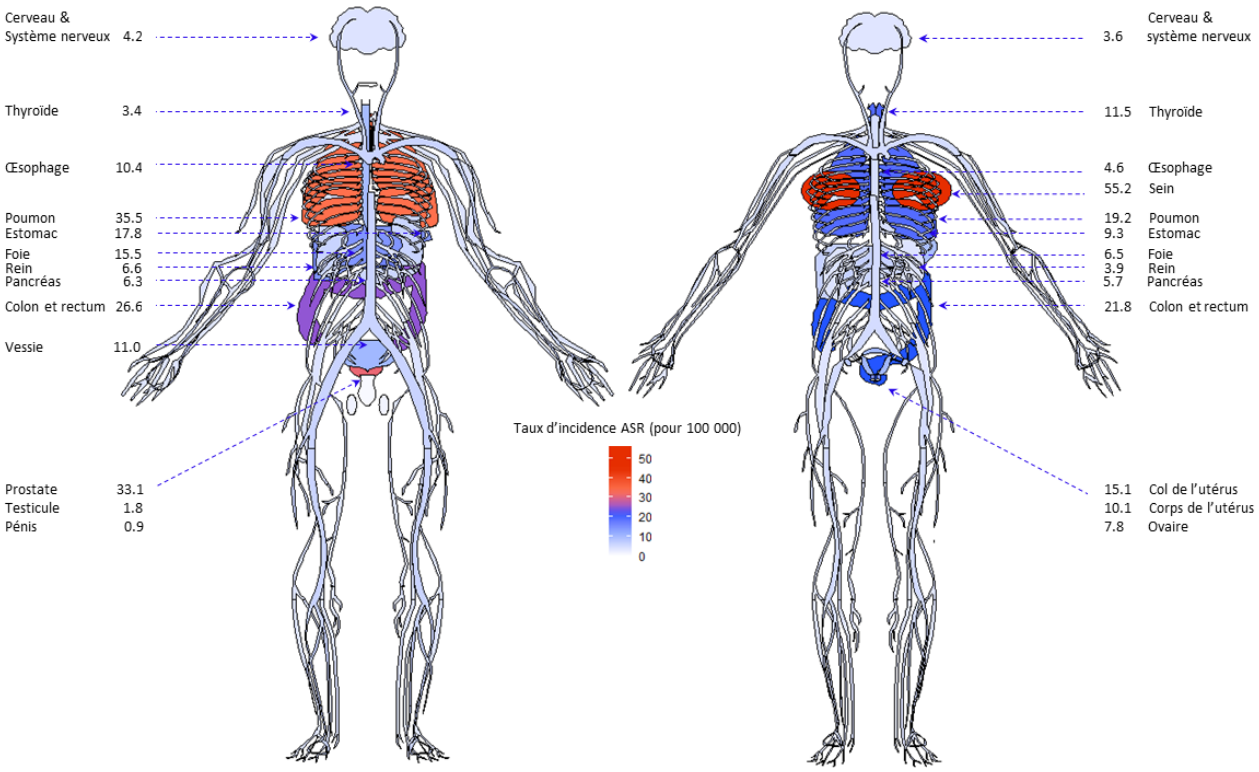


FIGURE 6: Taux d'incidence estimé du cancer par sexe et selon leur localisation en 2018

Les techniques de détection du cancer comprennent les tests sanguins, les biopsies et l'imagerie médicale. Des programmes de détection nationaux encouragent la détection précoce des cancer. Différents types de traitements existent, avec la chirurgie, la chimiothérapie, la radiothérapie, l'immunothérapie ou les greffes de cellules souches. La survie est traditionnellement mesuré à un horizon de 5 ans et varie fortement selon le type et la sévérité du cancer.

Le coût du cancer est une préoccupation à la fois sur le plan macroéconomique et individuel. Il engendre des coûts majeurs, représentant notamment 10% des dépenses de santé en France. Le coût des traitements est souvent sous-estimé par les individus. Des solutions d'assurance-cancer ont alors émergé dans certains marchés pour assurer une protection face aux différents reste à charge suite au diagnostic d'un cancer.

L'assurance-cancer

De multiple formes d'assurance-cancer existent, avec notamment les couvertures d'assurance santé. Cette dissertation traite de l'assurance-cancer comme produit à prestations définis sous forme de capital dans le cas d'un diagnostic du cancer. Deux marchés sont étudiés : la Corée, pour laquelle l'âge possible à la souscription varie entre 0 et 80 ans et dont la majorité des produits sont à couverture longue, jusqu'à l'âge de l'assuré de 80 ou 100 ans ; et le Japon, où la pénétration de ces produits sur le marché atteint 42,6% en 2019.

Un modèle de tarification d'assurance-cancer est ensuite décrit. Une seule prestation par localisation est possible, les prestations en cas de cancer sont des capitaux qui peuvent varier selon la localisation et mettre fin ou non au contrat. En notant $q^{(i)}$ le taux annuel d'incidence d'un cancer de type i, le produit suivant est utilisé comme illustration ($i \in A, B, C$) :

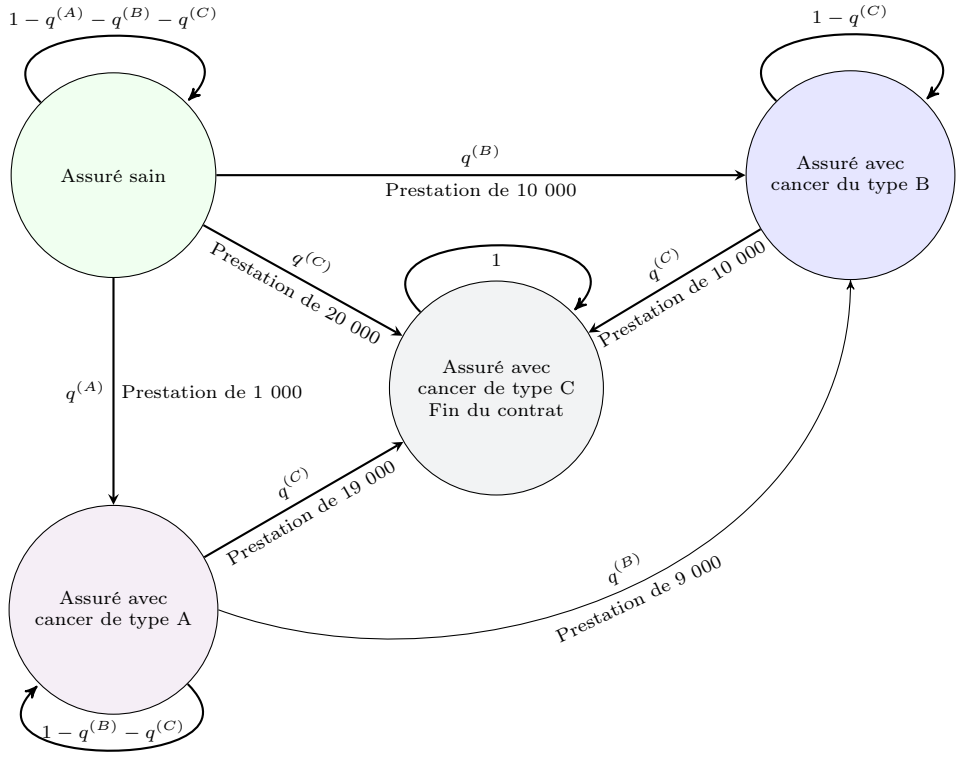


FIGURE 7: Assurance-cancer avec 3 garanties : Diagramme de transition du produit

Une application numérique pour la prime pure de l'illustration précédente est présentée dans le tableau suivant. Le cancer de type A est le cancer de la peau, le type C le cancer de l'oesophage, du foie ou du pancréas, le type B les autres cancers. La garantie prend fin à l'âge de l'assuré de 80 ans. Les taux d'incidence suivent les données nationales du Japon en 2015. Les résultats sont comme suit :

Age	Sexe	Prime unique	annuelle	mensuelle	Sexe	Prime unique	annuelle	mensuelle
40	Homme	4,662	131	11	Femme	2,921	81	7
60	Homme	4,156	244	21	Femme	2,138	117	10

TABLE 5: Assurance-cancer : Primes pures nivelées selon âge et sexe à la souscription (extrait)

Prédiction de l'incidence du cancer

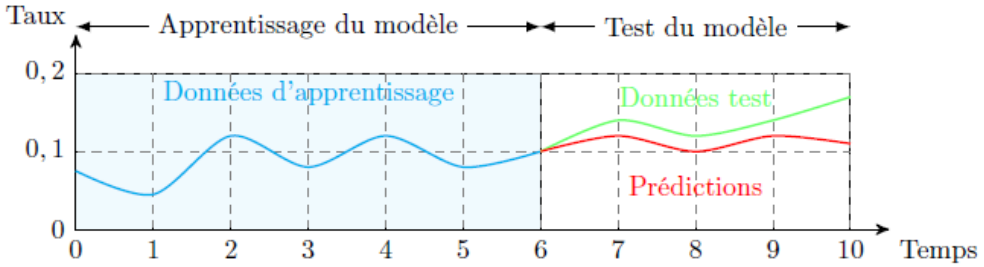
Deux bases de données sont utilisées pour l'étude :

(a) Données nationales coréennes, de 1999 à 2017, qui incluent 24 localisations de cancer
(b) Données nationales japonaises, de 1975 à 2015, qui incluent 37 localisations de cancer

Les modèles suivants sont comparés pour prédire l'incidence du cancer :

1. Modélisation naïve, qui répète la dernière valeur connue
2. Modèle autorégressif Sparse VAR, et VAR pour contrôler sa qualité
3. Réseaux de neurones, sans et avec régresseurs externes (NN I et NN II)
4. Modèle Lee-Carter, classiquement utilisé pour modéliser la mortalité

L'estimation repose sur une procédure de validation croisée imbriquée, illustrée ci-dessous, dans laquelle l'échantillon d'apprentissage augmente d'une observation à chaque étape :



L'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne en pourcentage (MAPE) sont utilisées pour évaluer la qualité des prédictions sur l'échantillon test, c'est-à-dire l'écart entre les valeurs prédites et réelles. Le critère de décision retenu est la MSE minimale. La MAPE, dont l'interprétation est plus aisée, permet de contrôler la qualité de la prédiction. Les résultats pour le cancer toutes causes pour le Japon sont comme suit :

	Hommes	MSE	MAPE	Femmes	MSE	MAPE
Cancer toutes causes	Naïve	6 760	5,5%	Naïve	4 523	8,5%
ICD C00-C96	Sparse VAR	3 985	5,0%	Sparse VAR	2 730	6,1%
	NN I	6 239	5,7%	NN I	5 279	8,8%
	NN II	9 282	5,6%	NN II	4 658	8,0%
	Lee-Carter	4 331	5,5%	Lee-Carter	4 290	7,8%

TABLE 6: Résultats de la procédure de validation croisée pour le cancer toutes causes (extrait)

Le modèle Sparse VAR est sélectionné pour les hommes et les femmes. La prédiction pour le Japon montre une anticipation d'une hausse des taux d'incidence, par exemple de 4,4% et 26,1% pour les hommes et les femmes d'âge 40-44 à un horizon de 20 ans :

Age/Sexe	Homme						Femme					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
40-44	1,1%	2,2%	4,4%	6,6%	8,8%	11,0%	6,5%	13,0%	26,1%	39,1%	52,1%	65,1%
60-64	4,5%	9,0%	18,0%	27,0%	35,9%	44,9%	5,0%	10,1%	20,3%	30,4%	40,6%	50,8%

TABLE 7: Evolution entre valeur historique et prédictions pour le cancer toutes causes (extrait)

Les résultats pour le cancer toutes causes en Corée ne dégage pas d'évolution future. Dans les deux cas, la faible taille de l'échantillon d'apprentissage est une limite certaine pour la prédiction à long terme. Ces prédictions ont pour hypothèse que le futur se comporte de manière analogue au passé, et ne prenne donc pas en compte de possibles avancées en termes de détection du cancer. Des sensibilités sur l'incidence sont donc requises.

Impact sur la Profitabilité

L'impact de la prédiction des taux futurs est analysé au Japon à travers la profitabilité d'un produit tarifé avec des taux constants de 2015, vendu en 2020, évalués dans trois scénarios :

1. Scénario I : projection avec des taux d'incidence constants de 2015 de la tarification
2. Scénario I*bis* : projection avec les taux d'incidence prédits de 2020
3. Scénario II : projection avec des taux d'incidence prédits à partir de 2020 (tables générationnelles)

Les hypothèses de produit et de vente sont les suivantes :

(a) Produit identique à l'illustration de la tarification : prestation de 10 000 en cas de cancer sauf ceux qui suivent (cancer général), 1 000 pour le cancer de la peau et 20 000 en cas de cancer de l'œsophage, pancréas ou foie (cancer spécial). Ce dernier évènement met fin au contrat.

(b) 10 segments de vente de 10 000 assurés chacun : (A) des hommes de 30 ans, (B) de 40 ans ... (E) de 70 ans, (F) des femmes de 30 ans, (G) de 40 ans ... (J) de 70 ans

(c) Fin du paiement des primes et de la garantie à l'âge de 80 ans

(d) Taux de marge sur le risque 20%, de chargement 10%, de rachat annuel 5%, d'actualisation 5%

Les indicateurs principaux sont comme suit :

Indicateurs	Scénario I	Scénario I*bis*	Scénario II	Gap de I à II
Nombre de sinistres	12 756	13 282	14 612	+ 1 854
dont Général ; Peau ; Spécial	10 894 ; 212 ; 1 650	11 330 ; 246 ; 1 706	12 366 ; 383 ; 1 863	+ 1,472 ; +171 ; +213
Valeur client	78,3%	80,8%	86,6%	+7,3 pts
Taux de sinistres sur primes	83,3%	86,8%	95,1%	+11,8 pts
VAP du profit (PVFP)	22,9 millions	19,3 millions	13,6 millions	-9,3 millions

Les indicateurs montrent une hausse de 7% des sinistres et de 11,8 pts du ratio de sinistres sur primes. Le profit décroît fortement mais reste positif. L'évolution annuelle du ratio de sinistres sur primes (sinistres survenus sur la prime de risque acquise) est comme suit :

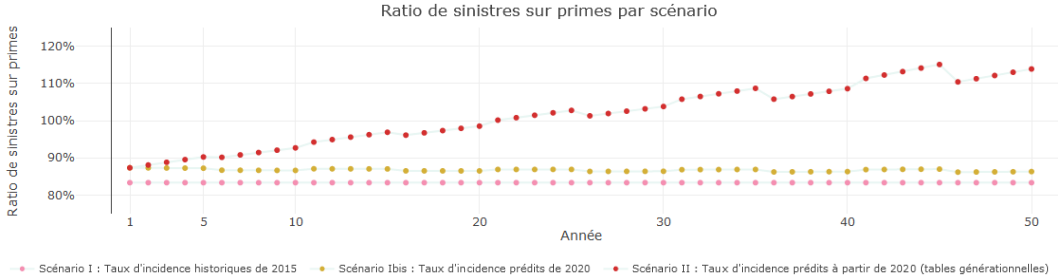


FIGURE 8: Ratio de sinistres sur primes dans les scénarios I et II

Alors que le taux de sinistres sur primes est de 83,3% dans le scénario I (dû à la marge pour risque de 20%), il augmente et atteint jusqu'à 115% dans le scénario II. L'analyse de mouvements suivante distingue l'évolution du profit entre scénario I et II.

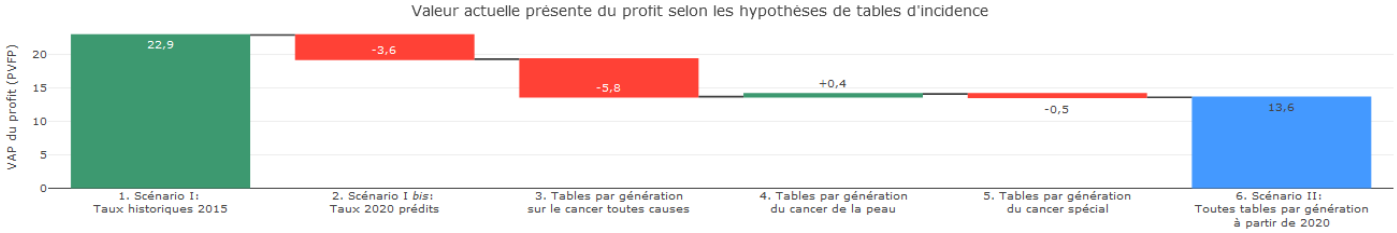
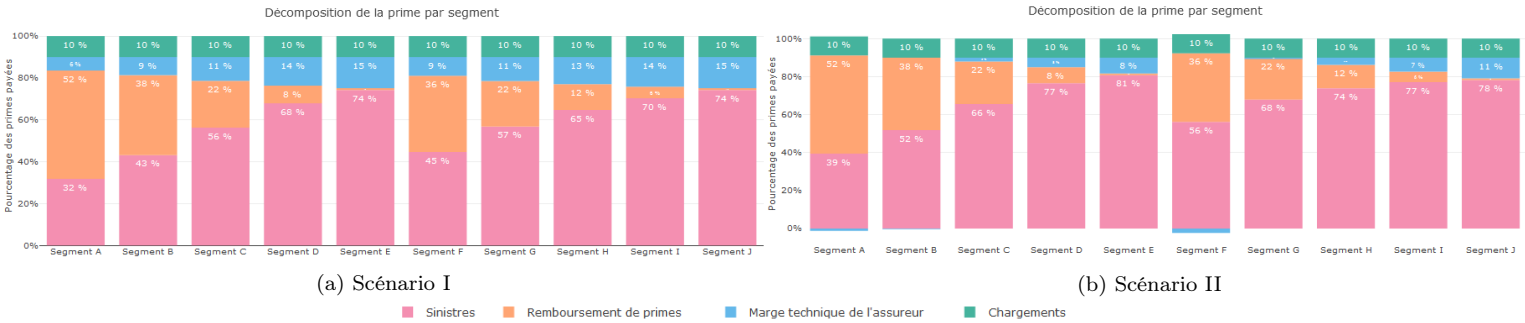


FIGURE 9: Analyse de mouvements de la PVFP selon les hypothèses d'incidence du cancer

L'impact du changement des taux de 2015 aux taux prédits de 2020 représente près de 40% de la variation totale, puis c'est le passage aux tables d'incidence du cancer toutes causes qui représente 60% de la variation totale. Les résultats par segment sont analysés ci-dessous à travers la destination de la prime reçue :



(a) Scénario I

(b) Scénario II

FIGURE 10: Décomposition de la prime dans les scénarios I et II

Alors que la part des chargements reste constante à 10%, la marge technique diminue fortement dans le scénario II par rapport au scénario I, notamment pour les âges jeunes pour lesquelles la croissance des taux d'incidence sera forte au cours de sa vie. La valeur client, défini comme la part de la prime revenant à l'assuré (sinistres et remboursement) augmentent dans le scénario II.

Des sensibilités sur le niveau des sinistres sont ensuite évalués avec des niveaux de 80 et 120% du niveau attendu, et la présence d'anti-sélection modélisé comme un niveau de sinistres supérieur de 150% la première année et 125% la seconde.

Sensibilités	Absence d'anti-sélection			Présence d'anti-sélection		
Niveau de sinistralité	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
Scénario I	36,2	22,9	6,5	35,1	16,9	-0,7
Scénario II	32,2	13,6	-4,2	27,0	7,3	-11,7

TABLE 8: Profit dans les sensibilités sur les sinistres (extrait)

Les illustrations numériques montrent l'impact des taux futurs sur la profitabilité, avec deux effets principaux : l'évolution des taux d'incidence entre la période de la base technique et la commercialisation d'un produit, puis et l'évolution de l'incidence au cours de la vie du produit. L'étude montre l'incertitude sur le niveau des sinistres dû à l'anti-sélection ou le niveau général des sinistres. Ces paramètres ont un fort impact sur la profitabilité et peuvent transformer des profits attendus en pertes.

L'assurance-cancer vise à réduire le fardeau financier lors du diagnostic d'un cancer. Il suppose pour l'assureur une connaissance fine du risque et une anticipation de l'incidence future. C'est aussi un outil pour sensibiliser à la lutte contre le cancer, et accompagner les assurés par des actions de prévention et en cas de sinistre. Le développement de l'assurance-cancer est alors un outil complémentaire dans la lutte contre le cancer.

요약 보고서

암

암은 체내의 세포가 비정상적으로 성장해서 발생하는 질병이다. 암발생률을 측정하는 데는 일반적으로, 다른 연령층의 암발생률의 비교가 가능한 연령준화 발생률(*age standardized rate*)을 사용한다. 그런가 하면 순발생률을 사용하기도 한다.

암의 분류는 종양의 전파성에 의거하는 분류도 있지만, 국제질병 분류상으로는 보통 암의 발생 부의에 따라 분류한다.

2018년도 전세계, 암의 발생 부의에 의거한 연령준화 발생률을 보여준다.

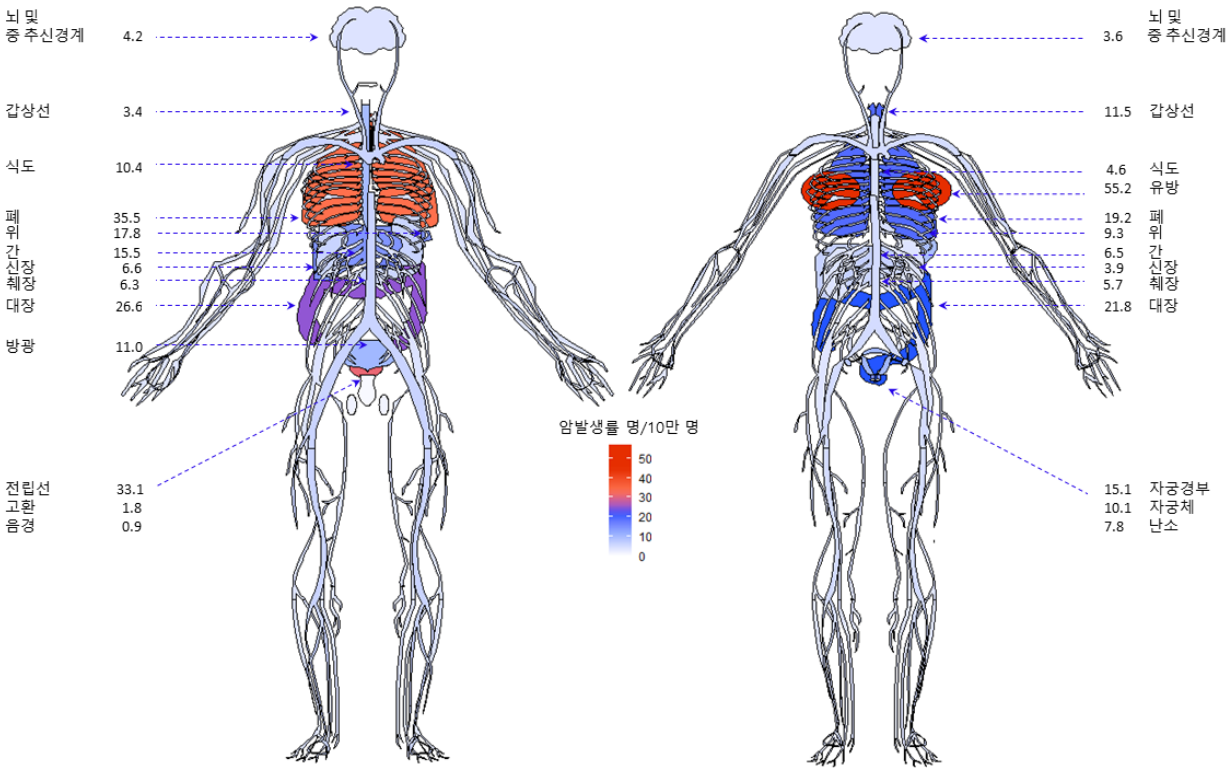


Figure 11: 발생 부위와성별 2018년의 연령준화 발생률

암을 발견하는 기술에는 혈액 검사, 생체조직 검사, 영상촬영 검사가 있다. 암의 조기 발견을 위해 많은 국가에서 국민 암검진을 실시하고 있다.

암을 치료하는 방법으로는 수술, 화학요법, 방사선 치료, 면역요법, 줄기세포 이식 등이 있다. 생존 여부는 일반적으로 5년 단위로 측정되며 암의 발병 위치와 단계에 다라 크게 다를 수 있다.

암은 개인적 치료 비용 부담이 큰 질병이다. 프랑스에서 암 치료비 지출은 의료 비용 지출 전체의 10%를 차지한다. 흔히 많은 사람들이 암 치료 비용을 과소평가하거나 충분히 예측하지 못한다. 그래서 암 진단 시 치료 비용 부담을 줄이기 위해 암 보험이 개발되었다.

암보험

암보험에는 건강보험을 포함해 여러 종류가 있다. 이 논문에서는 암진단판정 시 보험금을 지급하는 모든 암보험 상품에 관하여 논의한다. 연구 대상 시장으로 한국과 일본의 보험시장을 선정했다. 한국 보험 시장의 특징은 상품비교공시에 따라 0-80세 연령에서 가입이 가능하며, 대부분의 경우 장기 보장이 가능해 피보험자 나이 80세에서 100세까지 보장되기도 한다는 점이다. 일본 보험 시장의 경우, 암 보험 가입률이 42.6%에 육박한다는 특징이 있다.

본 논문에서는 이 두 시장의 사례를 분석한 다음 암보험 가격책정표를 제안한다. 보험 가격은 암 진단 보험금을 보장하는 상품으로 암진단 부위 및 계약 종결 여부에 따라 지급금이 결정되며, 부위별 암 진단시 1회 보장 가능한 상품을 기반으로 하고 있다.

i 부위의 연간 암 진단율이 $q^{(i)}$라고 하면 다음과 같이 도식화할 수 있다.

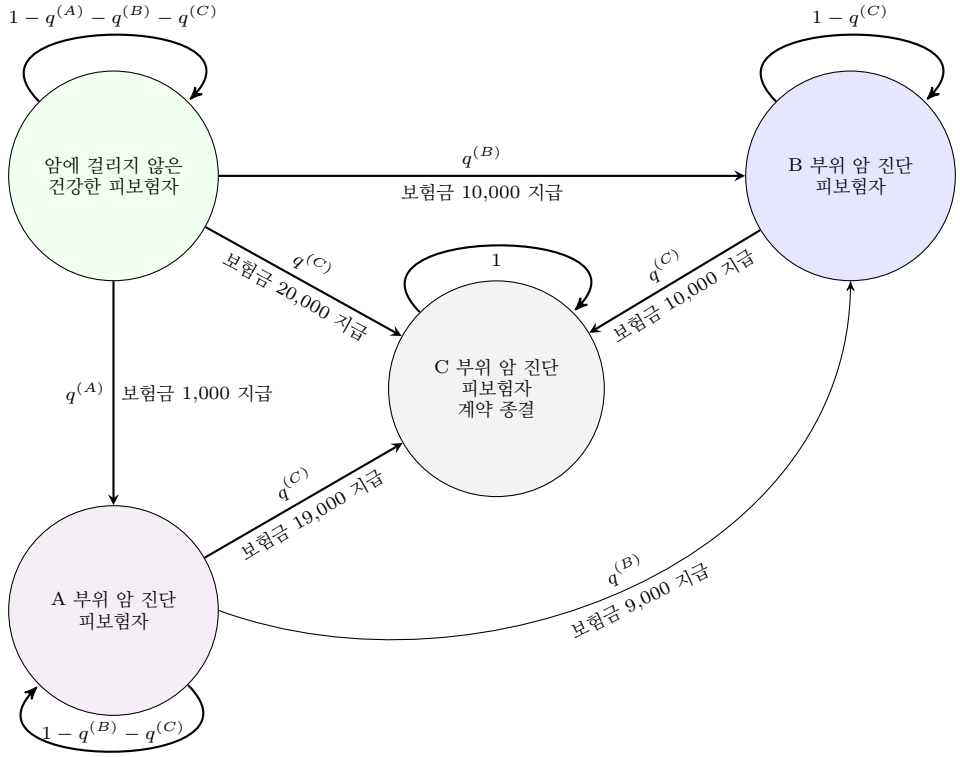


Figure 12: 3가지 암이 보장되는 암보험 상품의 변천 다이어그램 (본문에서 발췌)

A는 피부암, C는 식도암, 췌장암 및 간암, B는 그외 암으로 설정하였다. 다음 표는 위 상품의 암보험 가격 책정 예시 표이다. 보험기간, 보험료, 납입기간에 따른 계산이며, 보장 연령은 80세까지이다.

가입나이	성별	일시불 보험료	연납 순보험료	월납 순보험료
40	남자	4,662	131	11
60	남자	4,156	244	21
40	여자	2,921	81	7
60	여자	2,138	117	10

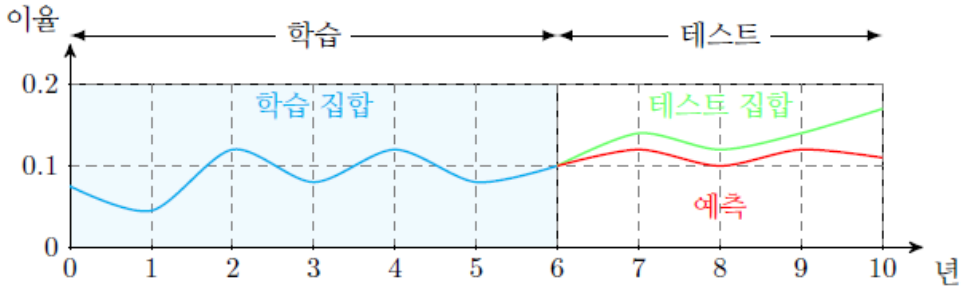
Table 9: 암보험: 순보험료 예시 (발췌)

암발생 예측 모형

본 논문의 암발생 예측 모형에는 두 개의 데이터베이스가 사용되었다.
(a) 24개 부위의 암 관련 1999~2017년 한국 데이터
(b) 37개 부위의 암 관련 1975~2015년 일본 데이터

이 두 가지 데이터에 대하여, 네 가지 측정 모형을 비교 연구한다.
1. 단순 측정 방법(*Naïve*): 마지막 진단 기록이 암 예측 여부를 결정한다.
2. *Sparse VAR* 모형
3. 외부 변수 추가 및 비추가 신경망 (*Neural network*) 모형 (NN I 및 NN II라고 명명)
4. *Lee-Carter* 모형: 사망률 예측에 많이 사용되고 있는 모형이다.

각 모형을 통한 예측값을 교차 검증 (*Nested cross-validation*)을 통해 비교한다. 이 검증 방법은 매 단계마다 학습집합의 기록 수가 하나씩 증가한다는 특징이 있다.



예측의 정확도를 평가하는 지수로는 *Mean Square Error* (MSE) 와 *Mean Absolute Percentage Error* (MAPE)를 사용한다. 최종 선정 모형은 최소 MSE값을 주는 모형이다. MAPE는 MSE에 비해 해석이 더 쉽다는 장점이 있어 예측의 정확도를 평가하는 추가적인 지표로 활용하였다. 일본 데이터를 통한 모든 종류의 암발생 예측 결과표는 다음과 같다.

암의 위치	남자	MSE	MAPE	여자	MSE	MAPE
모든 암	Naïve	6,760	5.5%	Naïve	4,523	8.5%
ICD C00-C96	Sparse VAR	3,985	5.0%	Sparse VAR	2,730	6.1%
	NN I	6,239	5.7%	NN I	5,279	8.8%
	NN II	9,282	5.6%	NN II	4,658	8.0%
	Lee-Carter	4,331	5.5%	Lee-Carter	4,290	7.8%

Table 10: 일본 데이터에서 모든 종류의 암발생 모형별 예측 교차검증 결과표 (본문에서 발췌)

남녀 데이터 모두에 *Sparse VAR* 모형이 선정되었다. 이 모형의 예측 결과에 따르면 일본의 암발생율은 앞으로 증가할 것으로 보인다. 예를 들어 20년 후에 40-44세 남자 암발생률이 4.4% 증가할 것으로 예측했으며, 동일 연령의 여자의 경우 암 발생 증가율은 26.1%로 예측하였다.

연령	남자						여자					
남자	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
40-44	1.1%	2.2%	4.4%	6.6%	8.8%	11.0%	6.5%	13.0%	26.1%	39.1%	52.1%	65.1%
60-64	4.5%	9.0%	18.0%	27.0%	35.9%	44.9%	5.0%	10.1%	20.3%	30.4%	40.6%	50.8%

Table 11: 모든 종류 암발생율의 미래 기간 예측값 (본문에서 발췌)

한국 데이터의 경우, 단순 측정 방법이 *MSE* 기준 가장 정확한 예측 결과를 보였으며 따라서 마지막 진단 기록이 예측값으로 선정되었다.

일본 및 한국 데이터 두 경우 모두, 학습 집합의 기록 수가 작아서 장기적 예측을 하는 것에 어려움이 있었다. 이 예측값들은 미래 암발생이 과거와 동일한 패턴으로 이루어질 거라는 가정 하에 측정된 것이며 연구 시점에서 상정하지 못한 미래 변수들은 고려하지 못한다는 한계가 있다. 이러한 한계를 측정값에 반영하기 위해 스트레스 테스트(*Stress test*)를 시행하였다.

암 진단증가율을 예측하는 것이 수익성에 미치는 영향

일본에서는 암발생율 예측의 효과를 2015년 정액 보험료 보험상품의 수익성을 통해 분석하였다. 수익성 분석은 3가지 시나리오를 통해 이루어졌다.

1. 시나리오 I: 사고율을 상수값으로 계산해 예측 (프라이싱 발생률)
2. 시나리오 I*bis*: 2020년에 대해 예측한 사고율을 통해 예측
3. 시나리오 II: 2020년부터 그 이후까지 예측한 사고율을 통해 예측

상품과 판매에 관한 가설은 다음과 같다.

(a) <암보험> 부분의 3가지 암 보장 상품 예시와 동일한 상품: 일반암 (피부암및 특별암 외의 암진단) 보험금이 10,000, 피부암 진단 보험금 1,000, 특별암 (식도암 췌장암 간암 중에서) 보험금이 20,000. 특별암 진단 시에만 보험계약이 종료된다.
(b) 각 1만명 피보험자가 있는 10개 판매부문: 부문 (A) 가 30세 남자, 부문 (B) 가 40세 남자, . . . , 부문 (E) 가 70세 남자, 부문 (F) 가 30세 여자, 부문 (G) 가 40세 여자, . . . , 부문 (J) 가 70세 여자
(c) 보험금 지급 및 보장은 80세에 종료된다.
(d) 위험마진 (*risk margin*) 20%, 예정사업비율 10%, 연해지율 5%, 연할인계수 5%

아래에 지표들이 있다:

Indicators	**시나리오 I**	**시나리오 I*bis***	**시나리오 II**	**Gap I-II**
보험금지급청구건수	12,756	13,282	14,612	+1,854
일반, 피부, 특별	10,894; 212; 1,650	11,330; 246; 1,706	12,366; 383; 1,863	1,472; +171; +213
소비자 가치	78.3%	80.8%	86.6%	+7.3 pts
손해율 (*Pure Technical Loss Ratio*)	83.3%	86.8%	95.1%	+11.8 pts
손익현가 (*PVFP*)	22.9 million	13.6 million	19.3 million	- 9.3 million

시나리오 II는 시나리오 I에 비해서 보험금 지급 청구건수가 10% 더 높고 손해율은 7점이 더 높다. 따라서 수익은 감소한다. 연간 손해율의 변화는 다음 그림과 같다.

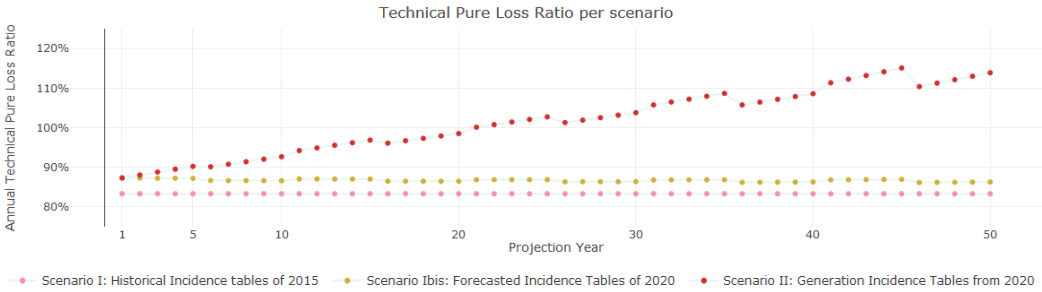


Figure 13: 시나리오 I&II에 손해율

시나리오 I의 경우 손해율이 83.3% (위험마진이 20%이므로)인데 반해, 시나리오 II의 경우 손해율이 2042년에 100%까지 증가하며 마지막에는 115%에 달한다. 다음의 변화 그래프는 시나리오 I과 II의 수익 증감을 나타낸다.

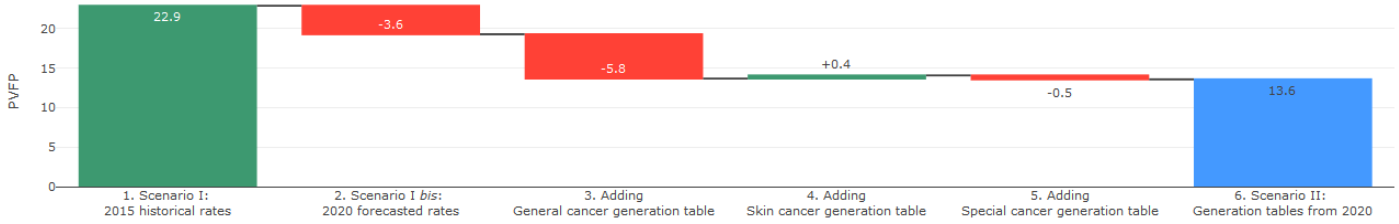
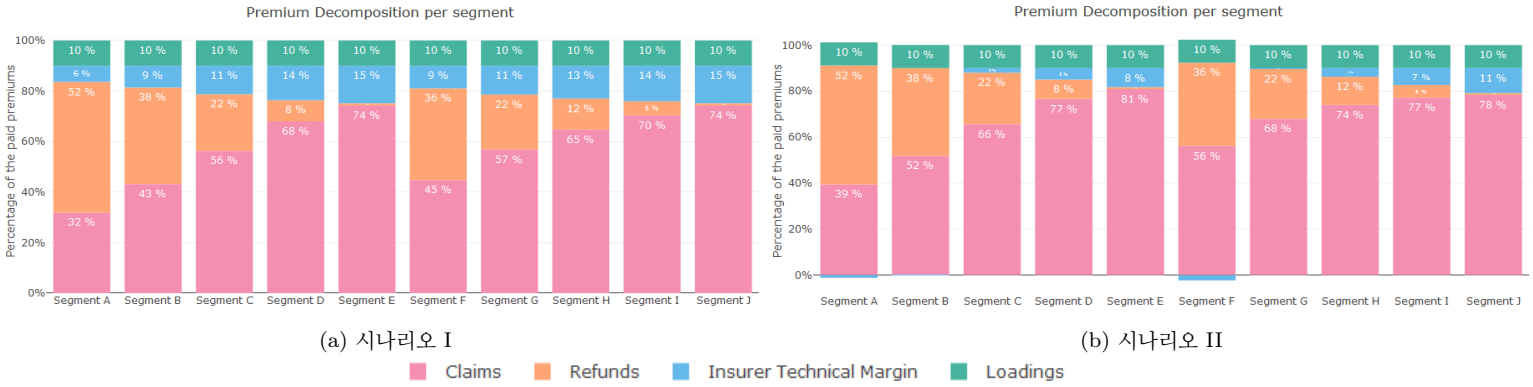


Figure 14: 시나리오 I과 II의 수익 변화

2015년 암발생율 변화가 2020년 예측값에 미치는 영향은 40%에 달한다. 사고율 표의 경우 모든 종류 암 발생율이 60% 가량 변동되었다. 부분별 결과는 다음과 같이 보험료의 지출 내역을 통해 분석할 수 있다.



(a) 시나리오 I

(b) 시나리오 II

Figure 15: 시나리오 I & II에 보험료 분해

예정사업비율이 10%인 반면 시나리오 II의 사차익(*technical margin*)이 시나리오 I에 비해 크게 감소한다. 이와 같은 현상은 미래 사고율이 크게 증가하는 30대에서 가장 두드러진다. 소비자 가치 (보험금과 해지환급금 나누기 보험료)는 시나리오 II에서 증가를 보인다.

다음으로 클레임의 크기에 대한 민감도를 측정하였다. 예상 클레임 크기의 80%와 120%를 테스트했다. 그리고 1년 째에 발생률 곱하기 150%과 2년 째에 발생률 곱하기 125%로 역선택이 있는지 평가한다.

유리한 시나리오	역선택 없음			역선택 있음		
발생률 레벨	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
시나리오 I	36.2	22.9	6.5	35.1	16.9	-0.7
시나리오 II	32.2	13.6	-4.2	27.0	7.3	-11.7

Table 12: 대안 시나리오들에 이익

측정값들은 미래 발생율이 수익에 미치는 영향을 보여주고 있다. 미래 발생율은 두 가지 측면에서 수익에 영향을 미칠 수 있다. 첫째는 분석 시점과 상품 개시 시점 사이의 사고율 변화이다. 둘째는 상품의 생애주기 동안의 사고율 변화이다. 본 연구는 역선택에 의한 클레임 수치의 불확실성을 보여주고 있다. 이와 같은 요소들은 수익성에 큰 영향을 미치며 기대 수익을 손실로 바꿀 수도 있다.

암보험 상품은 암진단 판정 시 재정적 부담을 감소하기 위한 것이다. 보험사에게는 위험에 대한 섬세한 이해와 미래 사고에 대한 정확한 예측을 요구한다. 또한 암보험은 암이라는 질병에 대한 인식을 개선하는 도구인 동시에 클레임과 예방활동 등을 통해 피보험자를 동반하는 과정이기도 하다. 결국 암보험 상품의 개발은 인간이 암과의 투쟁을 극복할 수 있도록 옆에서 돕는 조력자의 업무인 것이다.

Notations

The following table compiles the different notations used in this dissertation.

Actuarial Notations:

Notation	Description	Definition
x	Insured Policy Age at Policy Inception	Assumption
t	Time Period	Time Index
n	Policy Term	Assumption
m	Payment Term	Assumption
mm	Premium Frequency	1 if annual, 12 if monthly
q_{x+t}	Annual Mortality Rate	from Mortality Table
q^{0}_{x+t}	Annual Cancer Incidence Rate (cancers all site)	from Incidence Table
q^{i}_{x+t}	Annual Cancer Incidence Rate from Site i	from Incidence Table
w_{x+t}	Annual Premium Waiver Incidence Rate	from Incidence Table
l_{x+t}	Active Population	$l_{x+t} = l_{x+t-1} \times (1 - q_{x+t})$
l'_{x+t}	Payment Population	depends on waiver definition
ω	Terminal Age	Assumption
d_{x+t}	Number of deaths at year $x+t$	$d_{x+t} = q_{x+t} \times l_{x+t}$
i	Technical Rate	Assumption
ν	Discount Rate	$\nu = 1/(1+i)$
C_{x+t}	First Commutation Number for Claims	$C_{x+t} = \nu^{(t+0.5)} \times d_{x+t}$
M_{x+t}	Second Commutation Number for Claims	$M_{x+t} = \sum_{k \geq t} C_{x+t}$
D_{x+t}	First Commutation Number for Life	$D_{x+t} = \nu^{(t+0.5)} \times l_{x+t}$
N_{x+t}	Second Commutation Number for Life	$N_{x+t} = \sum_{k \geq t} D_{x+t}$
$N^{(m)}_{x+t}$	Monthly Approximation of $N_{x+t} - N_{x+t+m}$	$N_{x+t} - N_{x+t+m} - \frac{(mm-1)}{2mm} \times (D_x - D_{x+m})$
P	Premium	Calculation result
$P^{(m)}$	Premium paid m times a year	Calculation result
a	Risk Margin rate	Assumption
l	Loadings rate	Assumption
ϕ	Percentage of Return on Premiums	Assumption
SA	Sum Assured	Assumption
S	Incurred Claims	Calculation result
$\alpha^{(i)}$	Percentage benefit for cancer type i	Assumption
T_i	Dummy variable to assess if the policy stops after cancer type i diagnosis	Assumption
W_i	Dummy variable to assess if premiums are waived after cancer type i diagnosis	Assumption

Mini-Dictionary

English-French non-cognates medical terms are compiled below for French-speaking readers:

English	French
Bladder	Vessie
Blood	Sang
Bloodstream	Système sanguin
Bone	Os
Bone marrow	Moelle osseuse
Brain	Cerveau
Breast	Sein
Cervix uteri	Col de l'utérus
Corpus uteri	Corps de l'utérus
Ear	Oreille
Esophagus	Œsophage
Gallbladder	Vésicule biliaire
Heart	Cœur
Hoarseness	Enrouement, éraillement (de la voix)
Kidney	Rein
Leukemia	Leucémie
Lip	Lèvre
Liver	Foie
Lump	Bosse, protubérance
Lung	Poumon
Lymph nodes	Nœuds lymphatiques, ganglions lymphatiques
Skin	Peau
Sore	Lésion, plaie
Stem Cells	Cellules souches
Stomach	Estomac
Stroke	Accident Vasculaire Cérébral
Testis	Testicules
True Pelvis	Pelvis vrai, petit bassin

List of Figures

List of Tables

Introduction

Cancer denotes a large family of diseases involving abnormal cell growth. It is a major concern for people due to impacts on health, private occupations and cancer-related deaths. It is also a key point in the health systems and health policies, involving cancer awareness, screening programmes, treatments as well as research and use of advanced technologies for fighting Cancer.

Cancer varies by site and severity. Cancer survival rates have very much improved and therefore more people live with cancer than die due to cancer. However, treatments costs are getting higher, involving multiple long-term treatments types. Cancer treatment price perception is usually low and depending on health system and health insurance, out-of-pocket payments can be a financial burden.

Due to these factors, insurance solutions about Cancer have arisen. Two types of cancer insurance exist: medical expenses policies and lump-sum benefit policies. The aim of this thesis is to present and study the latter, named cancer insurance products. In such products, the insurer pays a lump sum to the insured at cancer diagnosis. The lump sum usually depends on Cancer type. Such products are especially very popular in Asia but not widespread in Europe, and especially in France. What are the main characteristics of cancer insurance products?

Designing a cancer insurance product especially requires a strong assessment of the risk. A major concern is the confidence of cancer incidence rates used in the pricing, especially for long-term products. Cancer incidence are generally rising and this evidence must be anticipated by the insurer. Then, which techniques can be used to predict cancer incidence?

Then, assessing the product profitability is very much associated with a good prediction of cancer incidence. Profitability testing implies best-estimate information, including cancer incidence rates. Therefore what is the impact of cancer incidence prediction on profit testing?

This document is organized as follow: the first chapter collects some elements about Cancer. The second chapter describes the development of cancer insurance, discuss product features and pricing, as well as a market analysis of Korea and Japan. Then, the third chapter discusses the cancer incidence modelling and prediction. Numerical illustrations are given using real data from Korea and Japan. The fourth chapter discusses the profitability with and without cancer incidence prediction using a projection model.

1 Cancer

1.1 Overview

Cancer is a disease of cells. This generic name denotes a large family of diseases involving abnormal cell growth [1]. Cancer has been known from Antiquity, with the first occurrence in the Edwin Smith papyrus - an ancient Egyptian surgical manuscript. The word  , translated by tumor or cancer, is used to describe a breast disease which has no treatment [2]. The word cancer is derived from the Greek *carcinos* (καρκίνος) which means *crab* and has been introduced by the Greek physician Hippocrates. The modern era of cancer research really began in the 19^{th} century [3]. Then, progress of medicine, biology have deeply improved the knowledge of these diseases.

Cancer incidence can be measured by crude incidence rates, which denotes the number of diagnosed people over the number of lives in a time period. The crude incidence rate, denoted incidence rate hereafter, helps to evaluate the cancer burden for a total population through time. However, it may hide population aging effects. Therefore crude rates can be used to compare rates at a given age.

Cancer incidence comparison per country use an indicator named *age-standardised rate* (ASR) which aims to compare areas with different age structures. The ASR is a weighted mean of the age-specific rates. The standardized weights are available in the appendix. The following map, from the International Agency for Research on Cancer (IARC) represents the cancer incidence using ASR in 2018 per country for both sexes:

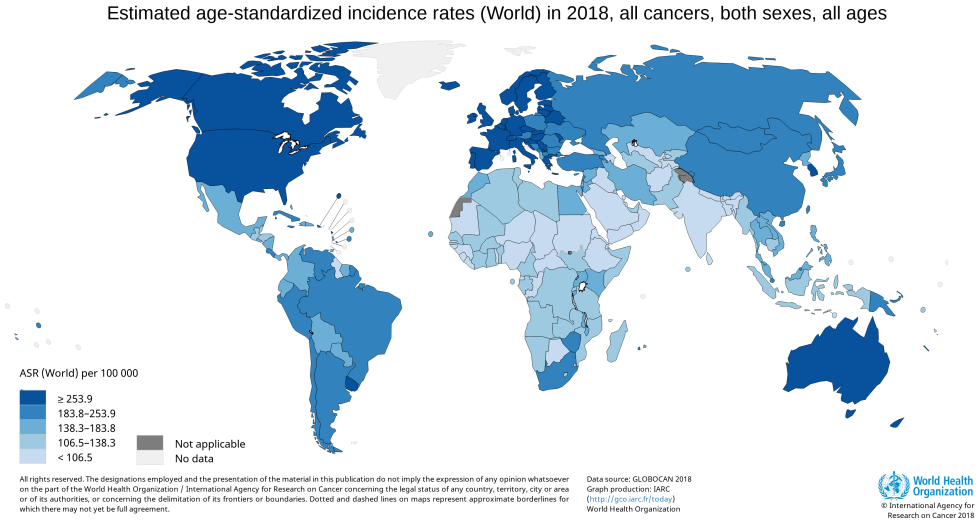


Figure 16: Cancer Incidence per Country in 2018

The map shows that at equivalent age structure, annual cancer incidences rates are higher in developed countries (North America, Europe, Australia) exceeding 2.5 per 1,000 and lower in Middle Africa, Middle East and India. Risk factors as well as cancer detection are two issues discussed later in this chapter.

The previous figure, regarding incidence rates, does not show the burden of cancer in terms of number of new cases. Despite that the ASR is higher in North America and in Europe, due to to demographic weight, it is in Asia that the cancer incidence in the highest in numbers. The figure below shows the number of cancer cases per region in 2018. Almost 1 cancer case out of 2 is in Asia, with more than 8 million new cases. Then, Europe counted 4.2 million new cases and North America almost 2.4 million, Latin America 1.4 million and Africa 1 million.

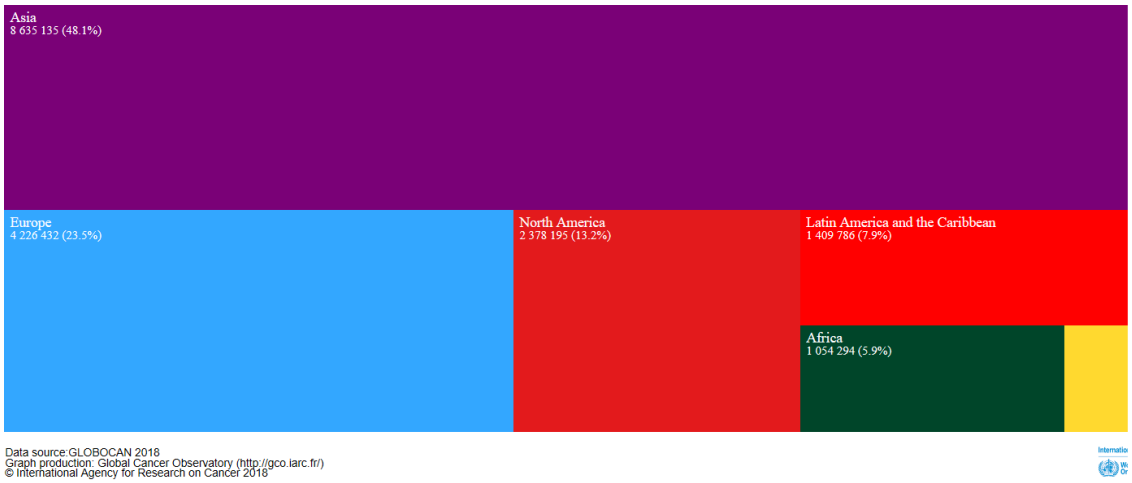


Figure 17: Number of cancers cases in 2018 per region

The burden of cancer can also be expressed using statistics for a given person to be diagnosed cancer over their lifetime. Such calculations requires to take into account both cancer incidence rates and mortality rates. The lifetime risk of cancer in the UK is 38.5% for men and 36.7% for women born in 1930, and 53.5% for men and 47.5% for women born in 1960 [4]: 1 in 2 people will be diagnosed cancer in their lifetime. The cumulative risk for a live aged 120 years reaches almost 90% for men and 70% for women. Almost all men would be diagnosed cancer over their lifetime in the case they would have not died of other causes first.

Figure 18: Probability to get cancer over lifetime in the UK for the cohort born in 1960

Such statistics consider cancer as one disease and does not take into account the variety of cancer types. The next section describes the different types of cancers.

1.2 Different Types of Cancers

Cancer types are usually classified depending on the cell type originally affected or the organ affected. This latter is named a classification per site. Also, some classifications involve the size or the spread of a tumor.

1.2.1 Cancer Classification per Cell Type Originally Affected

Using the cell type, the five types are:

- **Carcinomas** (from Greek καρκίνος *crab* and -oma *disease*) are tumors that appear in the tissues lining the body's organs. Most of cancer cases are carcinomas. Carcinoma in situ denotes a pre-cancer stage located in the epithelium (which lines the outer surfaces of organs) but not in the organ itself.
- **Sarcomas** (from Greek σάρξ *flesh*) are tumors that originate in the body's bone, muscle, cartilage, fibrous tissue or fat.
- **Leukemia** (from Greek λευκός *white*, and αἷμα *blood*) are cancers of the blood or bone marrow. It denotes a large family of diseases which attacks the different blood cells types. Acute leukemia produces a large number of abnormal white blood cells which interferes with the production of normal blood cells. Chronic leukemia denotes multiple diseases in which blood cells production is damaged.
- **Lymphomas** (from Latin lympha *water* and -oma) affect the lymphatic system, which distributes nutrients to the blood and tissue and prevents bacteria and other foreign "invaders" from entering the bloodstream. There are more than 20 types of lymphoma.
- **Myeloma** (from Greek μυελός *marrow* and -oma) denotes blood cancers arising from plasma cells.

The different types are illustrated below[1]:

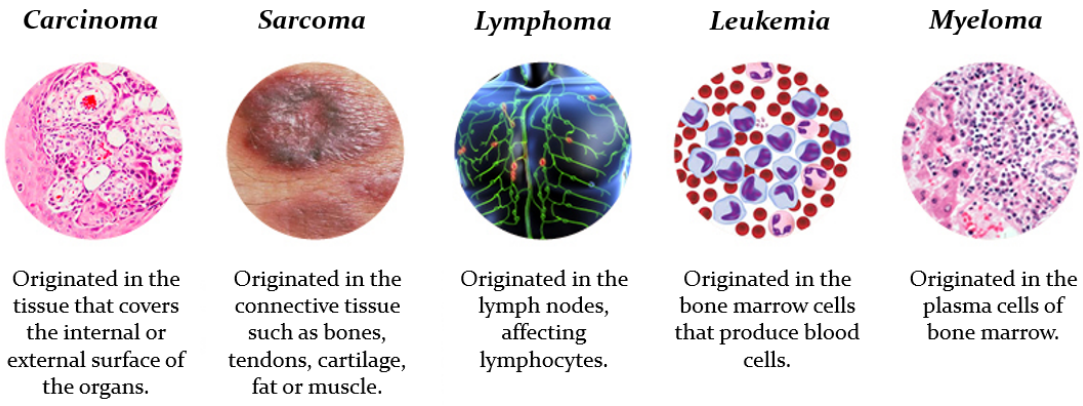


Figure 19: Different types of cancer

The classification per site is detailed hereafter.

[1] Illustration based on https://www.afmedios.com/2017/02/dia-mundial-contra-el-cancer/

1.2.2 Cancer Classification per Site

The most common classification is based on International Classification of Disease (ICD), edited by the World Health Organization (WHO). This classification defines a standard code for each disease type, including cancers. The ICD is regularly updated; ICD-10 was published in 2011 and will be updated by ICD-11 from 2022. ICD allows comparison between different areas and different time periods. Malignant neoplasms codes are from C00 to C96, and each number is dedicated to cancer site. For example, C16 denotes stomach cancer. Then sub-codes exists to denote the exact location in the organ.

The Internal Agency for Research on Cancer (IARC) collects some statistics from national agencies through the Cancer Incidence in Five Continents initiative[2] to provide estimations about the cancer burden per cancer site. At a world level, the most common cancers in 2018 are, by sex, as follows:

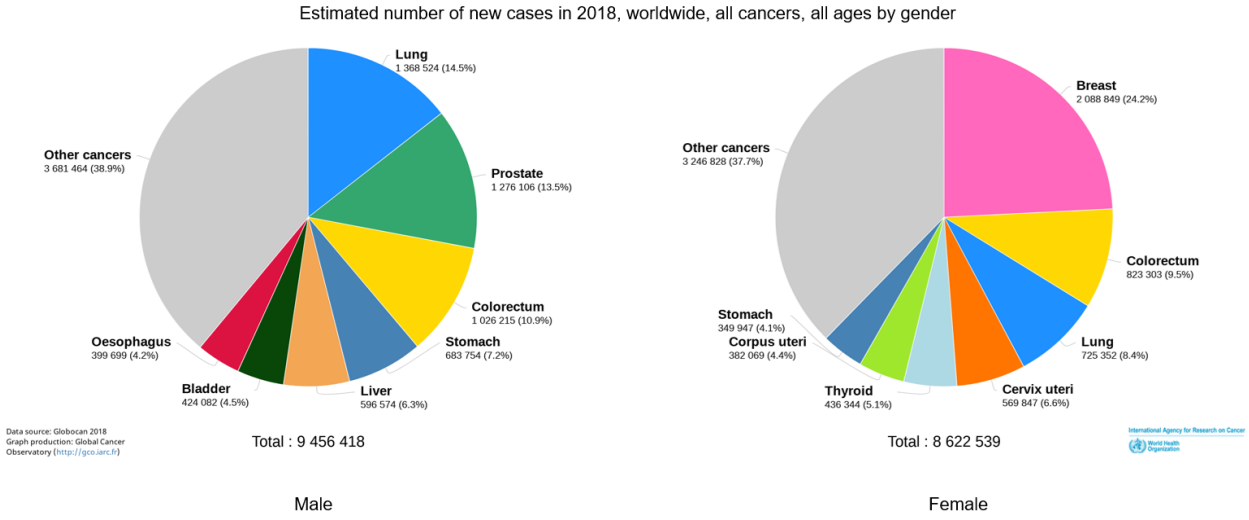


Figure 20: Estimated Cancer Incidence per gender and per site in 2018

Cancer incidence varies very much per gender. Most frequent cancers are lung, prostate and colorectum cancers for male, and breast, colorectum and Lung cancers for female. Almost 1 out of 4 new cases of cancer for female is breast cancer. For both gender, 7 cancer type represent almost 7 cancer incidence out of 10. Among them, only 3 are common between the two genders: Lung, colorectum and stomach. Others particularly include specific-gender cancer sites (prostate for male, breast, cervix uteri and corpus uteri for female), which represent 13.5% of cancer incidence for male and 35.2% for female.

Therefore, the differences in cancer sites per gender require to take into account the gender in the understanding and the analysis for cancer incidence.

[2] http://ci5.iarc.fr/CI5-XI/

Incidence rates using ICD classification are also available in the WHO data. It is illustrated in the following figure[3], in which cancer sites are colored depending the annual incidence rate at world level from the color white to a null incidence rate to the color red for a maximal incidence rate (of 55.2 reached by breast cancer for female). The full table is available in appendix A.

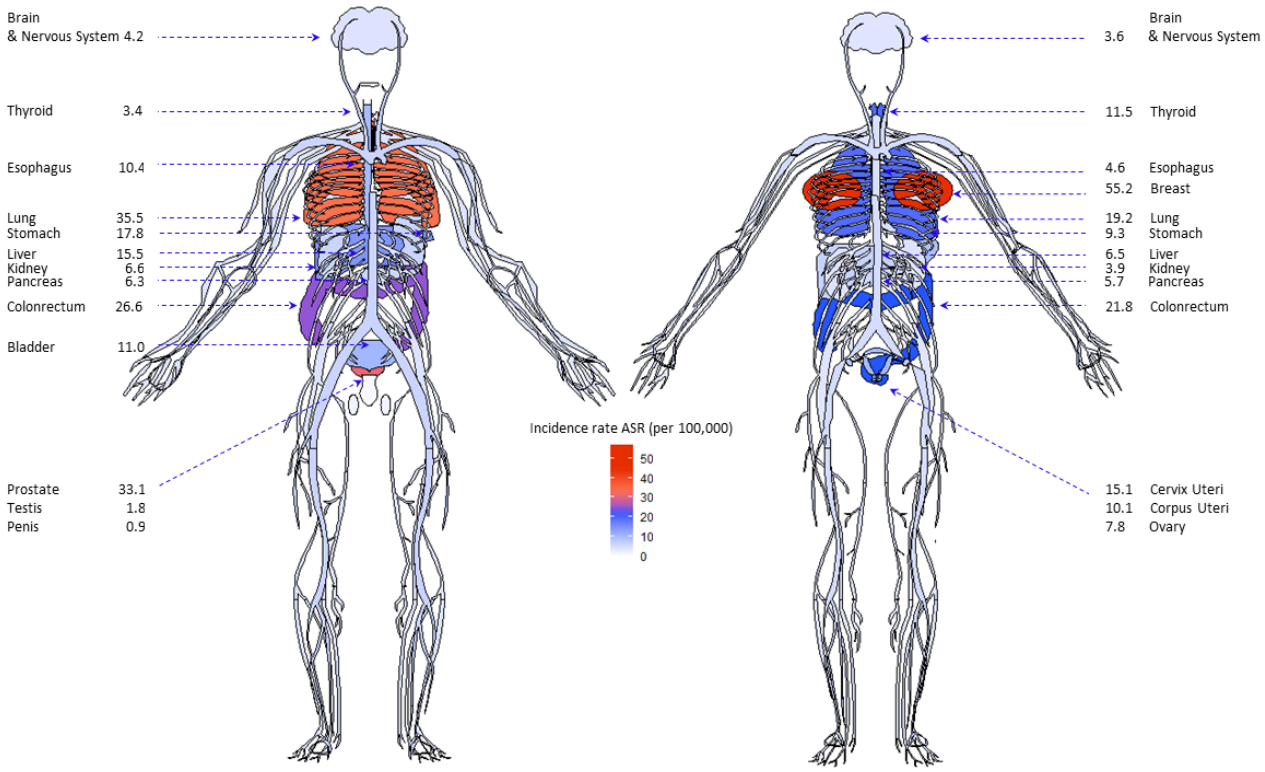


Figure 21: Estimated Cancer Incidence per gender and per site in 2018

This figure shows close results of the previous figure with the number cases. It measures the average incidence rate for population with the characteristics of the world population. Largest incidence rates are breast cancer for female, with an incidence of 55.2 per 100,00 and prostate cancer for male, with an incidence of 33.1 per 100,000. Lung & stomach cancer incidences strongly differs per gender (35.5 and 17.8 for male, versus 19.2 and 9.3 for female) whereas the difference is smaller for colonrectum (26.6 versus 21.8).

Despite that the age-standardized rates are a useful tool to summarize cancer incidence rates for a population, the cancer incidence distribution per age or per age group would be preferred. Such data will be presented and used in the next chapters, with data from Korea and Japan.

[3]Figure generated with the *R* package *gganatogram*

1.2.3 Additional Classifications

TNM Classification

The TNM classification[4] (TNM) is a classification which focuses of the stage of cancer using 3 components: the Tumor (T), the Nodes (N) and the Metastasis (M):

- T describes the primary tumor size and invasion of nearby tissue, with 4 levels (1 to 4).
- N describes nearby lymph nodes[5] that are involved, with 4 different values (0 to 3). N0 denotes no nearby lymph nodes involved, N1 1 or 2, N2 3 to 6 and N3 more then 7 lymph nodes involved.
- M describes the presence of distant metastasis, that is the spread of cancer from the primary location to another part of the body, with 2 different values (0 and 1). M0 denotes absence of distant metastasis, and M1 the presence.

Classification varies depending on cancer type. For example[6], a breast cancer tumor classified T2 N0 M0 denotes a single tumor 2.1 to 5 cm across, with no spread in the lymph nodes as well as no evidence of spread outside the breast. The TNM is illustrated below with esophageal cancer[7].

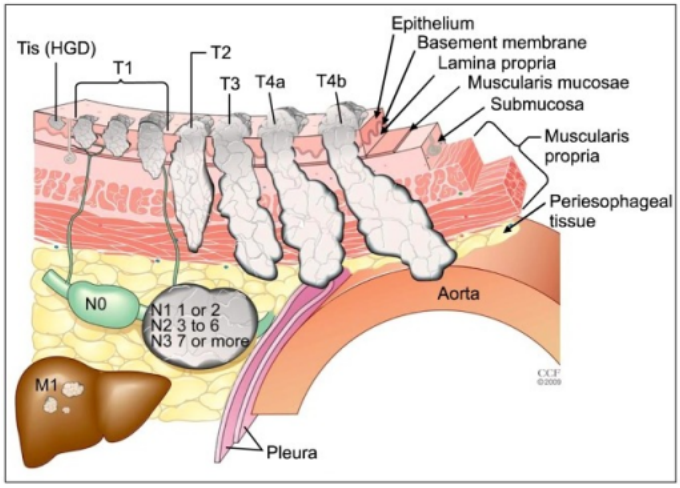


Figure 22: Esophageal Cancer Staging with TNM Classification

FIGO staging of Cervical carcinomas This classification is based on 5 stages. Stage 0 denotes carcinoma in situ, Stage 1 is confined to the initial organ, Stage 2 does not extend into the pelvic wall whereas stage 3 does, and stage 4 beyond the true pelvis.

Colorectal cancer Duke system This classification is based on 4 stages, from A to D, depending the grow on the muscle (if no stage A, else B) and the spread on lymph nodes (stage C) and other organs (stage D).

[4] `https://www.uicc.org/resources/tnm`
[5] Lymph nodes are part of immune system and creates B cells and T cells
[6] `https://breast-cancer.ca/tnm-class/`
[7] Figure from T. W. Rice, *Korean J Thorac Cardiovasc Surg*, 2015 Jun;48(3):157-63.

1.3 Cancer Causes

Numerous studies investigate cancer causes, that is the different factors which can provoke cancer or may favor cancer incidence, for example [5] and [6] for France or [7] for Asian Americans. Scope in terms of population and cancer site may vary. Main findings are as follows:

- **Gender**: cancer incidence is related to gender. While most cancers affects both gender, incidences strongly differ. Then, some cancers are gender-specific. For example prostate cancer affects only male and more generally all reproductive organs are gender-specific sites. Then, while breast cancer is the most common cancer for female, it is very rare form of cancer for male.

- **Age**: cancer incidence all causes increases with age. Cancer is generally related to advanced age[8], despite the diseases can occur at any age. Some cancer affects mostly young ages (such as leukemia, brain cancer). Some cancer affects mostly middle-aged adults (breast cancer for female). Incidence databases are usually given per gender and age (or age groups). Data presented in section 4 describes the variations in cancer incidence depending one's age.

- **Genetics**: cancer being a cell disease, cancer is by nature a genetic disease. It is caused by acquired gene mutation or inherited cancers. Acquired mutations denote gene mutation occurring after conception due to errors in cell division. Also, some faulty genes may be inherited and then cause cancer during one's lifetime. Though it is a strong belief that genetics and inherited genes are an important cancer cause, it may represent only 5% to 10% of cancer. Some specific genes have been related with higher risks of cancer. Especially two genes, named BRCA1 (BReast CAncer) and BRCA2 mutations are associated with a higher breast and ovarian cancer incidence due to differences in forming proteins. cases. While current tests may not be sufficient to detect all forms of mutations, the test result can both indicate predisposition and may impact the cancer treatment [8].

- **Infections**: the Human Papilloma Virus (HPV) infections are a family of infections, usually eliminated by the immune system. If the infection is latent, two types of HPV, named HPV-16 and HPV-18 are a common cause of cervical cancer. Other types can favor cancer, especially genital and oral cancer. Infection with Helicobacter pylori (H pylori) bacterial increases the risk of ulcer and stomach cancer.

- **Environment**: exposition to specific substances increases cancer risk. Arsenic is a source of different cancer types, especially skin, lung and bladder cancer. Exposition to Asbestos is an important risk factor for larynx and lung cancer. Polycyclic Aromatic Hydrocarbons (PAHs), which are elements resulting from combustion of organic products, is also related to a large variety of cancer types. Inhalation of Silica, produced from mining or heavy industry activities is a known cause of lung cancer. More generally, pollution (both urban air pollution and indoor pollution) affects cancer incidence.

[8] https://www.cancer.gov/about-cancer/causes-prevention/risk/age

- **Behavior**: Behavioral sources of cancer risk increase are numerous [5]:
  - Smoking (including passive smoking): increases lung, larynx & digestive organs incidence cancer due to inhalation of carcinogens, including PAHs, nitrosamines, aromatic amines and metals. It is the main cause of lung cancer. Studies show that 9 out 10 lung cancers in the US[9] and 7 out 10 in the UK[10] is due to smoking. People who smoke are about 25 times more likely to develop lung cancer than people who do not smoke.
  - Alcohol: tends to increase digestive organs & breast cancer incidence due to ethanol. As smoking, carcinogens damage cells and can cause changes in DNA replication. The current Dietary Guidelines for Americans[11] recommend not to start drinking for people who do not drink and to not drink more than one drink a day for women, and no more than two drinks a day for men.
  - Nutrition: high consumption of meat (especially red meat) tends to increase stomach & colon cancer. Salted and spicy food high consumption tends to increase gastric cancers. Oppositely, high consumption of fruits & vegetables tends to reduce cancer incidence, especially for digestive cancer. More generally, a varied diet is recommended to prevent cancer and stay healthy.
  - Excess of sun exposition increase skin cancer incidence, especially without sun protection, due to ultraviolet radiation.
  - Insufficient physical activities and sedentary lifestyle tend to increase cancer risk. The WHO general recommendation for physical activities are to perform at least 150 minutes a week of moderate-intensity aerobic physical activity [12] (or 75 minutes if vigorous).
  - Overweight, which can be defined as a BMI[13] greater then 25, and obesity tend to increase cancer risk.
- **Health Policy**: efficient medical system may increase the general health condition and may decrease cancer incidence, while screening programmes may increase cancer detection (but not incidence itself). Cancer awareness encourage cancer early detection and may increase estimated cancer incidence at short-term, but may change population's behaviors and then hopefully reduce both cancer incidence at long-term.

Due to usually multiple causes to cancer for a given individual, it is uneasy to correctly measure the importance of each risk factor. The different risk factors above constitute a base of cancer awareness policies.

[9] https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/pdfs/fs_smoking_cancer_508.pdf

[10] https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/smoking-and-cancer/how-does-smoking-cause-cancer

[11] https://health.gov/dietaryguidelines/2015/guidelines/appendix-9/

[12] https://www.who.int/dietphysicalactivity/factsheet_adults/en/

[13] Body Mass Index (BMI) is equal to $\frac{weight}{height^2}$, expressed in kg/m^2

1.4 Cancer Detection

1.4.1 Techniques

Cancer may be detected following some symptoms from a patient, or from tests, named screening, to a population without symptoms. Early signs of cancer may vary with cancer type. It generally includes pain, fatigue, unexplained weight loss, skin change (skin cancer), lumps (breast and testicular cancer), sores that fail to heal (skin or oral cancer), abnormal bleeding (especially lung and digestive cancer), persistent indigestion (digestive cancer), and chronic hoarseness (lung, larynx and thyroid cancer)[14] [15]. Though, those symptoms may also be caused by other diseases, and both patient medical history and patient examination are required to diagnosis cancer.

Cancer detection tests include blood test, biopsy and imaging (CT scan, MRI, mammography). In a biospsy, the removed tissue are examined under a microscope. CT (Computed Tomography) scans use X-rays to produce cross-sectional images of the patient. Like other imaging techniques, the examination focuses on a designed part of the patient body. MRI (Magnetic Resonance Imaging) scans are longer exams that use magnetic fields and radiowaves pulses to produce body images. PET (Positron Emission Tomography) scan is an additional technique with use radioactive materials, called radiotracers, injected in the patient body before scanning. Nuclear materials accumulate in tumors and help to identify a cancer tumor when imaging analysis. CT scan and PET CT are illustrated below[16].

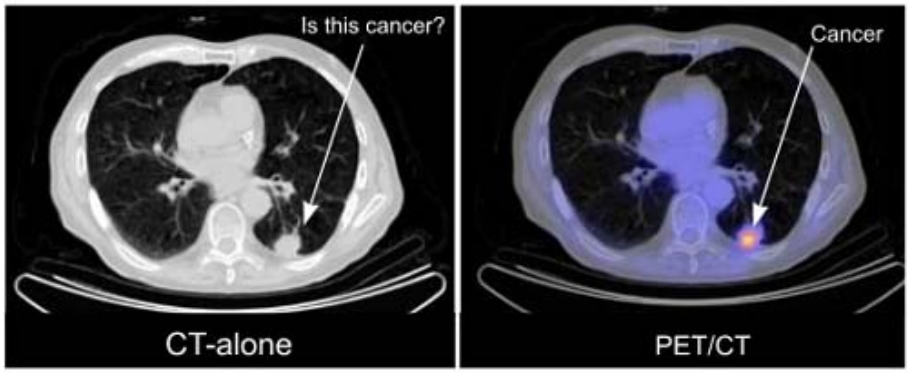


Figure 23: Lung Cancer detection with CT and PET scanning

Some promising blood tests aim to detect cancers. Blood tests focus on DNA methylation, that is additional cancer-related elements on DNA. While recent publications show promising results, they rely on cancer diagnosed populations and look retrospectively on previous blood samples to find specific DNA methylation changes [9][10]. Such tests also aim to detect the cancer site in case of detected cancer and to evaluate how many years before cancer diagnosis DNA methylation can be detected. A large-scale prospective programme involving 165,000 people including 140,000 with no symptoms will start mid-2021 in the UK[17]. Prospective blood test methods would be a milestone for cancer detection, as a non-invasive technique for early detection of multiple cancer sites.

[14] https://www.who.int/cancer/detection/en/
[15] https://www.cancer.org/cancer/cancer-basics/signs-and-symptoms-of-cancer.html
[16] https://petctberkeley.com/patient/lungcancer/
[17] https://www.england.nhs.uk/2020/11/nhs-to-pilot-potentially-revolutionary-blood-test/

1.4.2 Screening Programmes

Due to the heavy burden of cancer, some countries have launched detection programmes. The sooner a cancer is detected the better the treatment can be efficient.

Screening Criteria

Following [11], the standard screening criteria in the literature are as follows (Wilson and Jungner classic screening criteria):

1. The condition sought should be an important health problem
2. There should be an accepted treatment for patients with recognized disease
3. Facilities for diagnosis and treatment should be available
4. There should be a recognizable latent or early symptomatic stage
5. There should be a suitable test or examination
6. The test should be acceptable to the population
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood
8. There should be an agreed policy on whom to treat as patients
9. The cost of case-finding (diagnosis and treatment) should be economically balanced in relation to possible expenditure on medical care as a whole
10. Case-finding should be a continuing process and not "once and for all"

The following sub-sections give more details about existing screening programmes in countries used in this dissertation: France, Korea and Japan.

France

Three screening programmes are currently in-force in France:

Site	Offered from	Target Population	Frequency
Breast	2006[18]	women from age 50 to 74	2 years
Colorectal	2006[19]	from age 50 to 74	2 years
Cervical	2018[20]	from age 25 to 65	3 years

Screenings techniques are mammography for breast cancer, fecal immunological test for colorectal cancer and Papanicolaou test (also named pap test) for Cervical cancer. Mammography is an X-ray of the breast. Pap test is a medical exam in which cells are collected from the cervix, and then analyzed to detect abnormal cells. Fecal immunological tests are tests done from self-collected stool samples sent to a lab for analysis.

Screening participation rates reach 50.3% for breast cancer in 2018, 32.1% for colorectal cancer in 2018-2019 and 59.5% for cervical cancer in 2015-2018. The positive rate for colonrectal cancer is 4.0% (4.9% for male, 3.2 % for female)[21].

[21] http://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers

Korea

In Korea, 5 cancer screening programmes are in-force[22]: stomach, breast, cervical, liver and colorectal. They were initially offered to the beneficiaries of Medical Aid[23][12]. Then, the target population was expanded and included individuals of the lower 20% of National Health Insurance beneficiaries in 2002, 30% from 2003 and 50% from 2005.

Site	Offered from	Target Population	Frequency
Stomach	1999	from age 40	2 years
Breast	1999	women from age 40	2 years
Cervix	1999	women from age 30	1 year
Liver	2003	high-risk groups from age 40	2 years
Colon	2004	from age 50	1 year

Screenings techniques are upper gastrointestinal x-rays exam for stomach cancer, mammography for breast cancer, pap test for cervical cancer, abdominal ultrasonography and blood biomarkers for liver cancer and fecal immunological test for colorectal cancer.

Participation rates for stomach, colorectal, breast and cervical screening cancers have increased year by year and reached 47.3%, 25.0%, 51.9%, 40.9% in 2012, respectively [13].

Japan

In Japan, 6 cancer screening programmes are currently in-force: stomach and cervix (started in 1980), lung, breast and corpus uteri (from 1987) and then colonrectum (from 1992) [14].

Site	Offered from	Target Population	Frequency
Stomach	1983	from age 40	1 year (6 months from 2015)
Cervix	1983	women from age 40 (from age 30 from 1998) (from age 20 from 2004)	1 year (6 months from 2004)
Lung	1987	from age 40	2 years
Breast	1987	women from age 40 (from age 30 from 1998) (from 40 from 2004)	1 year (6 months from 2004)
Corpus Uteri	1987	women from age 30	1 year
Colorectal	1992	from age 50	1 year

Screening rates have increased from 2007 to 2016 and reached 46.4% (males) and 35.6% (females) for stomach cancer, 44.5% (males) and 38.5% (females) for colorectal cancer, 51.0% (males) and 41.7% (females) for lung cancer, 44.9% and 42.3% for breast and uterine cancer (females only) in 2016 [15]. It however hides some heterogeneity per region both for screening levels and trends, with for example for breast cancer in 2016 a screening rate of 61.2% in Yamagata prefecture (+7.4 points from 2010) and only 36.1% in Yamaguchi prefecture (+3.1 points from 2010).

[22] `https://ncc.re.kr/main.ncc?uri=english/sub04_ControlPrograms03`

[23] Medical Aid covered 3.7% of the population in 2006. National Health Insurance covered 96.3% of the population in 2006.

1.4.3 Cancer Registries

Cancer registries collect data regarding cancer cases. The European Network of Cancer Registries (ENCR) publishes recommendations for data to be collected[24]. The essential data to be collected concerns both individual and tumor information. About individual, it is the personal identification, date of birth, sex, ethnic group, vital status and last follow-up date. About tumor, it is the incidence date, primary tumour site (according the ICD classification), behaviour, the basis of diagnosis, the stage, the initial therapy and the source of information.

Registries existence and quality vary by country. Population based cancer registries[25] have a good quality in North America, Australia and some European countries. In Europe, quality varies by country as well as by areas. The evolution of the existence of registries is presented below:

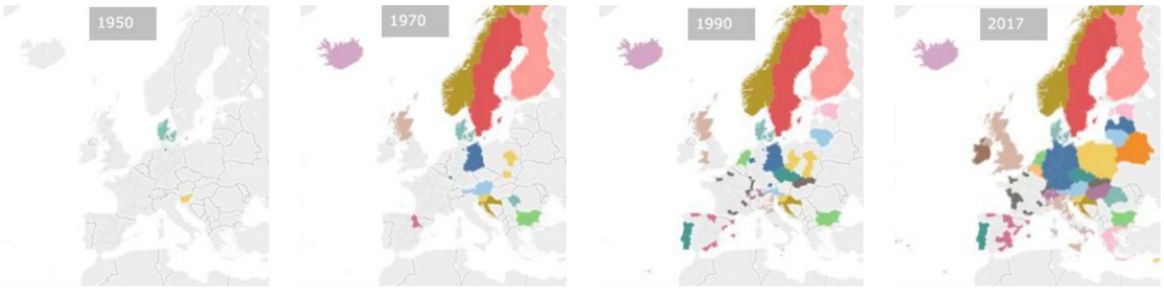


Figure 24: Cancer registries in Europe from 1950 to 2017

In France, cancer registries exist in only 22 departments (out of 100 in 2019) with geographic disparity. For example, there is no cancer registry in Paris area[26]. Quality of registries are poor or incomplete in a lot of countries in South America, South Asia and Africa.

Cancer registration started in Korea on a voluntary basis in 1975. The first official programme was the Korea Central Cancer Registry (KCCR), launched in 1980 through a network of 47 hospitals. Currently more than 190 hospitals are involved in the programme, which covers over 90% of cancer cases in Korea[27].

Japan has a long history in cancer databases, with the first registry established in 1951 in Miyagi Prefecture. From 2016, the *Act on Promotion of Cancer Registries* enforces hospital managers to report information on any primary cancer that was first diagnosed in their institutions to the prefectural government. There is then one registry per prefecture, that is 47 prefectural cancer registries.

[24] https://www.encr.eu/sites/default/files/pdf/recommendations.pdf

[25] http://canceratlas.cancer.org/assets/uploads/2014/10/31_CancerRegistries_rev_2.pdf

[26] https://www.lemonde.fr/sciences/article/2019/01/22/cancers-aucune-donnee-pour-78-de-la-population-francaise_5412764_1650684.html

[27] https://ncc.re.kr/main.ncc?uri=english/sub04_ControlPrograms02

1.5 Treating Cancer

1.5.1 Treatment Types

This section details existing general treatments for cancer. For example, among patients diagnosed with cancer in England in 2013-2015, 45% experienced surgery, 29% experienced chemotherapy and 28% radiation therapy.[28] A treatment generally imply multiple treatments types[29].

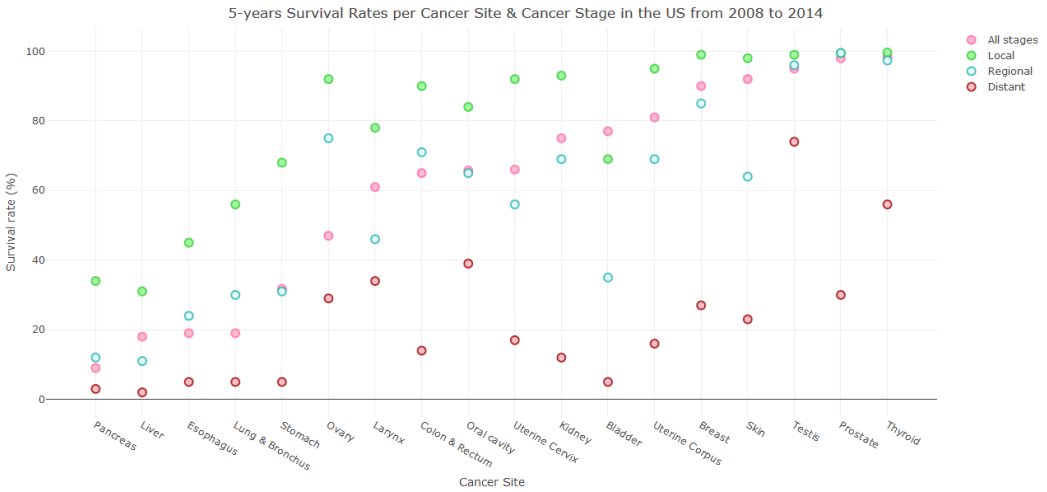
- **Surgery**: Surgery is used to remove a part or a entire solid tumor. Surgery was originally the main treatment for cancer [16]. Surgery may involve some techniques to remove tumors minimizing pain or damages, including cryosurgery or laser use. It is especially used for breast, gastric, lung and skin cancers. Aside surgery for curative purposes, surgery may be used at diagnosis stage (removing body cells for biopsies), at preventive stage (removing cells which may develop cancer) as well as reconstruction stage (for example breast reconstruction).
- **Chemotherapy**: Chemotherapy denotes regular treatments using drugs containing cytotoxic agents designed to stop cancer cells proliferation or kill cells. Cancer cells dividing more often than normal cells, chemotherapy is more likely to kill them. Chemotherapy is usually done using intravenous administration. Oral medication is a less frequent mode. Both produce adverse side effects which cause hair loss, fatigue and gastric disturbance.
- **Radiation therapy**: Ionizing radiation is radiation which remove electrons from the orbit of an atom in the cells (causing the atom to be ionized) [17]. Then, the damaged cells may be killed or simply repaired with time. Radiation has effects on both healthy and cancer cells. Despite that the goal of radiation therapy is to expose cancer cells to radiation, healthy cells are also exposed. However, healthy cells have a better recovery capacity than cancer cells, which then are killed. The machine used for such treatments is a linear accelerator (LINAC) device. Radiation therapy is used in approximately 50% of cancer treatments, such as skin, cervix, lung cancers, breast, rectum, prostate, bladder and advanced cancers.
- **Immunotherapy**: Immunotherapy, also called biological therapy, is a set of techniques designed to enhance the immune system of a patient to fight cancer. It includes cancer vaccines, used for fighting Human Papilloma Virus (HPV) which may cause cervical, vaginal, vulvar cancer and anal cancer.
- **Stem Cell Transplant**: Transplant allows to replace damaged stem cells and then grow into healthy cells. Such treatments may be used to cure leukemia, lymphoma and myeloma. It requires a donor who must be a good match, that is the donor stem cells match the patient stem cells.

Cancer treatments may involve physical changes (tumor removal, loss of hair, fatigue, loss of weight) and stress. Patient support (especially from medical staff and family members), balanced diet, physical activities are also important factors during treatment phase.

[28] http://www.ncin.org.uk/cancer_type_and_topic_specific_work/topic_specific_work/main_cancer_treatments
[29] https://www.cancer.gov/about-cancer/treatment/types

1.5.2 Survival Cancer

Cancer is a leading cause of death in many countries. Survival statistics are used to assess the survival of cancer diagnosed people. Survival cancer statistics are calculated by cancer type, usually per cancer stage, at a defined time horizon (usually 5 years). In the past decades, survival rates have significantly risen. In the US, 5-years survival rates have improved from 49% in 1975-1977 to 69% in 2008-2014 [18]. The following figure present the survival rates in the US in 2008-2014 per cancer site and stage (local[30], regional[31] and distant[32])



Four cancer types have a very low 5-year survival rates: pancreas (9%), liver (19%), esophagus and lung (21%) cancer. Oppositely, prostate (99%), thyroid (98%) and testis cancer (97%) have a very high 5-year survival rates, which demonstrate strong differences between cancer sites. Survival decreases sharply with the spread of the disease, and survival rates for distant cancer are low (and close to null in many sites) for all cancer sites.

Survival rates have generally evolved positively among time. International comparisons in 7 developed countries from 1995 to 2014 show that 5-years survival rates have evolved positively for colon, rectum, stomach and ovarian cancer, whereas improvements are very small for pancreas and esophagus cancer [19], which remain very severe. Some cancer are classified as cancer of unknown primary, which code in ICD classification is C77-C80. Such cancers are advanced cancers with metastasis for which it is not possible to identify the primary cancer site. Survival rates are very low for such types of cancers, with some studies having an median survival time of 33 days [20], some 2 to 3 months.

The public perception of survival rates may differ from epidemiological estimates. In Korea, 85% of the adults underestimate survival rates for cancer locations with high survival rates (prostate, breast, thyroid), whereas 50% overestimate survival rates for cancer locations with low survival rates (lung, liver) [21].

[30] Local: an invasive malignant cancer confined entirely to the organ of origin.

[31] Regional: a malignant cancer that 1) has extended beyond the limits of the organ of origin directly into surrounding organs or tissues and/or 2) involves regional lymph nodes

[32] Distant: a malignant cancer that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis.

1.6 Cost of Cancer

Cost of cancer is an increasing concern at both macroeconomic and individual level. In the US, costs of cancer care are estimated to be 125 billion$ in 2010 and 158 billion$ in 2020 [22]. In France, the macroeconomic cost is estimated to be 15.6 billion€ in 2017 (against 12.4 billion€ in 2012) [23] in the general health scheme, that is 11.1% of the total healthcare expenses.

Costs depend on cancer type, that is cancer site and cancer severity. The cost of care at the initial phase of care is estimated in the US to be 25,694$ for a bladder cancer, 72,639$ for lung cancer and 129,802$ for brain cancer. Aside the high cost of cancer treatments, there is a mismatch between the perception of such costs and real treatments costs. In France[33], 67% of the population under-estimate the cost of a chemotherapy for breast cancer, which costs between 5,200 and 31,200€. Especially, 24% believe that such treatment costs less than 500€. Some recent advances in immunotherapy have led to new drugs for which treatments exceed 100,000$, raising concerns on cancer treatment costs. Such costs, that the pharmaceutical industry justifies by the costs of years of research and development, lead to debates about the fair price of treatments and financial accessibility of treatments to the patients in need.

The existing coverages are usually incomplete and patients must face out-of-pocket payments (direct payments made by individuals to health care providers[34]) and face a financial burden fighting cancer. The out-of-pocket payments for one individual can be split in 3 types [24][35]:

1. Out-of-pocket medical expenses: consultations, medical examinations, blood tests, drugs, surgeries or hospitalization fees, depending the coverage from the national health services or private health insurances.
2. Out-of-pocket payments related to a particular cancer type: breast implant for breast cancer, colostomy bags for colorectal cancer, incontinence protection products for prostate cancer.
3. Out-of-pocket payments that may arise for all cancer types: transportation costs, childcare, wigs to face hair loss, skincare, dietary supplements, dental complications due to chemotherapy or radiation therapy to the head and neck.

Cancer diagnosis may also force an individual to stop partially or completely his current professional activity. Treatments may require some time off, due to hospitalization or rest, and some treatments make patients unable to work or may face limitations at work. This implies a productivity loss at macroeconomic level and can lead to a decrease of income at individual level.

For all these reasons, depending the cancer awareness and the risk averse preference, the healthcare services, healthcare coverages and income protection schemes, cancer insurance can meet societal and individual needs.

[33] https://curie.fr/page/observatoire-cancer-2017-couts-des-traitements-innovants-en-cancerologie

[34] https://www.who.int/health_financing/topics/financial-protection/out-of-pocket-payments/en/

[35] The reference deals with cancer out-of-pocket payments in France, where cancer-related medical expenses are reimbursed by the national health service based on statutory fees. The reference presents only two types of out-of-pocket payments (types 2 and 3 below), and includes extra-medical expenses in the type 3 below. In order to generalize this tyopology for other health systems, the first type about medical expenses is isolated compared to the reference

2 Cancer Insurance

2.1 Insurance Solutions regarding Cancer

Insurance solutions regarding cancer can be classified as follows:

- **Medical Expenses Policy**: the benefit is based on paid medical expenses exceeding the national health service coverage. Benefits can be proportional to the medical expenses and include deductibles or a payment limit. Covered events may include payments for medical examination, drugs, medical transportation and hospitalization expenses. Such policies usually cover all types of diseases, including cancer. Some can be specially designed to cover only medical expenses from cancer.
- **Lump-sum Policy** : the benefit is a lump-sum at cancer diagnosis. The benefit may vary depending cancer site. The benefit is paid to the insured independently of possible medical expenses. Because the benefit is a sum paid to the insured, it may be used for both medical and non-medical costs. Many policies stop after the first cancer diagnosis.

In this dissertation, cancer insurance denotes lump-sum products. The product examples and the market description hereafter discuss products for which the main coverage is lump-sum in case of cancer diagnosis. However, the study of cancer incidence rates can be used in all types of cancer insurance products.

2.2 Cancer Insurance & Critical Illness Insurance

Critical Illness (CI) products aim to cover dread diseases as an accelerated benefit of a death benefit or an additional benefit, as a rider or a stand-alone product [25]. The first critical illness policy, launched by Crusader Life in 1983 in South Africa covered 4 diseases: heart attack, cancer, stroke and coronary artery surgery [26]. Such products have then spread worldwide and coverage has been extended to a larger number of diseases [27].

Cancer coverage is usually one of the most important coverage in a CI policy, due to the relative important incidence rate. Some policies include a death benefit. In that case, CI benefits are sometimes considered as an early payment of a portion of the death benefit. Such coverages are called accelerated benefits. The final death benefit is then the amount deduced from the accelerated benefits previously paid.

Cancer insurance has actually preceded CI insurance, with the first policies being introduced by American Family Life Insurance Company of Columbus (now named AFLAC)[36] in 1958 in the US. They then introduced cancer insurance in 1974 in Japan [28]. Cancer products have globally expanded, with an introduction in Korea and Taiwan in the 1980's [37] and in India in 1985 [38].

[36] https://www.aflac.com/about-aflac/our-company/our-history.aspx
[37] https://www.rgare.com/docs/default-source/newsletters-articles/cancer_asia.pdf
[38] http://www.indiancancersociety.org/what-do-we-do/cancer-insurace-schemes.aspx

2.3 Products

Cancer insurance products may cover all cancers or only specified cancer types. Benefits may differ per cancer sites as well as additional benefits (hospitalization) may be included in some products. This section details two products to illustrate cancer insurance products:

Example 1: 6180실버 암보험 (6180 Silver Cancer)[39]
This insurance product is from BNP Paribas Cardif in Korea.
Main Benefit: lump sum in case of cancer diagnosis. Lump sum depends on cancer type. Maximum of one benefit per cancer type
Main Benefit Range: from 5,000,000 to 10,000,000 wons (from 4,000 to 20,000€), varying per cancer type. Benefit is reduced by 50% during the first 2 years.
Maturity Benefit: 10% or 20% of the sum assured if no claim at maturity
Age at Inception: from 61 to 80 years
Policy Term: 10 years
Payment Term: same as policy term
Premium Type & Frequency: constant Monthly Premium

The main benefit is limited compared to a total possible cost of cancer treatment. Therefore this product (as well as the one hereafter) is designed to cover extra-costs due to a cancer diagnosis. Age at inception is limited to an aged population.

Example 2: 生きるためのがん保Days1 (Cancer insurance for Life - Day 1)[40]
This insurance product is from Aflac in Japan.
Main Benefit: lump sum in case of cancer diagnosis. Lump sum depends on cancer type. Maximum of one benefit per cancer type.
Main Benefit Range: 250,000 or 500,000 yens (2,000 or 4,000€)
Additional Benefit: Hospitalization day: 10,000 yens; Surgery Benefit: 200,000 yens; Radiotherapy Benefit: 200,000 yens
Product type: 3 coverage types for cancer diagnosis are available:
(a) Plain type: main benefit is constant during policy term
(b) 2-steps: main benefit is reduced by 50% after age 60
(c) Age 60: policy term stops at age 60
Age at Inception: 0-85 for type (a), 0-55 for type (b), 0-50 for type (c)
Policy Term: Whole Life for types (a) & (b), age 60 for type (c)
Payment Term: same as policy term
Premium Type & Frequency: Constant Monthly Premium

This product can offer a whole life cover. Main benefit is reduced compare to the the first product, but this product offers additional benefits following cancer diagnosis depending the duration of hospitalization and treatments types.

The next section, about markets, also details some products features.

[39] https://www.cardif.co.kr/web/main/product/medical-protection/6180-silver-cancer
[40] https://www.aflac.co.jp/gan/days1/

2.4 Markets

No statistical public source compares cancer insurance markets. This section details some characteristics of French, Korean and Japanese markets. The discussion about the Korean market relies on an analysis of cancer insurance products' disclosure information.

2.4.1 France

Cancer medical expenses, if recognized as long-term illness (*affection de longue durée*) are reimbursed on the basis of the reference price of medical care and drugs[41][42]. The national health service does not reimburse charges exceeding the statutory fees and a portion of hospitalization fees. Additional health insurance plans may waive those fees totally or partially. The out-of-pocket payments due to a cancer diagnosis have been estimated by the patients as 817€ in 2008 [24].

There is no wide market for cancer insurance products in France, oppositely to Korea and Japan detailed in the next sections. Some of the existing products in France are as follows:

- Allianz offers the product named *Assurance Prévoyance*[43], which is a comprehensive protection product. The product may include a lump sum benefit in case of cancer. Entry age range is from 18 to 62 years old, and lump sum benefit may vary between 7,500 and 125,000€. Waiting period is 6 months, and some cancer exclusions apply on skin cancer.

- La Banque Postale offers a product named *Assurance Coups Durs Santé* [44] which covers cancer and additional critical illnesses. Entry age range is from 18 to 65 years old. Benefit is a lump sum from 5,000 to 30,000€ in case of stroke, myocardial infarction or cancer diagnosis, with additional assistance benefits. Waiting period is 3 months. Some cancer exclusions apply, especially in case of skin cancer, cancer in situ and thyroid cancer.

- Malakoff Humanis offers the product *Prévoyance Cancer*[45] whose benefit is a lump sum from 5,000 to 30,000€, chosen at policy inception, with additional benefit for assistance. Entry age range is from 18 to 66 years old and coverage stops at age 70. Waiting period is 6 months and a cancer diagnosis during the waiting terminates the policy. Premiums are refunded in that case.

- Metlife launched in 2011 a product named *Serena*[46] for women between 18 and 60 years old. Benefit is a lump sum, with an amount between 10,000 and 70,000€ in case of breast or cervix uteri cancer diagnosis, with additional benefit for assistance and cancer surgeries.

[41] https://www.service-public.fr/particuliers/vosdroits/F34068

[42] https://www.e-cancer.fr/Patients-et-proches/Demarches-sociales/Prise-en-charge-financiere

[43] https://www.allianz.fr/prevoyance-dependance/prevoyance/

[44] https://www.labanquepostale.fr/particulier/produits/assurances_prevoyance/Assurance-Coups-Durs-Sante.avantages.html

[45] http://www.malakoffmederic.com/particuliers/prevoyance/solutions-prevoyance/prevoyance-cancer/index.jsp

[46] https://www.argusdelassurance.com/marches/produits-services/couvrir-les-cancers-feminins-sans-tabou.55989

2.4.2 Korea

The public Korean health system is based on two programmes: the Medical Aid (to the lowest incomes), and the National Health Insurance (NHI) which beneficiaries are respectively 3.7% and 96.3% of the population in 2006. The system is based on co-payments. Patients share 20% of the cost of insured inpatient care services and 30% in a Public Health Center [29]. However, the rate has been decreased to 5% in the 2000s for critical illness such as cancer.

Cancer is the leading cause of death in Korea. Out-of-pocket payments represent about 35% of total health expenditure in 2011. It is the second high rate in the OECD (after Switzerland); however there is no detailed statistics for cancer out-of-pocket payments.

Cancer insurance can be sold by both life and non-life companies in Korea. Detailed public data[47] about cancer insurance is available online. An extract of the database is presented in the appendix. The data is a list of cancer products and includes the company name, product name, distribution channel, entry age, possible policy and payment terms, a short description of the benefits and a premium amount example. Without distinction of the existence of maturity benefit and the distribution channel, 143 cancer insurance products are currently commercialized in April 2019 in Korea (98 through life insurers, 35 through non-life insurers). All life insurers (except one) are offering cancer insurance, with an insurer having up to 10 different products.

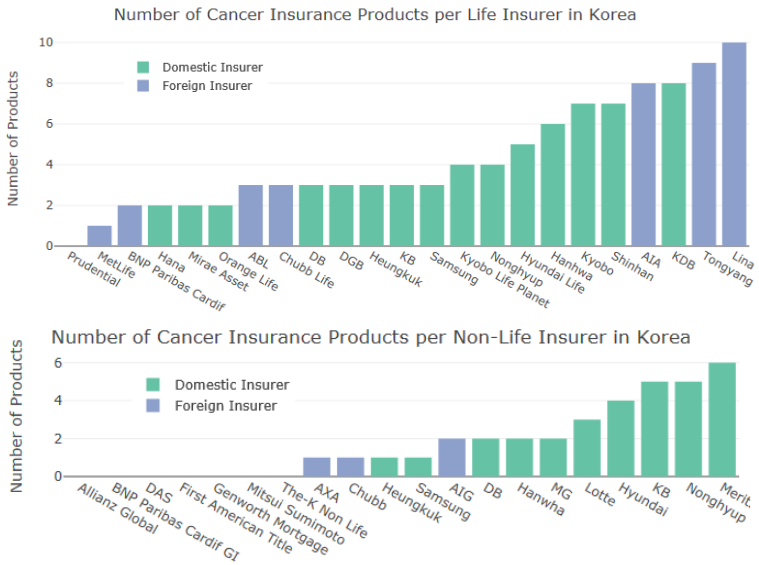


Figure 25: Cancer Insurance Products per Insurer in Korea

The insurer with the largest number of disclosed products is then Lina, which does not reflect the market share of this company and the others. The figure distinguishes domestic insurers (in green) and foreign insurers (in blue - subsidiaries of foreign-based insurer).

[47] `https://pub.insure.or.kr`, for Life insurers, accessed on April 15th, 2019
`http://kpub.knia.or.kr`, for Non-Life insurers, accessed on May 1st, 2019

The following graphs derived from the public disclosure data represent the minimal and maximal entry age (plain) and the maximum possible covered age (in dot) for each cancer insurance product in Korea in April 2019.

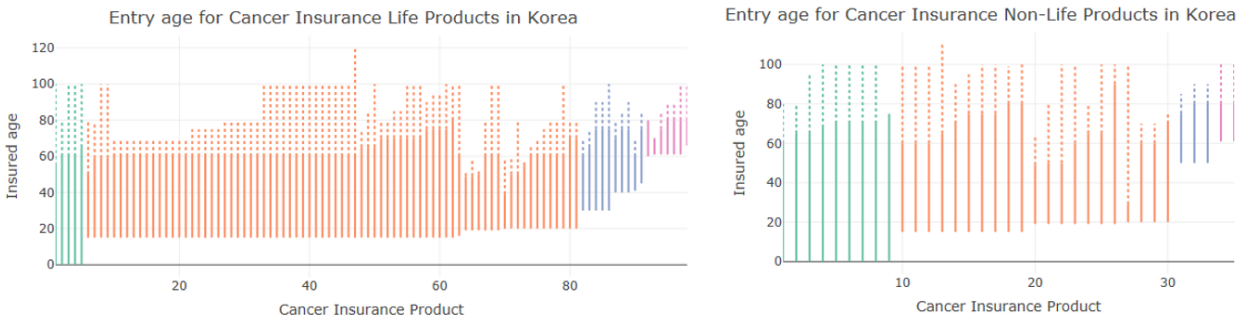


Figure 26: Minimal and maximal entry age for each cancer insurance product in Korea in March 2019, sorted by ascending minimal entry age

The figures must be read as follows: a vertical line represents one cancer insurance product. For example, the first (green) line in the left figure denotes one cancer insurance product from a life insurance company in Korea for which the inception age must be between 0 and 60 (plain line) and the coverage age can be up to 100 years old (dotted line).

The different colors in the graph group products per age range in 4 groups:

- in green, products for which a newborn can be insured
- in orange, products designed to adults, from aged 15 or 20
- in blue, products designed to middle adults only
- in pink, products designed for older ages, from age 60

Insured entry age varies from 0 to 80. 5 products can insure newborns. Most of products target people aged from 15 to 60. A particular market segment is the silver generation, with entry age from 60 to 80. Then, the insured can be covered until 100. One product is a whole life product, represented with a maximal cover age of 120 in the figure. Therefore, cancer incidence can be insured at (almost) all ages, and the study of cancer incidence for all ages make sense.

It is not easy to compare the sum assured in the different products. Some products have one sum assured for all cancer types. Some distinguish small cancer, middle-cancer and high-cost cancer. The benefit in case of a small cancer is usually 10% of the sum assured and such a claim does not stop the policy. The benefit in case of high-cost cancer diagnosed is usually 200% of the sum assured. Using the premiums illustrations and corresponding policy characteristics, the average sum assured is 22,000,000 wons, that is roughly 17,000€.

The different insurance associations do not publish some statistics about cancer insurance market in terms of number of policies or amount of premiums.

2.4.3 Japan

There are two main public insurance schemes in Japan: Employee's Health Insurance - for both public and private sectors - and National Health Insurance (NHI) - for self-employed or unemployed). The ratio of out-of-pocket in the NHI scheme is 30% for people younger than 75 years old, and then decreases at 10% [30]. Out-of-pocket payments represent in average about 11.7% of total health expenditure in 2014. Cancer is the leading cause of death.

The Life Insurance Association of Japan publishes annual statistics about the domestic insurance market, including cancer insurance. 24.5 million cancer policies are in-force in March 2018 [31]. 1.9 million policies were sold in the fiscal year 2018[48], that represents 13.4% of the insurance policies sold that year. The corresponding premiums amount is 1.8 trillion yen (13.7 million €)[49]. The market penetration in 2019 is 41.2%, almost doubled since 2001 (21.3%), leaded by private insurance companies (other actors being consumer's & agricultural cooperatives) [32].

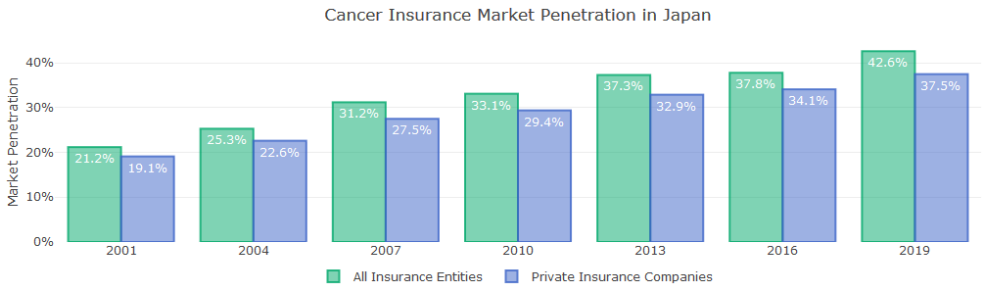


Figure 27: Evolution of Cancer Insurance Market Penetration in Japan

The Japanese cancer insurance market is dominated by the US-based insurer Aflac. Cancer insurance was introduced in 1974 by this insurer, which gained a 4-years monopoly, then extended for 4 more years. It is today the strong market leader with 15 million stand-alone cancer policies in-force [33], that is a 62.9% market share and 25% of the new policies. Aflac distribution strategy is based on two channels: agents (with 11,000 representatives) and partners (Japan Post, Insurers and Banks). Aflac sells two cancer products[50] including 生きるためのがん保Days1 (Cancer insurance for Life - Day 1), described in the section Products and a specific product for cancer survivors.

Japanese market for cancer insurance is recently diversifying, with Aflac in-force policies growing from 14 millions in 2020 to 15 millions to 2018 whereas other competitors total in-force policies has grown from 2.5 to 10 millions.

Cancer insurance products in Japan are characterized by multiple benefits: lump sum at cancer diagnosis, daily hospitalization benefits and benefits for surgery or other cancer treatments. Products are long term (Aflac sells whole-life cancer insurance products). Some products include benefits which differ per cancer stage[51].

[48] Fiscal year in Japan starts in April, 1st and ends in March, 31st
[49] Using the closing exchange rate at March 30th, 2018: 1 Yen = 0.00763 €
[50] as in October 2019
[51] https://www.manulife.co.jp/Kodawari-cancer2

2.5 A Cancer Insurance Pricing Model

This section describes the pricing of basic cancer product aimed to cover cancer diagnosis allowing specific benefits per cancer site. Notations are from [34] and [35], though they are adapted here to describe multiple cancer types. A first simple uniform cancer coverage is then discussed, and then a variation with an accelerated benefit for a subset of cancer types. Finally, a 3-type coverage product design is exposed.

2.5.1 Definitions

Assume the insured event is a cancer diagnosis of type j, and the corresponding benefit is a lump sum defined as a percentage α_j of a sum assured SA. The different insured events are assumed disjoints. Let x, $0 \leq x \leq \omega$, the insured age at policy inception and $q_{x+t}^{(j)}$ the incidence rate of the cancer type j of a live aged $x+t$. Policy term (expressed in years) is denoted by n and payment term by m ($m \leq n$).

One benefit can be claimed for each type and the policy can continue after some cancer diagnosis. We proceed hereafter by illustrations and then derive a theoretical framework. For the sake of simplicity, the different notations about insured age and time period are omitted in the following three illustrations, and discussed in the calculations section.

2.5.2 Illustration 1: Uniform Cancer Coverage

This simple product has one unique coverage. Product characteristics are as follows:

- Sum Assured SA: 10,000
- Cancer type A ($j = A$) and $\alpha_A = 100\%$
- Policy stops after cancer type A diagnosis

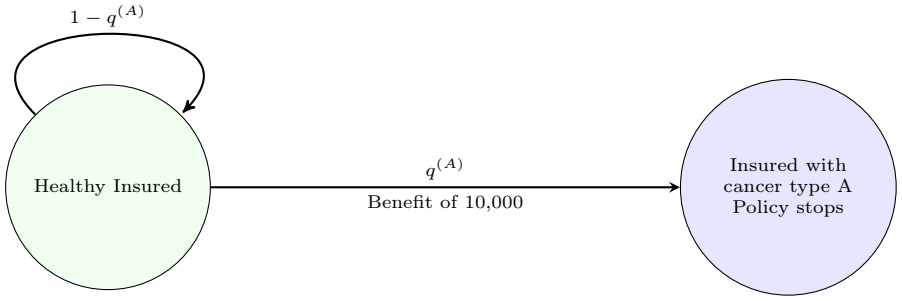


Figure 28: Uniform cancer coverage product transition diagram

Two possible product designs are then possible:

- Limited cancer coverage: cancer type A denotes one specific cancer, or a sub-set of possible cancers. Some cancer types are not insured through this product design.
- Comprehensive cancer coverage: cancer type A denotes all-causes cancer coverage. All types of cancer cause the same benefit.

While this product design is straightforward, it may not face insured needs because it leads to uniform benefits whatever the cancer site or cancer severity. A first refinement is to add another cancer type with a smaller benefit.

2.5.3 Illustration 2: Cancer Coverage with Accelerating Coverage

A limited benefit for cancer type B is added in this product design:

- Sum Assured SA: 10,000
- Accelerated coverage: cancer type A $\alpha_A = 10\%$ ($\alpha_A < \alpha_B$)
- Main coverage: cancer type B $\alpha_B = 100\%$
- Policy stops after cancer type B diagnosis
- The maximum total amount of benefits is capped by the sum assured

An healthy insured diagnosed with cancer type B will receive a benefit equal to the sum assured. An healthy insured diagnosed with cancer type A will received a first benefit. In that case, the policy continues and a second benefit is possible in case of cancer type B diagnosis. This latter benefit is reduced by the accelerated benefit already received.

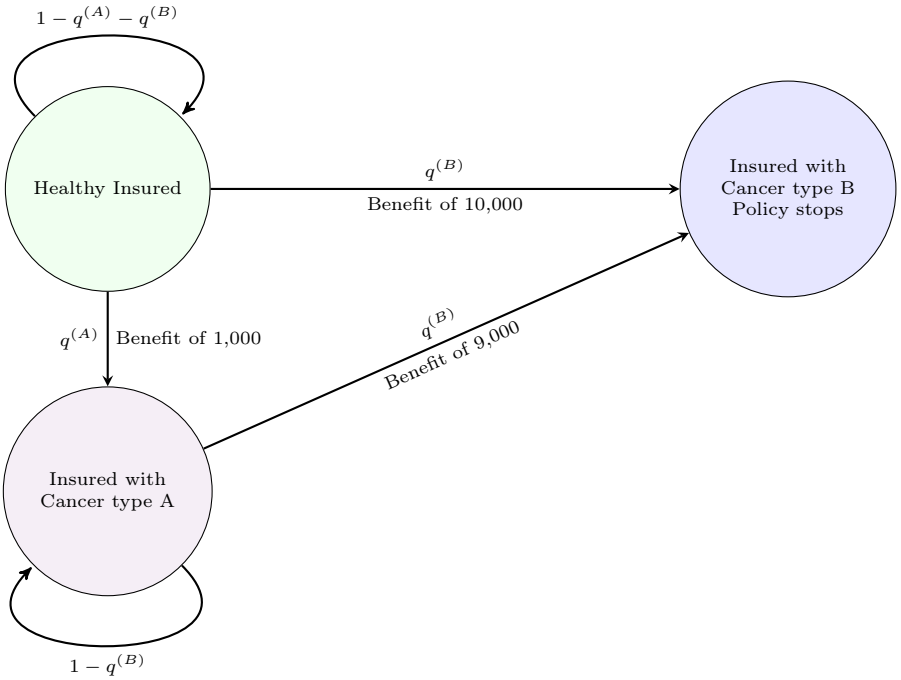


Figure 29: Cancer product with accelerating coverage transition diagram

In this transition diagram, it is assumed that the probability to get cancer type B is the same for healthy insured and insured diagnosed with cancer type A. The validity of this assumption must be assessed depending type A and type B specifications. However, some probabilities (such as the transition from diagnosed with type A to diagnosed with type B) may be hard to estimate due to no available public data and very limited experience data.

Again, the definition of cancer types A and B can lead to comprehensive or limited cancer coverages. This product allows to define two types of benefits, for example for *small* cancers and *general* cancers (all cancers types excluding *small* cancers). A wider range of benefits is however needed to segment coverages and corresponding benefits.

2.5.4 Illustration 3: Cancer Coverage with Nested Accelerated Coverages

An additional coverage, named cancer type C, is added in this product design, and it is this additional benefit which will terminate the policy. It allows to define a benefit for severe or costly cancers, while providing the possibility of comprehensive coverage:

- Sum Assured SA: 10,000
- Sub-accelerated coverage: cancer type A $\alpha_A = 10\%$ ($\alpha_A < \alpha_B$)
- Accelerated coverage: cancer type B $\alpha_B = 100\%$
- Main coverage: cancer type C $\alpha_C = 200\%$ ($\alpha_C > \alpha_B$)
- Policy stops after cancer type C diagnosis
- The maximum total amount of benefits is capped by the cancer type C benefit ($200\% \cdot SA$)

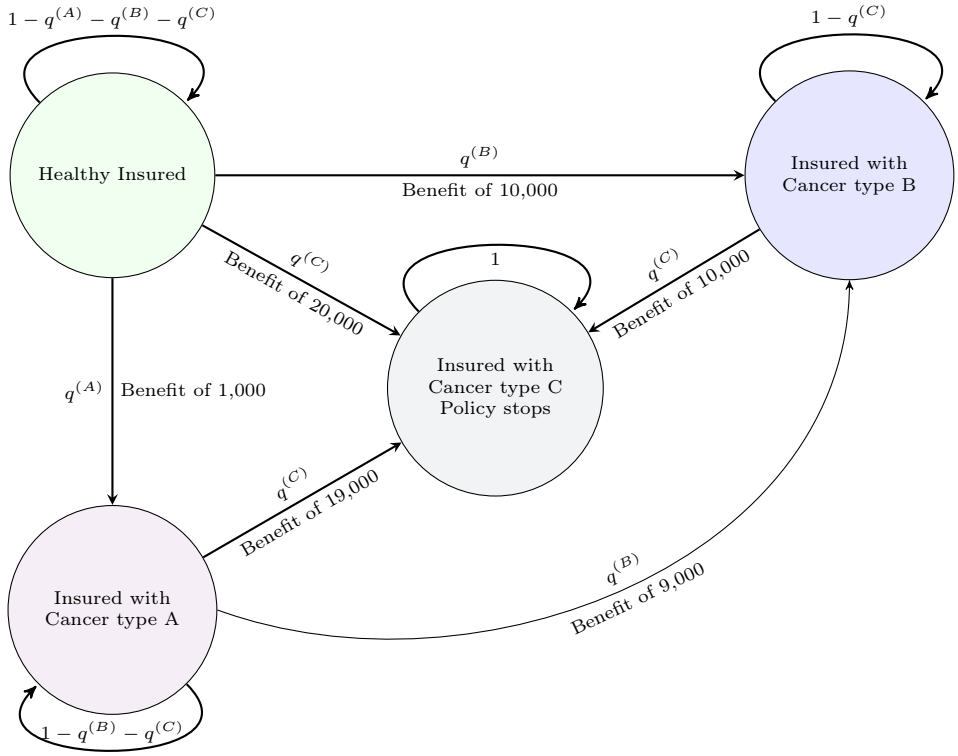


Figure 30: Cancer product with accelerating and additional cancer coverage transition diagram

These types of products allow more sophisticated product designs. An insured diagnosed cancer type C will receive a 20,000 benefit and the policy terminates. An insured first diagnosed cancer type A will receive a 1,000 benefit and remains covered in case of diagnosis of cancer types B or C. The insured will receive 9,000 in case of cancer type B diagnosis and eventually 10,000 in case of cancer type C afterwards. Again, the transition probabilities rely on the assumption that the probabilities transition to another state do not depend on current insured state. The next section generalizes the pricing of such products.

2.5.5 Basic Calculations

Let $l_{x+t}^{(j)}$ be the population aged x at policy inception without diagnosis of cancer type j until time t. A coverage for cancer all causes is denoted by $j = 0$. Let $q_{x+t}^{(j)}$ the incidence rate of cancer type j at age $x + t$ and $q_{x+t}^{(0)}$ the incidence rate of cancer all causes at age $x + t$. We assume that insured events are disjoints; for example the incidence rate of cancer all causes except skin cancer ($j = 1$) is $q_{x+t}^{(0)} - q_{x+t}^{(1)}$.

At policy inception, for all $j \in \{1; J\}$, $l_x^{(j)} = l_0$. For example $l_0 = 100,000$. Let T_j a dummy variable equals to 1 if the policy stops after this cancer type j diagnosis and 0 if not. Then $l_{x+t}^{(j)}$ evolves as follows: $l_{x+t}^{(j)} = l_{x+t-1}^{(j)} \times (1 - (q_{x+t}^{(0)} - \sum_{k=1}^{J} q_{x+t}^{(j)} \cdot T_j))$. The number of cancer cases of type j at time t is then $d_{x+t}^{(j)} = l_{x+t}^{(j)} \times q_{x+t}^{(j)}$.

Let i the technical interest rate and $\nu = 1/(1 + i)$. Assuming mid-year events, the related commutation number C is then $C_{x+t}^{(j)} = d_{x+t}^{(j)} \times \nu^{t-1/2}$ and commutation number M is $M_{x+t}^{(j)} = \sum_{k=0}^{k=\omega-x-t} C_{x+t+k}^{(j)}$ with ω the terminal age. The commutation number D is $D_{x+t}^{(j)} = l_{x+t}^{(j)} \times \nu^t$ and commutation number N is $N_{x+t}^{(j)} = \sum_{k=0}^{k=\omega-x-t} D_{x+t+k}^{(j)}$.

2.5.6 Single Pure Premium

The pure premium P is the premium component dedicated to face benefits. It is determined using the equivalence principle, that is the expected present value of pure premiums is equal to the expected present value of benefits. Recall that the benefit for cancer type j is the percentage α_j of the sum assured SA and n the policy term, the single pure premium is:

$$P = SA \cdot \sum_{j=1}^{J} \alpha_j \cdot \frac{(M_x^{(j)} - M_{x+n}^{(j)})}{D_x^{(0)}} \tag{1}$$

2.5.7 Single Gross Premium

The gross premium, denoted by GP, is the amount paid by the insured. It includes the additional following elements:

- Risk margin: the risk margin is assumed to be based on the pure risk premium using a risk margin rate, denoted by a.
- Loadings: the loadings rate captures sub-loadings rates, such as commission loadings, acquisition loadings, administration loadings, capital cost loadings or insurance tax rate. We assume that the loadings are expressed as a gross premium rate using a loadings rate denoted by l. The loadings rate l can be seen as the sum of L sub-loadings components l_i such as $l = \sum_{i=1}^{L} l_i$

The single gross premium is therefore:

$$GP = \frac{(1 + a)}{(1 - l)} \times P \tag{2}$$

2.5.8 Periodic Premium

The premium may be paid regularly, usually annually or monthly, during the payment term m. The expected present value of premiums to be paid must be assessed. In the case of all the insured pay an annual premium, then this amount is equal to $\frac{(N_x^{(0)} - N_{x+m}^{(0)})}{D_x^{(0)}} \cdot P^{(1)}$.

The product may include a premium waiver in case of adverse events, such as total and permanent disability or critical illness (cancer, stroke, heart attack). In that case, the insurer allows the insured to stop paying premiums. Then the insured population differs from the payment population. We consider in this model a premium waiver in case of specified cancer diagnosis.

Let W_j a dummy variable equals to 1 if the cancer type j diagnosis is a premium waiver event and 0 if not. Then the payment population l'_{x+t} evolves annually as follows: $l'_{x+t} = l'_{x+t-1} \times (1 - (q_{x+t}^{(0)} - \sum_{k=1}^{J} q_{x+t}^{(j)} \cdot T_j \cdot W_j))$[52]. That is, the payment population is reduced by the events that stop the policy or the events that allow a premium waiver. Analogously to the previous definitions, the corresponding commutation numbers are $D'_{x+t} = l'_{x+t} \times \nu^t$ and $N'_{x+t} = \sum_{k=0}^{k=\omega-x-t} D'_{x+t+k}$. Note that at policy inception, insured and payment population coincide and then $D'_x = D_x^{(0)}$.

Considering a premium $P^{(1)}$ paid annually during m years, the equivalence principle states that

$$\frac{(N'_x - N'_{x+m})}{D'_x} \cdot P^{(1)} = SA \cdot \sum_{j=1}^{J} \alpha_j \cdot \frac{(M_x^{(j)} - M_{x+n}^{(j)})}{D_x^{(0)}} \tag{3}$$

Then the annual premium $P^{(1)}$ is

$$P^{(1)} = SA \cdot \sum_{j=1}^{J} \alpha_j \cdot \frac{(M_x^{(j)} - M_{x+n}^{(j)})}{(N'_x - N'_{x+m})} \tag{4}$$

Considering a premium $P^{(12)}$ paid monthly during m years, the expected present value of received premiums derived from Woolhouse's formula [36] is:

$$12 \cdot \frac{(N'_x - N'_{x+m} - 1/24 \cdot (D'_x - D'_{x+m+1}))}{D'_x} \cdot P^{(12)} \tag{5}$$

Then the monthly premium $P^{(12)}$ is

$$P^{(12)} = SA \cdot \sum_{j=1}^{J} \alpha_j \cdot \frac{(M_x^{(j)} - M_{x+n}^{(j)})}{12 \cdot \frac{N'_x - N'_{x+m} - \frac{11}{24} \cdot (D'_x - D'_{x+m+1})}{D'_x}} \tag{6}$$

Finally, the corresponding periodical gross premiums are derived from similar formula based on gross single premiums instead of single pure premiums.

[52] A simplification would be that W_j includes both policy stop and premium waiver. Then only W_j is required instead of $T_j \cdot W_j$. The current formulation aims to explicit the two possible events.

2.5.9 Accelerated Benefits

Accelerated benefits are benefits paid to the insured in anticipation of the terminal event of the policy. It is widely used in life insurance in which a critical illness diagnosis can allow the insured to receive a portion of the death benefit. The terminal death benefit will be then reduced by this early payment. In this section, accelerated benefits are benefits for selected cancer site diagnosis which as early payment will reduce the benefits of other cancer sites. The discussion hereafter describes how accelerated benefits can be managed using the pricing formulation from the previous section.

From the product designed presented in Illustration 2, assume a coverage of cancer type B with an accelerated benefit for cancer type A. The formulation $J = 2$ with $j = 1$ for general cancer type B with $\alpha_1 = 100\%$ and $j = 2$ for cancer type A with $\alpha_2 = 10\%$ does not hold for such product design. It does not include the features the benefit for cancer type B in case of previous diagnosis of cancer type A is reduced, and that a benefit for cancer type A cannot be claimed after a benefit from cancer type B. The correct formulation is to consider $J = 2$ with:

- $j = 1$ for cancer type A or B with $\alpha_1 = 10\%$
- $j = 2$ for cancer type B with $\alpha_2 = 90\%$

In that case, an insured diagnosed cancer type B get $100\% \cdot SA$ due to two benefits from $j = 1$ and $j = 2$. An insured diagnosed cancer type A gets $10\% \cdot SA$ due to the benefits from $j = 1$. The insured remains eligible to a benefit from $j = 2$, which would be $90\% \cdot SA$.

The product design presented in Illustration 3 has introduced a sublevel of accelerated benefit. As in Illustration 2, the cancer type A is an accelerated benefit for cancer type B, but in Illustration 3 the cancer type B is also an accelerated benefit for cancer type C. The formulation is to consider $J = 3$ with:

- $j = 1$ for cancer type A or B or C with $\alpha_1 = 10\%$
- $j = 2$ for cancer type B or C with $\alpha_2 = 90\%$
- $j = 3$ for special cancer type C with $\alpha_3 = 100\%$

In that case, an insured first diagnosed with cancer type C gets $200\% \cdot SA$ due to three benefits from $j = 1$, $j = 2$ and $j = 3$. An insured first diagnosed with cancer type A gets $10\% \cdot SA$ due to the benefits from $j = 1$. The insured remains eligible to a benefit from $j = 2$, which then would be $90\% \cdot SA$, or from $j = 3$, which then would be $190\% \cdot SA$. An insured first diagnosed with cancer type B gets $100\% \cdot SA$, while remaining eligible for a benefit for cancer type C with a corresponding benefit of $100\% \cdot SA$.

Such formulations allow to use the pricing formula presented in the previous section for accelerated benefits, while removing the assumption that the insured events are disjoints. A large variety of product designs can be derived by introducing additional or accelerated benefits. Accelerated benefits limit the total sum assured (i.e. the sum at risk for the insurer) and therefore limit the premiums paid by the insured.

2.5.10 Numerical Illustrations

We present here numerical results of both pure and gross premiums from the 3 illustrations performed using the the pricing model exposed in the previous sections:

- P1 (Product 1) - Illustration 1 with Cancer type A all causes cancer (100% SA)
- P2 (Product 2) - Illustration 2 with cancer type A skin cancer (10% SA) accelerated benefit and cancer type B all causes cancer except cancer type A (100% SA)
- P3 (Product 3) - Illustration 2 with cancer type A skin cancer (10% SA) sub-accelerated benefit and cancer type B all causes cancer except cancer type A (100% SA) accelerated benefit and cancer type C pancreas, liver or esophagus cancer (200% SA)

with the following characteristics:

- Sum Assured SA of 10,000
- Guaranty and payment term is insured age 80 years old
- Gross premiums include a 20% risk margin rate and 10% loadings rate
- Incidence rates from Japan Cancer Information Service (diagnosis year 2015)

Pure Premiums

		Single Premium			Annual Premium			Monthly Premium		
Age	Gender	P1	P2	P3	P1	P2	P3	P1	P2	P3
30	Male	4,089	4,041	4,697	90	89	104	8	7	9
40	Male	4,056	4,009	4,662	114	113	131	10	9	11
50	Male	3,965	3,919	4,559	153	151	176	13	13	15
60	Male	3,630	3,586	4,156	214	211	244	18	18	21
70	Male	2,565	2,529	2,901	290	286	328	26	24	28
30	Female	2,816	2,777	3,020	62	61	67	5	5	6
40	Female	2,718	2,680	2,921	76	75	81	6	6	7
50	Female	2,434	2,397	2,632	90	89	97	8	7	8
60	Female	1,961	1,928	2,138	107	105	117	9	9	10
70	Female	1,217	1,192	1,337	129	126	141	11	11	12

Table 13: Cancer Insurance Product: Pure Premiums Illustration

Pure premiums for male are greater than premiums for female, with a larger gap for old ages. Single premiums vary very little for younger ages. Premiums for P2 are logically smaller than premiums for P1, whereas premiums for P3 are greater than for P1.

Gross Premiums

		Single Premium			Annual Premium			Monthly Premium		
Age	Gender	P1	P2	P3	P1	P2	P3	P1	P2	P3
30	Male	5,453	5,388	6,262	121	119	138	10	10	12
40	Male	5,408	5,345	6,216	153	151	175	13	13	15
50	Male	5,287	5,226	6,078	204	201	234	17	17	20
60	Male	4,840	4,782	5,541	285	281	325	24	24	27
70	Male	3,420	3,371	3,868	387	381	437	34	32	37
30	Female	3,755	3,703	4,026	83	82	89	7	7	7
40	Female	3,625	3,573	3894	101	100	109	8	8	9
50	Female	3,246	3,196	3,509	120	118	130	10	10	11
60	Female	2,615	2,570	2,850	143	140	156	12	12	13
70	Female	1,622	1,589	1,782	171	168	188	15	14	16

Table 14: Cancer Insurance Product: Gross Premiums Illustration

Gross premiums logically increase compared to pure premiums.

2.5.11 Model Extensions

Additional Premium Waivers

The product may include additional premium waiver in case of adverse events, such as total and permanent disability or critical illness (stroke, heart attack), which are not part of the product coverage. In that case, one must considered the annual incidence rate of those adverse events, and then the payment population also decreases by the event.

Benefit Pattern

It is common that the benefits may vary with time. Usually the benefit is smaller during the first months or years of the policy term. Formally, it can be expressed as having the parameters $(\alpha_J)_{1\leq j\leq J}$ time-dependant, i.e. having a set of parameters $(\alpha_{i,t})_{1\leq j\leq J, 1\leq t\leq n}$. This formulation allows to consider limited or augmented benefits, as well as an waiting period.

Return on Premiums

A product may include a maturity benefit, that is a lump sum at policy term. The lump sum may be fixed, for example as a percentage of the the sum assured, or can be a percentage of premiums received. In that case, it is called Return on Premiums (ROP) benefit.

Let $\phi \geq 0$ the percentage of paid premiums to be paid as a maturity benefit. At policy term, the remaining number of insured is $l^{(0)}_{x+t}/l^{(0)}_x$. Therefore the expected present value of a 1-euro maturity benefit is $D^{(0)}_{x+n}/D^{(0)}_x$. For a single Premium and assuming no premium waiver, the expected present value of benefits is:

$$(M^{(0)}_x - M^{(0)}_{x+n})/D^{(0)}_x \times SA + \phi \times D^{(0)}_{x+n}/D^{(0)}_x \times P_{ROP} \tag{7}$$

The amount of the single premium is determined following the equivalence principle :

$$P_{ROP} = \frac{M^{(0)}_x - M^{(0)}_{x+n} + \phi \times D^{(0)}_{x+n} \times P_{ROP}}{D_x} \times SA \tag{8}$$

Rearranging the terms, we get:

$$P_{ROP} \times (1 - \phi \times \frac{D^{(0)}_{x+n}}{D^{(0)}_x}) = \frac{M^{(0)}_x - M^{(0)}_{x+n}}{D^{(0)}_x} \times SA \tag{9}$$

Finally,

$$P_{ROP} = \frac{M^{(0)}_x - M^{(0)}_{x+n}}{D^{(0)}_x \times (1 - \phi \times \frac{D^{(0)}_{x+n}}{D^{(0)}_x})} \times SA \tag{10}$$

Rewriting the expression with $\frac{D^{(0)}_x}{D^{(0)}_{x+n}} = \frac{1}{\nu^n \times {}_npx}$, and recalling that P is the single premium without ROP feature, the single premium with ROP feature can be expressed as:

$$P_{ROP} = \frac{P}{(1 - \frac{\phi}{\nu^n \times {}_npx})} \tag{11}$$

The ROP pure premium increases with ϕ and decreases with technical rate and mortality level.

2.6 Underwriting Issues

2.6.1 Cancer Definitions

A clear set of cancer types covered by the policy is required for cancer insurance. For example, a minimum tumor size can be clearly express in some insurance conditions. While the tumor size may be included in some classification (such as TNM classifications), the classifications may be updated in the future, which is a risk for the insurer. Some policies exclude pre-cancer stage (carcinoma-in-situ). In some countries, regulatory definitions apply.

2.6.2 Adverse Selection

Anti-selection must be clearly addressed for such products. Insured may be aware of risk factors unknown by the insurer, such as sedentary life, smoking or risky professional activities or a cancer family history.

Some evidence of anti-selection has been found in the cancer insurance in Taiwan using questionnaires on policyholders [37]. Especially, they show the importance of knowledge of cancer family history in cancer incidence. In Taiwan, following the Taiwan's Insurance Act, an insurer cannot use the individual's family cancer history for underwriting purposes. Evidences of anti-selection has been also found in Japan with insured incidence greater than national registries' incidence [27].

Genetic testing can also be source of anti-selection. Many companies offers DNA testing for the presence of cancer-related genes. Especially, breast, ovarian, colorectal and prostate cancer are cancer types in which genetic is an identified factor[53]. Prices range are from one hundred to thousands of dollars. Such tests offer an information to an individual which is not available to the insurer and can encourage to buy (or not) a cancer insurance.

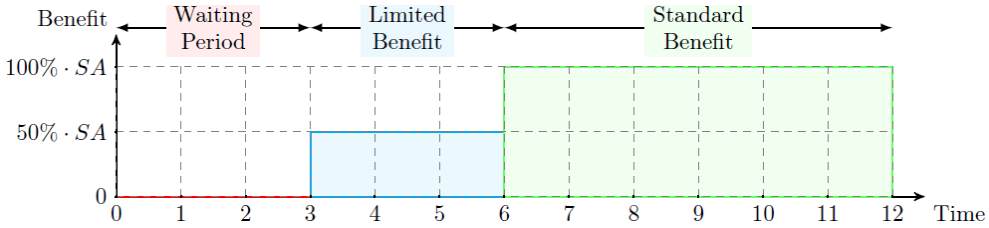
From a product design perspective, typical solutions are

- Medical questionnaire or medical selection, including cancer family history questionnaire (if legally applicable) provide the ability for the insurer to select the risks, to enrich his insured data and then assess and review medical selection for future sales.
- Pre-existing conditions exclusion. In a wide sense, we all have pre-existing conditions. Then, this exclusion can exclude benefits for cancer site with previous cancer diagnosis, declared (or not) to the insurer throughout the medical questionnaire.
- Main product or rider feature: a cancer coverage as a compulsory rider of a given death policy may reduce adverse selection. Still, stand-alone products are very popular in some markets.

[53] https://www.nhs.uk/conditions/predictive-genetic-tests-cancer/

- Waiting period, that is a time period from policy inception before the coverage starts. It is up to 6 months in the previous product illustrations. While it is a strong limitation to anti-selection, it is also a benefit limitation and can make the product hard to sell
- Limited Benefit at policy early stage (for example 2 years), in order to limit the benefit amount for early claims

As an illustration, the benefit pattern can then be split in 3 parts combining waiting period and two different levels of benefits through time. The first part is the waiting period during which there is no benefit in case of cancer diagnosis. The second part is designed to limit the benefit in case of cancer diagnosis. The benefit is then maximal in the third part. The following figure illustrates this benefit pattern in which the limited benefit is 50% of the sum assured SA.



Adverse selection can also be detected after underwriting processes. From a risk management perspective, it should be addressed with the following axes:

- Morbidity: detection and analysis of a general level of claims, at least by gender, age and cancer site, as well as cancer stage if available
- Time: detection of general level of claims, early claims occurring right after policy inception or the waiting period) and late claims occurring right before policy term)
- Spatial: detection of areas with high claim rates

Following such reviews, business decisions can try to mitigate the risks. Possible decisions include review of product design, underwriting process, sales limitation or reinsurance solutions.

2.6.3 Moral Hazard

Due to complex cancer causes, moral hazard is difficult to control. A insured may have a riskier behavior because he is covered for the disease. Public screening programmes may mislead an individual that such public policies solve cancer problems [28] whereas fighting cancer require individuals to mitigate cancer risk factors.

In order to prevent moral hazard, general prevention information can be provided to insureds.

3 Cancer Incidence Prediction

3.1 Motivation

Cancer insurance relies on a deep knowledge of cancer incidence, that is the number of individuals diagnosed with cancer over a population during a defined time period (usually an annual incidence rate per gender, per age or age group and per country or province).

Cancer incidence impacts both pricing and profitability. Especially, the evolution of cancer incidence through time will impact the profitability of a given cancer incidence product, which can be designed as a product with a large profitability, but expected profits can turn into losses due to a cancer incidence increase compared to the technical assumptions used in the pricing.

This chapter discusses different techniques to predict the future cancer incidence rates. Formally, let $q^{(j)}_{y,x+t}$ be the yearly incidence rate at year y for cancer of type j of an insured aged $x+t$. The aim of this chapter is to evaluate some models to predict the incidence rates for the next σ years, i.e. $(q^{(j)}_{y+s,x+t+s})_{1\leq s\leq \sigma}$ for all cancer sites as well as cancer all causes.

Due to long-term coverage mentioned in the previous sections (up to age 80, 100 or whole life coverages), forecasts must be done at long-term horizons. Long-term coverage is a motivation of cancer incidence prediction, because the incidence at a distant horizon is more likely to differ from the incidence at short horizon compared to the current information. It is also a challenge because long-term forecasts might face insufficient historical data.

3.2 Data

Historical data is needed to perform incidence rates prediction. Data source can be private or public. An operating insurer in cancer insurance business has a knowledge of both insureds and claimants and can build experience tables if his portfolio is large enough. A reinsurer can compare or merge databases from different insurers and then assist a newcomer or a player in this market. Also, international organizations or national institutions may publish statistics about cancer and cancer incidence at national or regional level.

We use in this study public data available online from Korean and Japanese national institutions. Both countries have a long history in terms of cancer registries. A summary of the datasets is presented hereafter and additional graphs par gender and age groups in the appendix.

3.2.1 Databases

Korea

The data source is cancer incidence database from the KOrean Statistical Information Service (KOSIS[54]). Available data is yearly incidences rates from cancer types defined using ICD-10 Classification from 1999 to 2017, per sex and age group. The data has 18 age groups from 0-4, 5-10, (...), 80-84 and 85+ years old. Data per age group is very common for cancer incidence data. While some data per age would add precious information, age grouping should bring more robustness through time.

	A	B	C	D	E	F	G	H	I
1	Site	ICD-10	Sex	Year at diagnosis	Total	0-4s old	5-9s old	10-14s old	15-19s old
2	All Cancers	C00-C96	Total	1999	215,4	16	8,5	9,6	13,1
3	All Cancers	C00-C96	Total	2000	216	15,5	9,5	9,3	13,6
4	All Cancers	C00-C96	Total	2001	234,3	17,6	9,3	9,5	14,5
5	All Cancers	C00-C96	Total	2002	246,2	18	8,7	10,9	14,2

Figure 31: Extract of the database for incidence rates in Korea

The database can be summarized as follows:

Column	Meaning	Number of values	Possible values
Site	Cancer Site	24 values	All cancers (..), Stomach, Colon ...[55]
ICD-10	ICD code	24 values	from C00 to C96
Sex	Gender	3 values	Male, Female or Both
Year at diagnosis	Diagnosis Year	19 values	from 1999 to 2017
Total	All ages	1	
a_1-a_2	Age group $[a_1; a_2]$	18 age groups	17 of 5 years interval, 1 for 85 years & more

Table 15: Database Summary: Korean data from KOSIS

The available incidence rates are cancer all causes (C00-C96) and the following sites: oral (C00-C14), esophagus (C15), stomach (C16), colonrectum (C18-C20), liver (C22), gallbladder (C23-C24), pancreas (C25), larynx (C32), lung (C33-C34), breast (C50), cervix uteri (C53), corpus uteri (C54), ovary (C56), prostate (C61), kidney (C64), bladder (C67), brain (C70-C72), thyroid (C73), hodgkin disease (C81), non-hodgkin lymphoma (C82-C86 C96), multiple myeloma (C90), leukemia (C91-C95) and other and unspecified (Re. C00-C96). It gives the cancer incidence for the most common cancer sites while rare cancers are group into the category other and unspecified (Re. C00-C96).

For each age class and diagnosis year, the rate of C00-C96 is the sum of the different incidences rates per site (see also section 3.2.2). The rates are actually built using a number of cases and the population at mid-year.

[54] `https://kosis.kr`

Japan

Datasource is the Center for Cancer Control and Information Services website from the National Cancer Center (NCC[56]). Available data is yearly incidences rates from cancer types defined using ICD-10 Classification from 1975 to 2015, per sex and age groups. The figure below presents an extract of the database.

	A	B	C	D	E	F	G	H	I
1	コード	部位	ICD-10	性別	診断年	粗率	0-4歳	5-9歳	10-14歳
2	1	全部位	C00-C96	男女計	1975	184,7	13,2	8,0	7,9
3	1	全部位	C00-C96	男女計	1976	185,1	13,1	7,3	7,6
4	1	全部位	C00-C96	男女計	1977	189,4	13,9	7,4	7,3
5	1	全部位	C00-C96	男女計	1978	194,5	13,1	7,1	6,1

Figure 32: Extract of the database for incidence rates in Japan

The database can be summarized as follows:

Column	Meaning	Number of values	Possible values
コド	Code	40 values	range from 1 to 167
部位	Site	37 values	All cancers (..), Stomach, Colon ...[57]
ICD-10	ICD code	40 values	from C00 to C96 and D00 to D49
性別	Gender	3 values	Male (男), Female (女) or Both (男女計)
診年	Diagnosis Year	41 values	from 1975 to 2015
粗率	All ages	1	
a_1-a_2	Age group $[a_1; a_2]$	18 age groups	17 of 5 years interval, 1 for 85 years & more

Table 16: Database Summary: Japanese data from NCC

The data structure is similar to the database from KOSIS, especially for age groups segmentation, but differs for some cancer sites definitions. The NCC data distinguishes colonrectum cancer in two items colon (C18) and rectum (C19-C20). The cancer site skin (C43-C44) is available. The female reproductive organs are merged in one item uterus (C53-C55) except ovary (C56) available separately. Then, the kidney site includes additional sub-cancer kidney (C64-C66 C68) and malignant lymphoma are merged in one site malignant lymphoma (C81-C85 C96). There is no item other or unclassified site in the NCC data.

Breast cancer (C50) incidence rates are only available from 2003. However, breast cancer including cancer in situ (C50 D05) incidence rates are available for the whole period. The ratio between the two items have been calculated per age group. With the assumption that this ratio is constant through time, this mean ratio is then used to estimate breast cancer (C50) incidence rates before 2003 using breast cancer including cancer in situ (C50 D05) incidence rates. This methodology could be challenged by alternative modelling but offer a simple rule to re-build non available rates. Moreover, there is no change of level in incidence rates for cancer all causes in 2003. Then, this cancer all causes incidence rate includes properly breast cancer incidence although this value is not available in the database before 2003.

[56] https://ganjoho.jp/reg_stat/statistics/dl/index.html

Moreover, some external data is available for Japan [15]: smoking prevalence, daily calory intake and daily consumption of salt, vegetables and fish per capita. For smoking, the data is available yearly from 1989 for each age groups. For other items, the data is available at a 5 year interval from 1975 to 2005 and then yearly from 2006. Linear interpolation have been used to generate annual data.

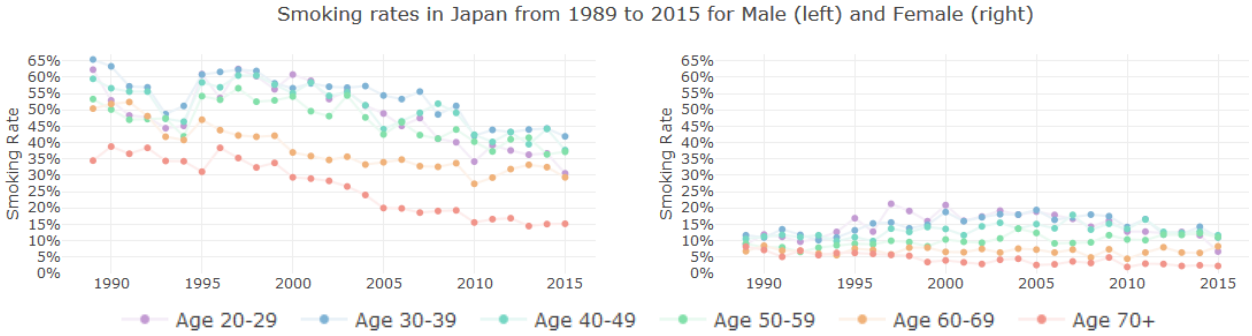


Figure 33: Smoking rates in Japan per gender

Smoking rates have been historically very high for male and have strongly decreased for all age groups. For example, it decreased from 65.4% to 42% for male in their 40's. For female, the rates have a bit increased between 2000-2010 compared to 1989-1995 and then have generally slowly decreased from 2010.

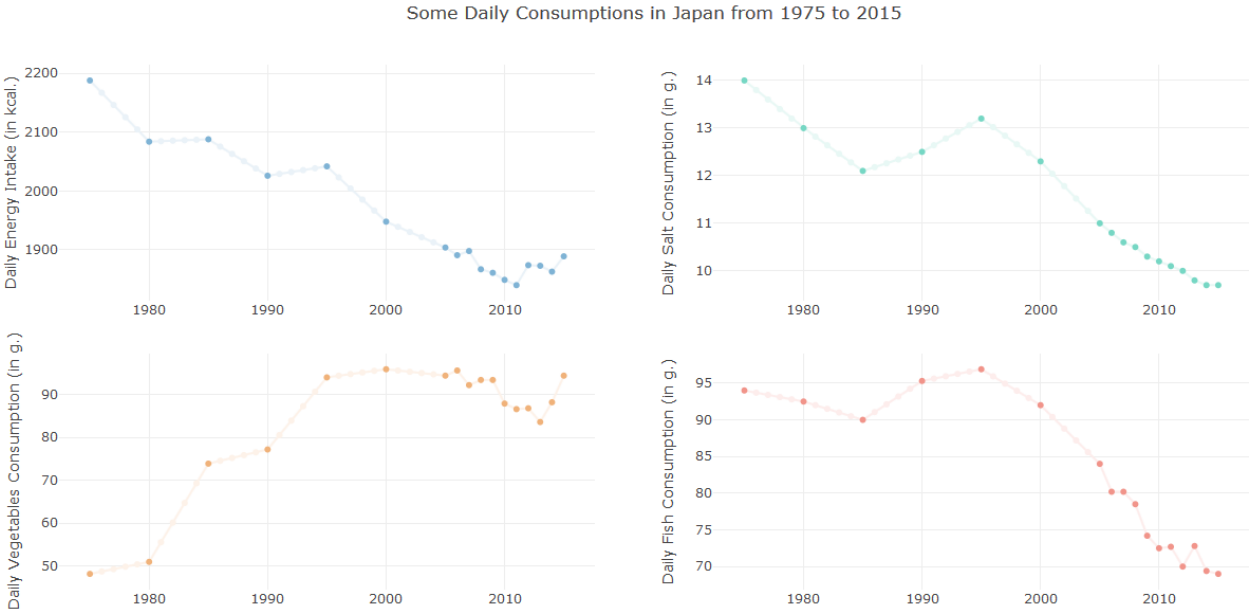


Figure 34: Daily Energy Intake, Salt, Vegetables & Fish consumption in Japan

Daily energy intake has decreased from 1975 to 2010 and then stable with a small upward trend. Consumption of salt and fish have globally decreased, despite a bounce in 1990 and 1995. Consumption of vegetables have strongly increased from 1975 to 2000 and is then stable, with a small downward trend. In all graphs, interpolated points have a lighter marker colour compared to the original data.

3.2.2 Data Visualization

Incidence rates for cancers all causes (i.e. ICD code C00-C96) and their evolution through time are presented below per gender and age class (x-axis). The color denotes the year, from green (year 1975) to orange (year 1999) and purple (year 2017). Incidence rates (y-axis) are presented using a log-normal scale:

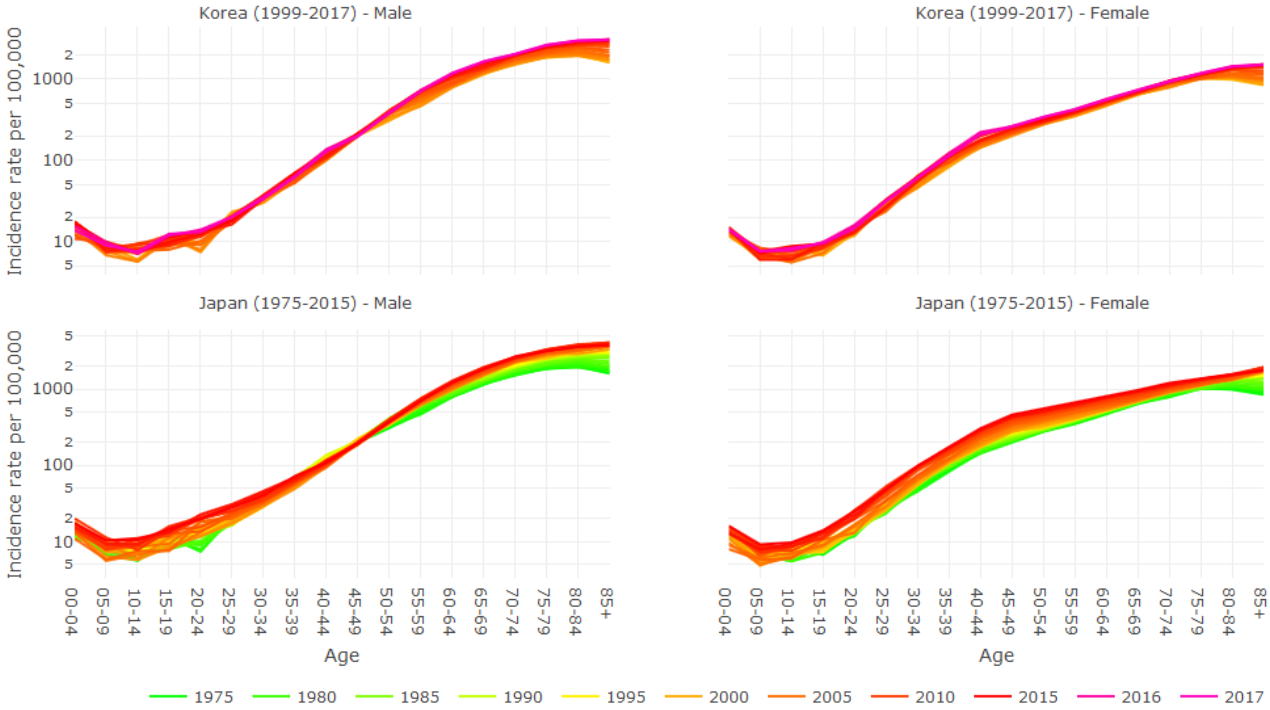


Figure 35: Cancer incidence rates in Korea & Japan per gender and age group

The first following stylized facts can be derived from the two databases:

1. Cancer incidence increases with age, except for age group 0-4.
2. Cancer incidence has increased with time, especially for old ages: the incidence rates curves has translated up. This is particularly visible for ages greater than 50 years old.
3. Cancer incidence is rather unstable for young ages: rates tend to be volatile through time. This may be due to the relative low incidence rates and cases before 25 years old.
4. Cancer incidence is similar for young ages per gender, then greater for female for middle ages and greater for male at old ages.
5. Cancer incidence differs per country: the patterns are relatively similar - at least at a visual level - but the log-normal scales flatten the differences. For example, in 2015 the cancer all causes incidence rates for female aged 65-69 is 1 per 100 in Korean and 0.8 per 100 in Japan.

In order to describe more precisely the databases, the next section presents some data per cancer site.

The figures below present the distribution of cancer incidence per site and per gender in the last available year in the databases (2017 for Korea, 2015 for Japan).

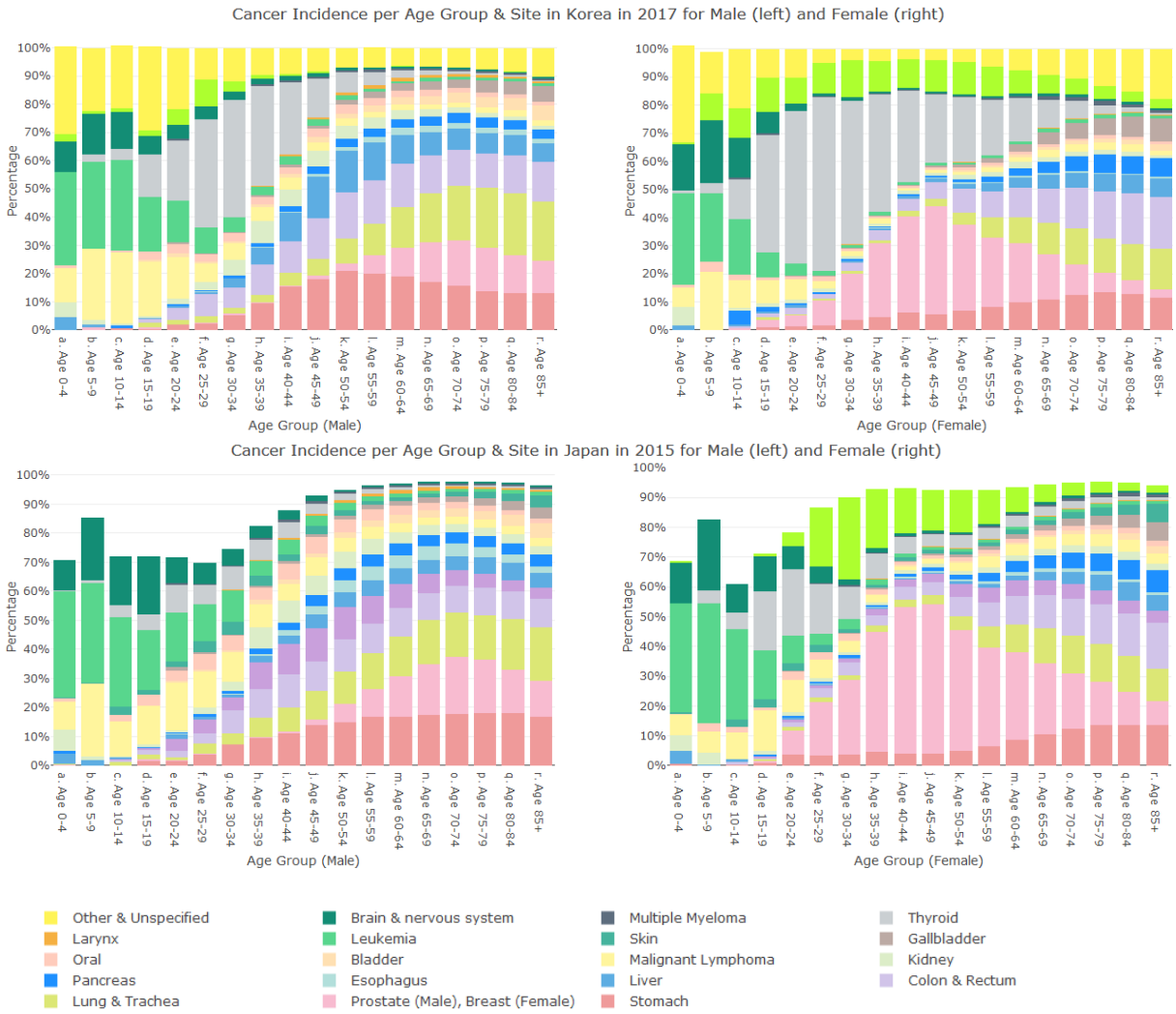


Figure 36: Incidences rates per site age in Korea in 2015

Most frequent cancers at young ages are leukemia, malignant lymphoma and brain cancer. Then, for male, gastric cancers (especially stomach, colon, rectum) are the main cancers and lung and prostate cancers rise at old ages. For female, breast cancer is the most prevalent cancer at middles ages, followed by uterus cancer. At old ages, breast cancer incidence relatively decreases and stomach, breast, lung and colon-rectum are the four main cancers. Figures show that thyroid cancer has a very large relative incidence in Korea for ages 15-59. This phenomenon has been called thyroid-cancer epidemic, due to the inclusion of thyroid cancer detection in health checkups and private screening programmes [38]. However, while thyroid cancer incidence has increased 25 times from 1993 to 2011, cancer mortality remained stable during that time period.

3.2.3 Discussion about Cancer All Causes

The cancer all causes incidence rates is a specific case because it includes sub-cases for which some data is available in the different databases. There are then two alternative options for forecasting cancer all causes:

(a) Forecast future cancer all causes predictions from cancer all causes time series

(b) Forecast future cancer all causes predictions from the combination of predictions from all different cancer sites.

For option (a), the estimation can be directly done using the available time series. While this option is easily implementable, it may hide different dynamics in cancer incidence per site.

For option (b), the estimation relies on the right estimation on each component. While some errors on each component may compensate, it could also lead to cumulative errors leading to large errors at aggregated level. For the Japanese data, the cancer all causes incidence rates is not equal to the sum of incidence per site. Then, the category other and unspecified sites can be deduced as the difference between cancer all causes and the sum of cancer incidence per site.

For Japan, this latter option leads to unexpected results on other and unspecified sites incidence rates for the years 2001 and 2002 for male. For that two data years, calculated other and unspecified sites incidence rates are evidently too low, especially in 2002 with negative incidence rates. There is no indication of methodology change in that year in the database notes. The chosen option for Japan is the option (a).

For Korea, analysis of cancer incidence rates show that thyroid cancer incidence is very high for middle ages and has increased steadily in the past years. Therefore an option (c) is designed to challenge option (a) in Korea: in option (c), cancer all causes is split between non-thyroid cancer incidence and thyroid cancer incidence rates. Each component is evaluated separately and the sum of the predictions is compared to the values cancer all causes test sample.

3.3 Modelling

3.3.1 Literature Review

A wide area of research about incidence rates prediction in actuarial science is mortality forecasting. It will be a source of models for analyzing cancer incidence rates. Then, the cancer forecasting literature will be discussed.

Literature Review for Mortality Rates Forecasting

Forecasting mortality rates has been used for both macroeconomic previsions and evaluation of insurance portfolios. A large class of models have been derived from the Lee-Carter model in which mortality is notably based on the combination of age-specific mortality and a general level of mortality. Model extensions include cohort effects (RH model) or multiple period effects (CDB model). Lee-Carter model extensions features can provide better fits on the overall population and some specific sub-populations (especially old ages) but require more parameters to estimate and may lack on parameters interpretation. Li and Lu (2018) [39] use VAR modelling with sparse modelling to avoid over-fitting in which mortality rates for each age class i_x can be only predicted by his neighbor age groups i_{x-1} and i_{x+1}. For a model with I age groups, a such constrained model required $2*(I-1)$ parameters whereas an unconstrained model requires $I(I+1)/2$ parameters. Guibert & al [40] use a penalized VAR method to obtain sparse parameters, without the a priori constraint to limit interactions between neighbor-ages group only. A new family of models applied to mortality forecasting are machine learning models such as neural networks [41].

Literature Review for Cancer Rates Forecasting

Forecasting cancer incidence rates is usually based on simple techniques or techniques applied to mortality rates. Simple techniques include constant or linear trends. For example, the IARC provides prediction forecasts (in number of cases) for 2020-2040 using the strong assumption that national rates, as estimated in 2018, do not change in the future[58]. Rahib and al. use a delay-adjusted average annual percentage change [42]. Such techniques offer a clear understanding of the forecasting method and an easy implementation. Australian Institute of Health and Welfare use an Ordinary Least Squares regression to forecast the trend on future incidence rates [43]. Jointpoint regressions, which summarize trend changes using segmented line regression between points (jointpoints) identified as trend change cite has been used in numerous studies. Mistry and al use 1975 to 2007 UK data to extrapolate to 2030 cancer incidence rates using a Age-Period Cohort (APC) model [44]. External variables are rarely used. A survey about lung cancer [45] reports that only 16 models out of 101 used data on smoking.

The next section details some models cited above which will be compared and discussed in the numerical illustrations.

[58] `https://gco.iarc.fr/tomorrow`

3.3.2 Regression Models

3.3.3 Time Series Models

ARIMA Models

ARIMA models express the d-lag of the explanatory variable Y_t as the combination of p weighted past value of Y_t and q weighted past value of a normal distributed variable [46]. Denoting by B the lag operator, $(\phi_{t-k})_{1\leq k\leq p}$ and $(\theta_{t-k})_{1\leq k\leq q}$ some coefficients and $(\epsilon_{t-k})_{1\leq k\leq q}$ some error terms, then the general form of ARIMA models is as follows:

$$(1-B)^d Y_t = \sum_{k=1}^{p} \phi_{t-k} Y_{t-k} + \sum_{k=1}^{q} \theta_{t-k}\epsilon_{t-q} \tag{12}$$

For a given p and q dimensions, the parameters to be estimated are $(\phi_{t-k})_{1\leq k\leq p}$ and $(\theta_{t-k})_{1\leq k\leq q}$. ARIMA($p$,0,0) are named AR-models, and ARIMA(0,0,q) MA-models. In a ARIMA model, there is no interaction between other time series. This can be modelled using a VAR model.

VAR Models

Vector Auto Regression (VAR) model is multi-dimension generalization of AR models [47]. In a VAR model, the explanatory variable Y_t is a n-dimension vector, with $Y_t = (Y_{1,t}...Y_{n,t})^T$, and the dependant variables are $(Y_{t-k})_{1\leq k\leq p}$. A VAR(p) model is as follows:

$$Y_t = c + \sum_{k=1}^{p} A_{t-k} Y_{t-k} + \epsilon_t \tag{13}$$

VAR are estimated using OLS (Ordinary Least Square). For a two-dimensional VAR(1),

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{14}$$

that is, 6 parameters are required to be estimated, i.e. 2 more than compare to two AR(1) models. Note that one condition is that $COV(\epsilon_{1,t}, \epsilon_{1,s}) = \sigma_{1,2}\mathbb{1}_{s=t}$.

A sparse VAR is a VAR in which many elements of A are equal to zero. The number of estimated parameters is therefore smaller compared to a unrestricted VAR, which should prevent over-fitting. Sparsity is obtained considering a minimization which includes a penalty function g, that is:

$$\min_{c,A} \sum_{t=1}^{T} \|Y_t - \sum_{k=1}^{p} A_{t-k} Y_{t-k}\|^2 + g(A) \tag{15}$$

Multiple penalty mechanisms can be considered [48]. We use *elastic net* penalty (BasicEN) which combines Lasso L1-penalty and the ridge L2-penalty, such as $g(A) = \lambda(\alpha\|A\|^1 + (1-\alpha)\|A\|^2)$, with λ and α two hyper-parameters. $\lambda \geq 0$ is the general weight of the penalty function, and $0 \leq \alpha \leq 1$ is the weight of the lasso penalty component compared to the ridge one.

3.3.4 Neural Networks

Neural networks are a popular class of non-linear models in supervised learning. Such models are based on hidden layers composed of neurons, based on inputs or previous layers. Each neuron produces a result using a transfer function, used in the neurons of the next layer.

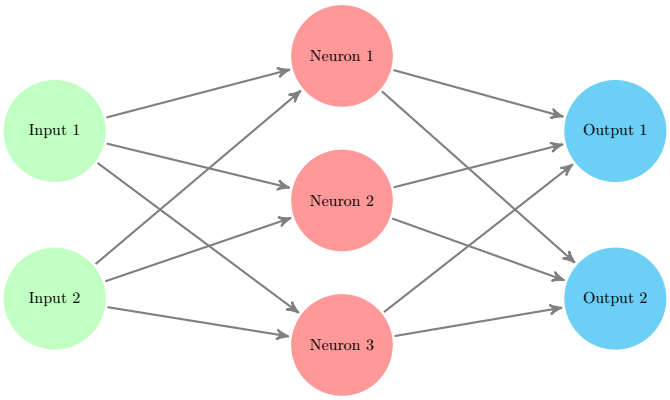


Figure 37: Neural Network with 1 hidden layer composed of 3 hidden neurons

A large variety of neural network exists. We use Multi-Layer Perceptron (MLP), which use lagged inputs for forecasting univariate time series. Inputs are lagged values of the explained variables and can also include external regressors.

MLP takes several items as parameters:

- Activation function: This is a non-linear function, usually a sigmoïd ($f(z) = \frac{1}{1+e^{-z}}$), hyperbolic tangent ($f(z) = tanh(z) = \frac{e^{z}-e^{-z}}{e^{-z}+e^{-z}}$) or ReLU ($f(z) = max(0, z)$)
- Number of lagged inputs: the number of lagged values to be incorporated in the model can be either fixed, or retained by the model following an algorithm. Such algorithms require iterative networks estimations subsetting lags in order to reproduce the seasonality of the series [49].
- Number of hidden layers: a network with multiple layers requires more parameters to estimate.
- Number of neurons per layer: each layer can be formed by different numbers of neurons
- Number of repetitions of the training process

Estimation is done using back-propagation algorithms. Initial weights of the different neurons are random. Then, from inputs of the training set, the network is propagated until the output and an error function is evaluated. The network is then evaluated backwards computing the gradient of the loss function in order to update weight values. And this procedure repeats using updated weights and stops when then number of repetitions in inputs is reached. A complete description of MLP is available in [50].

Lee-Carter

Lee-Carter model is a traditional model for mortality forecasting, originally based on US mortality rates [51]. Denoting by m_{at} the log of the mortality rate in age group a ($a = a_{min}, ..., a_{max}$) and time t ($t = t_{min}, ..., t_{max}$),

$$m_{at} = \alpha_a + \beta_a k_t + \epsilon_{at} \tag{16}$$

with:

- α_a is an age-specific component independent of time
- β_a is an additional age-specific component
- k_t is a time-varying parameter summarizing the general level of mortality
- ϵ_{at} a normal distributed random variables

Adapting this model to a cancer incidence, this model allows to model incidence rates using an age and time component. Forecasting is done by extrapolating k_t, usually with random walk with trend or ARIMA modelling.

The model being non-identifiable, the estimation is done using parameters constraints, usually:

$$\sum_{a=a_{min}}^{a=a_{max}} \beta_a = 1, \sum_{t=t_{min}}^{t=t_{max}} k_t = 0 \tag{17}$$

Many models have been since derived from Lee-Carter models, including models with cohort effects and multiple interactions between age and time. Such models requiring more parameters to estimate, only Lee-Carter model is considered here.

Comparison

The following table compiles the pros and cons of the different models previously exposed:

Model	Pros	Cons
Naïve	Does not require historical data	Depends on one value
	Easily understandable	No trend forecast
VAR	Multivariate forecasting through age class	Large number of parameters
Sparse VAR	Multivariate forecasting through age class	Management of Penalty mechanism
Neural Network	Handle multiple forms of non-linearity	Structure Choice (neurons and layers)
		Non-interpretable model
Lee-Carter	Multivariate forecasting through age class	Fixed nonlinear structure
	Interpretable model	

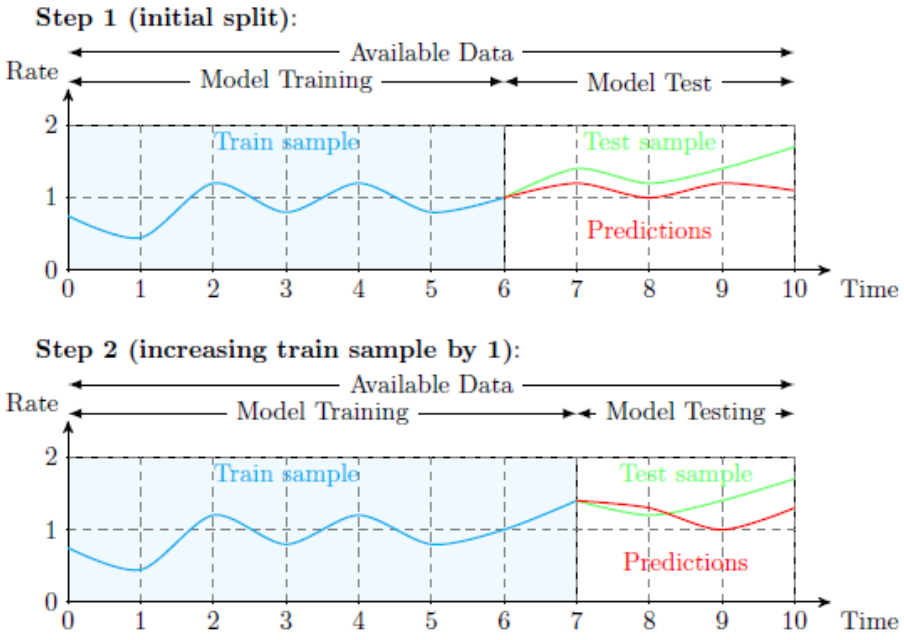
3.4 Numerical Results

3.4.1 Methodology

The usual methodology is to consider some randomized splits to get a train and test sample. However, the randomization approach does not hold for time series due to time-dependence of observations.

The methodology used is a nested cross-validation. First, we perform an initial split in which the k-first observations belong to the train sample, and the remaining n-k future observations to the test sample. See for example the initial split in the figure below. We then adjust the models on the train set and predict value for the remaining n-k future observations, and evaluate the forecasting quality comparing forecasts to the test sample. The level of the initial split (percentage of train sample in the available data) is arbitrary. Then, to assess the stability of the model performance, we consider rolling-origin evaluations using k+*1*-first observations to the train sample, and repeat the product until having only 1 observation in the test sample. This method is applied independently for each gender and cancer site.

The methodology is illustrated below:



In this illustration, the whole procedure includes two additional steps.

While the first estimation steps are based on a smaller train data, it is in such steps that the forecast horizon is the largest. Oppositely, in the last estimation steps, the train sample is larger, more recent, but the test sample size decreases, and ends with a single value at the final estimation step.

3.4.2 Scope

The different models are used on the

- Japanese data, for which the historical data range is 1975-2015. Additional data is available from 1989.
- Korean data, for which the historical data range is 1999-2017

3.4.3 Parameters

We consider an arbitrary initial split with 75% of the data in train sample. For the data from Japan, the first estimation step uses data from 1975 to 2005 as train sample, and values from 2006 to 2015 for the test sample. The methodology includes 10 steps. For the data from Korea, the first estimation step uses data from 1999 to 2013 as train sample, and values from 2014 to 2017 for the test sample. The methodology includes 4 steps.

3.4.4 Models

The different models are compared in the estimation procedure:

1. Naïve modelling: the forecasts are equal to the last value of of the train sample. It is a common forecasting method when there is no enough information for alternative modelling. It is also used as a benchmark to compare the performance of the other models presented hereafter.

2. VAR model: a VAR(1) is selected, in order to limit the number of parameters to estimate. Still, this model requires too many parameters to be estimated with the Korean data. It is then only used with Japan data and helps to control to quality of Sparse VAR models.

3. Sparse VAR model: the selected form is also a VAR(1), and the sparse feature is obtain using the $elastic - net$ penalty to select the estimated coefficients.

4. Neural Network I (NN I), with 2 layers of 2 neurons and a selection of lags among the last 10 lags. More complex networks (with more layers, or more neurons per layer) have been tested without forecasting improvement.

5. Data-restricted Neural Network I (NN I$*$) : same as previous neural network, but the train data is reduced to sub-sample. This model is only used with the data from Japan, to assess the impact of external regressors. The historical data range starts from 1989. Results of this model are available in the appendix only.

6. Neural Network II (NN II), with external regressors, with the same characteristics as the previous one. This model in only used with the data from Japan, for which external regressors are available. The historical data range starts from 1989.

7. Lee-Carter Model, using the whole historical data.

3.4.5 Indicators

Some indicators are needed to assess the performance of the different models in the test samples. Denoting by $\hat{Y}_{d,k}$ the estimate of the value of the age group d at time k $Y_{d,k}$, the two indicators Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) are used. They can be computed for a given age group or calculated over the different age groups. The following expressions give the definition of the MAPE and the MSE over the different D age groups over T time periods:

- $MAPE = \frac{1}{DT} \sum_{d=1}^{d=D} \sum_{k=1}^{k=T} \frac{|\hat{Y}_k - Y_k|}{Y_k}$
- $MSE = \frac{1}{DT} \sum_{d=1}^{d=D} \sum_{k=1}^{k=T} (Y_k - \hat{Y}_k)^2$

MAPE is not affected by scale, which make such errors measures easier to interpret. While MAPE is necessarily positive, a MAPE close to 0 is a sign of a good prediction. A limitation of MAPE is the training values must be different than 0, which is not met for some types of cancers, usually at young age.

MSE is affected by scale but is also pertinent because an error of 1% for a rate of 5% has much more consequences that for a rate of 0.5%. Both MAPE and MSE consider absolute errors and therefore no error compensation are possible.

Model Selection

The quality of prediction is determined by a small MAPE and a small MSE for each age group and each estimation step. The model selection rule used hereafter is the model with the smallest MSE to penalize the forecast errors on large incidence rates. At similar performance, model simplicity and parsimony (small number of estimated parameters) would be preferred.

The age range for model selection is 30-79, that is all age groups between 30-34 and 75-79. It corresponds to a large band for possible cancer cover and the results will be illustrated in the next chapter with a product having a guaranty term until insured age 80. Still, results of age groups 20-24 and 25-29 will be detailed, as well as 80-84 and 85+ for comparison purposes.

Discussion

A decision rule for selecting best models through the different age groups requires to specify the error weight on each age group. A classical rule is to use the same weight for each age groups, giving the same important to each groups. However one may be interest in focusing on some specific age groups, or to specify weights in a decision rule.

3.4.6 Implementation

Numerical implementation has been done with R. The different packages used are *vars* (VAR estimation), *bigVAR* (Sparse VAR estimation), *nnfor* (neural networks), *stMomo* (Lee-Carter estimation) and *plotly* (graphs).

3.4.7 Results - Japan

Illustration

The following figure illustrates the predictions of the different models for cancer all causes for male from age group 65-69. The train sample starts in 1975 and ends in 2005, and test sample starts in 2006 and ends in 2015. The data from the database, named *real data* in the figure legend is denoted by black markers, whereas the different models predictions are denoted by coloured markers.

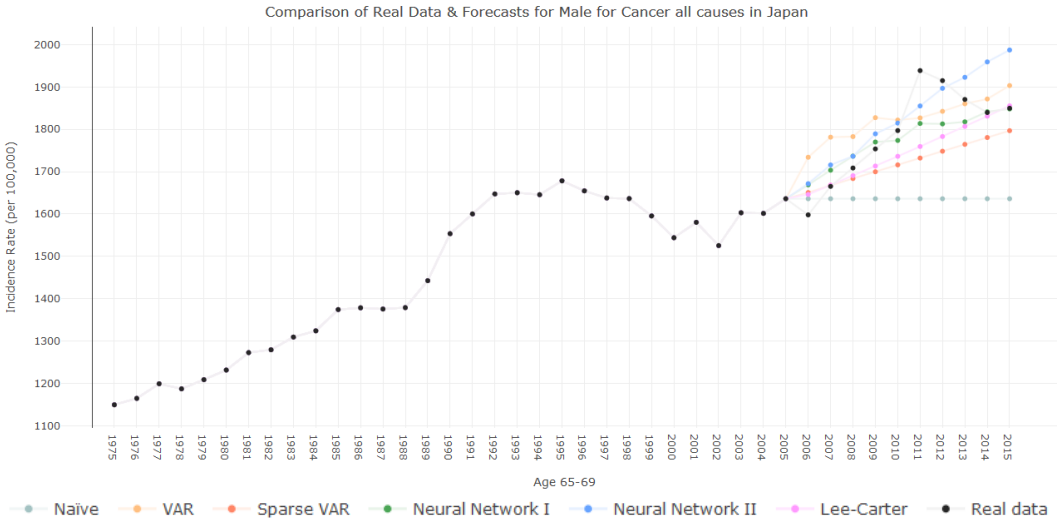


Figure 38: Model comparison (Japan data for male and all cancers) - Different forecasts and real data

In this example, the upward trend from 2004 to 2013 in the test sample is captured by all models except Naïve predictions which are necessarily constant. The *real* incidence rate is 1.64 per 100 in 2005 and reaches 1.85 per 100 in 2015. Sparse VAR model forecast a log-linear increase reaching 1.6 per 100 in 2015, while Neural Network I and Lee-Carter matches the real data in 2015, but with different evolutions throughout the test sample. VAR and Neural Network II forecasts the most important trends. The different forecasts are relatively smooth. Especially, the jump of incidence rates from 2010 to 2011 is not predicted by any models.

The next sections mainly detail the results for the cancer incidence all sites. The first step of the cross-validation, and then the remaining steps. Then, details about model estimation are provided. Finally, the indicators of the different models are compared in order to select the best model for each cancer site and each gender.

Quality of Forecast - Initial Step

In this section, we present the MAPE comparison between all models for the initial step (train test up to 2005, test sample from 2006 to 2015).

Mean Average Percentage Error (MAPE) per age group per model for Cancer all causes for Male in Japan

	1. Naive Prediction	2. VAR	3. Sparse VAR	4. Neural Network I	5. Neural Network II	6. Lee-Carter
Age 20-24	0.13	0.11	0.10	0.24	0.13	0.17
Age 25-29	0.18	0.10	0.18	0.24	0.15	0.26
Age 30-34	0.17	0.24	0.17	0.10	0.15	0.16
Age 35-39	0.10	0.17	0.09	0.10	0.11	0.07
Age 40-44	0.04	0.07	0.05	0.05	0.04	0.09
Age 45-49	0.06	0.08	0.08	0.07	0.06	0.03
Age 50-54	0.02	0.10	0.03	0.05	0.03	0.02
Age 55-59	0.10	0.11	0.06	0.07	0.10	0.03
Age 60-64	0.08	0.06	0.03	0.05	0.09	0.03
Age 65-69	0.09	0.04	0.04	0.03	0.03	0.03
Age 70-74	0.07	0.02	0.02	0.03	0.09	0.02
Age 75-79	0.06	0.03	0.03	0.04	0.06	0.05
Age 80-84	0.11	0.04	0.05	0.09	0.09	0.04
Age 85+	0.07	0.07	0.05	0.10	0.07	0.14

Figure 39: Model comparison (Japan data for male and all cancers) - MAPE at initial step

Ages 25 to 39 give poorer results with MAPE greater than 10%. Then, results for age group between 40 and 84 are relatively good with MAPE smaller than 10%, especially for Sparse VAR and Lee-Carter models. Results may deteriorate for age group 85 and more.

Mean Average Percentage Error (MAPE) per age group per model for Cancer all causes for Female in Japan

	1. Naive Prediction	2. VAR	3. Sparse VAR	4. Neural Network I	5. Neural Network II	6. Lee-Carter
Age 20-24	0.26	0.20	0.23	0.33	0.30	0.30
Age 25-29	0.25	0.16	0.24	0.29	0.23	0.31
Age 30-34	0.26	0.17	0.22	0.31	0.24	0.24
Age 35-39	0.23	0.09	0.20	0.28	0.17	0.17
Age 40-44	0.18	0.04	0.13	0.30	0.18	0.18
Age 45-49	0.17	0.09	0.12	0.16	0.06	0.19
Age 50-54	0.18	0.07	0.14	0.14	0.17	0.20
Age 55-59	0.16	0.07	0.11	0.14	0.16	0.20
Age 60-64	0.14	0.07	0.11	0.12	0.14	0.18
Age 65-69	0.13	0.08	0.11	0.15	0.13	0.14
Age 70-74	0.12	0.07	0.10	0.13	0.11	0.10
Age 75-79	0.10	0.05	0.08	0.10	0.10	0.08
Age 80-84	0.08	0.02	0.03	0.11	0.08	0.03
Age 85+	0.08	0.05	0.06	0.08	0.08	0.07

Figure 40: Model comparison (Japan data for female and all cancers) - MAPE at initial step

For females, all models show poor results for ages 20-39. The VAR model is almost always the best MAPE (except for age group 45-49). The average values of MAPE for female is larger than the ones for male.

Quality of Forecast - All Steps

We then extend the analysis and consider all the estimation steps, with increasing train samples, to assess if results are stable over the nested cross-validation procedure. For a given age group, the x-axis denotes the step (1 is the initial step with 11 values to predict, 10 the final step with only 1 value to predict), and the y-axis the MAPE value.

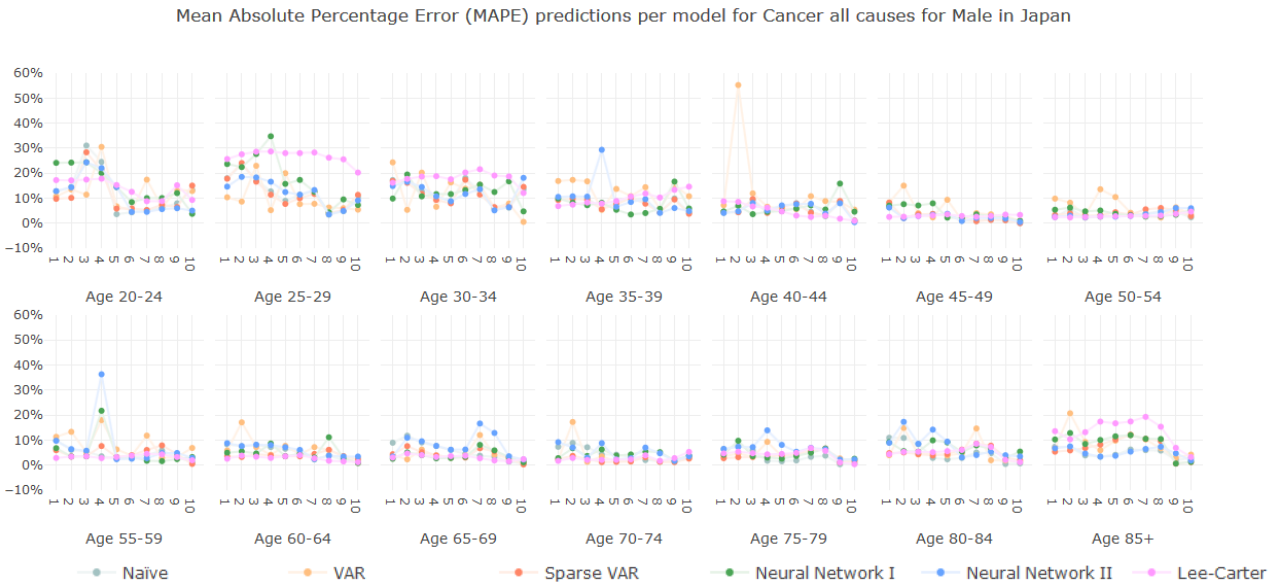


Figure 41: Model comparison (Japan data for male and all cancers) - MAPE criteria

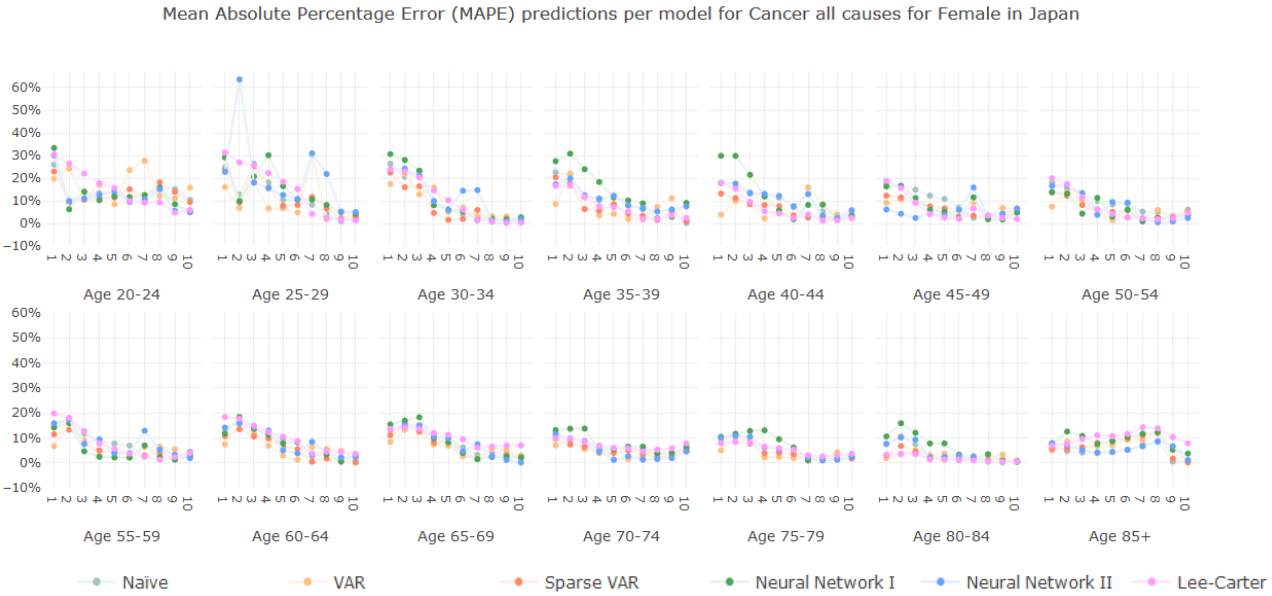


Figure 42: Model comparison (Japan data for female and all cancers) - MAPE criteria

Results confirm that the forecast error is greater for young ages with relatively poor results up to age 35-39. For young ages there is an important instability in results and the quality of forecast of each model, which MAPE reaching more than 20% in the first steps of the cross-validation procedure. From age 40-44 to age 80-84 the MAPE is relatively stable around 10% and the different models share the same trend in MAPE pattern.

MAPE decreases during the cross-validation procedure. For example, the MAPE is smaller in step 10 compared to step 1. This feature is generally expected because the forecast horizon is reduced throughout the cross-validation procedure. For middle ages, this decrease is more important for female compared to male.

Forecast is poorer for the ultimate age class compared to other old ages. This would indicate that some specific modelling could be applied to cancer modelling at old ages. Especially, because of the life expectancy increase, the average age for people aged more than 85 could not be constant over time (whereas it could be assumed for other age class). The impact of longevity is not modelled in the different models used here.

Estimation Details

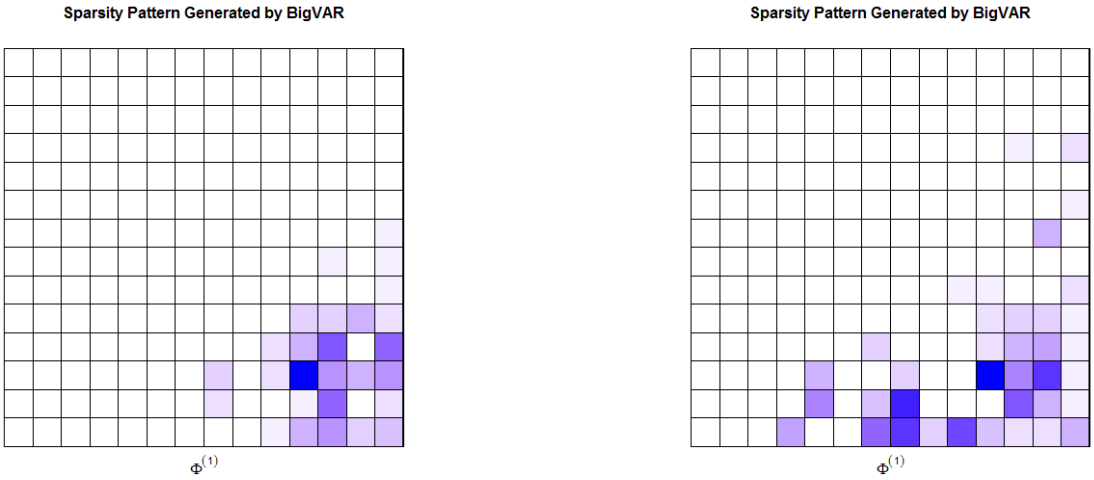
This section gives some details about the different models estimation for cancer all sites in order to assess the validity of each modelling and provides estimation results. More details are available in the appendix.

VAR

The multivariate Box-Pierce tests accept the hypothesis of non-autocorrelation of residuals for both male and female models, with a *p-value* of 1. The normality of residuals are both rejected by the Jarque-Bera test with a *p-value* of 0.

Sparse VAR

One critical element of Sparse VAR estimation is to assess the number of non-null coefficients in the matrix A, presented in the figure below. The first row represents the value of coefficients $a_{1,j}$, with the color white for null coefficients and a shade of purple for non-null values. More than 80% of estimated coefficients are null. Non-null coefficients concentrate on old ages. For young ages, the constant coefficients c (not represented in the figures below) only cause the dynamics for forecasted values. The normality of residuals are both rejected by the Jarque-Bera test with a *p-value* of 0.



(a) Male (b) Female

Figure 43: Sparse VAR estimation: coefficients of the matrix A

Sparse VAR have been modelled with *bigVAR* package, which does not provide tests for the presence of autocorrelation in the residuals. Univariate tests using Ljung-Box tests have been performed in each age group and 13 age groups out of 14 accept the hypothesis of non-autocorrelation for both male and female models.

Neural Network I

This table collects the selected lags for the neural network I per age group and gender:

Age	Selected lags for Male	for Female	Age	for Male	for Female
20-24	1, 3, 6	1, 4, 5, 8, 9, 10	55-59	1, 7	4, 6
25-29	1 to 10 (except 6)	1, 2, 3, 5, 9, 10	60-64	2, 5, 7, 10	2, 9
30-34	1, 2, 5, 7, 10	1, 2, 4, 6, 7	65-69	9	7
35-39	1, 10	1	70-74	1, 3, 4, 6, 7, 9, 10	1, 5, 6
40-44	1, 4, 5, 8	1, 2, 4, 5, 6, 7	75-79	1, 2, 6, 8, 10	1, 2, 3, 7, 9
45-49	1, 4, 5	7	80-84	1 to 7	1, 2, 5
50-54	1, 5, 6, 10	1, 3, 6, 7, 8, 9	85+	1	8

Table 17: Model Selection - Lags Selection for Neural Network I

It results a large number of different selected lags between the different segments. The lag 1 is often selected (21 out 28). However, there is no common pattern among the different estimations.

Neural Network II

This table collects similar data for the neural network II which includes external regressors:

Age	Lags for Male	for Smoking	for Salt	for Vegetables	for Fish	for Meat
20-24	1 to 10	1,4	5	none	none	none
25-29	4,5	none	7	1	2	none
30-34	1 to 10	4	none	1	2,3	none
35-39	1	none	none	none	none	none
40-54 (*)	1	none	none	none	none	none
55-59	1	1	none	none	1,2	none
60-64	1	1	none	none	1,2	none
65-69	1, 2, 9, 10	1, 7, 10	7	1	none	none
70-74	1	none	none	none	none	none
75-79	2	3	none	none	none	none
80-84	1	2, 4, 5	none	none	none	none
85+	1	none	none	none	none	none
Age	**Lags for Female**	**for Smoking**	**for Salt**	**for Vegetables**	**for Fish**	**for Meat**
20-24	1 to 10	1, 4	5	none	none	none
25-29	4, 5	none	7	1	2	none
30-34	1 to 10	4	none	1	2, 3	none
35-39	1	none	none	none	none	none
40-54 (*)	1	none	none	none	none	none
55-59	1, 10	1	none	none	1, 2	none
60-64	1, 10	1	none	none	1, 2	none
65-69	1, 2, 9, 10	1, 7, 10	7	1	none	none
70-74	1	none	none	none	none	none
75-79	2	3	none	none	none	none
80-84	1 to 10	4, 5	none	none	none	none
85+	1	none	none	none	none	none

Table 18: Model Selection - Lags Selection for Neural Network II

Selected lags in Neural Network I for the incidence rates strongly differs than Neural Network II. Recall that the network differs by the presence of external regressors as well as the training data which starts in 1975 for the former and 1989 for the latter. Age groups 40-44, 45-49 and 50-54, denoted by a star (*), have the same results. External regressors are selected for you. Meat consumption is not selected in the different estimations.

Lee-Carter

The following figures represent the values of the estimated parameters α_x, β_x and k_t:

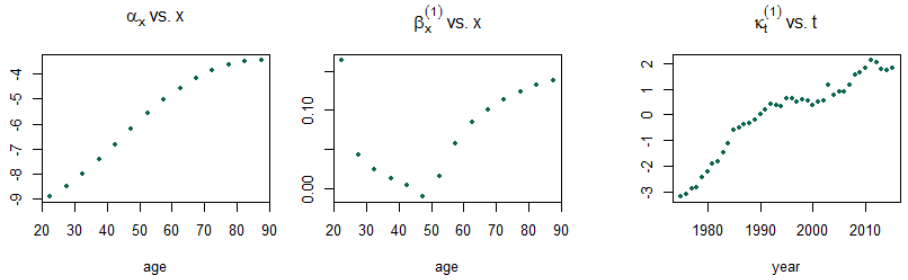


Figure 44: Lee-Carter estimation results for Cancer all sites for Male

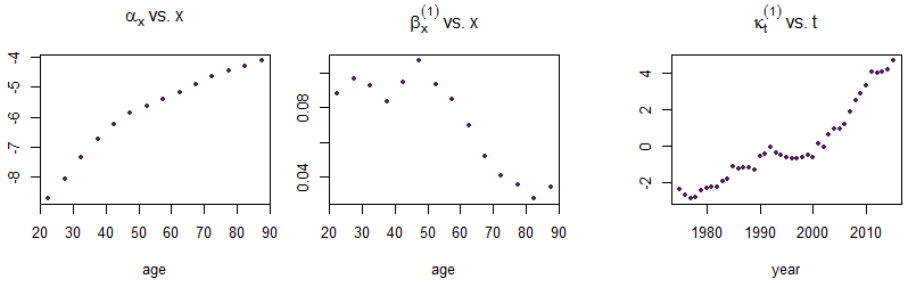
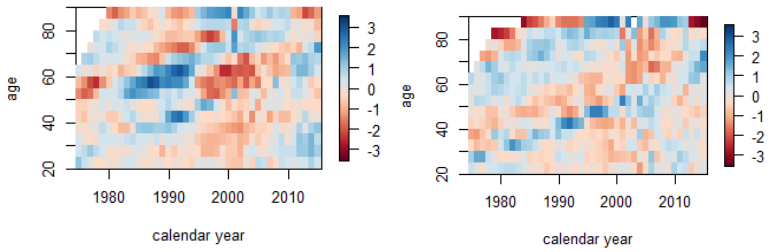


Figure 45: Lee-Carter estimation results for Cancer all sites for Female

α_x represents the average log-incidence ($ln(-9) = 0.1\%$ and $ln(-4) = 1.8\%$). Values of β_x and k_t must be analyzed following parameter constraints required by the model ($\sum_{a=a_{min}}^{a=a_{max}} \beta_a = 1$ and $\sum_{t=t_{min}}^{t=t_{max}} k_t = 0$). The values of β show mixed results. For male, the impact is greater at young & old ages compared to a quasi-null impact at middle ages. For female, the impact is more important at young ages. The values of k_t, which is the parameter used in the forecasts, show an upward trend for both male and female.

The following figures represent the residuals:



(a) Male (b) Female

Figure 46: Lee-Carter residuals for Cancer all sites in Japan

Some cohort effects (diagonals pattern, denoting residuals with a constant sign) are not captured by the model. Residuals are large for the extreme age group (85 and more) for female. Using a Jarque-Bera test, normality is rejected for both male and female residuals. More details are available in the appendix.

Model Selection

Details about 5 cancer sites This section presents the detailed results for 5 cancer sites which will be used in the numerical illustration. The model with the minimal MSE is highlighted in green.

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Cancer All causes	Naïve	6,760	5.5%	Naïve	4,523	8.5%
C00-C96	VAR	8,930	7.6%	VAR	2,598	6.4%
	Sparse VAR	3,985	5.0%	Sparse VAR	2,730	6.1%
	NN I	6,239	5.7%	NN I	5,279	8.8%
	NN II	10,017	5.6%	NN II	4,658	8.0%
	Lee-Carter	4,331	5.5%	Lee-Carter	4,290	7.8%
Skin	Naïve	31	20.2%	Naïve	12	23.3%
C43-C44	VAR	33	22.8%	VAR	10	23.5%
	Sparse VAR	27	18.5%	Sparse VAR	10	21.3%
	NN I	37	26.1%	NN I	26	31.2%
	NN II	43	25.0%	NN II	15	24.3%
	Lee-Carter	19	16.8%	Lee-Carter	6	18.6%
Pancreas	Naïve	24	16.4%	Naïve	13	18.7%
C25	VAR	31	20.5%	VAR	11	18.4%
	Sparse VAR	17	17.0%	Sparse VAR	8	18.5%
	NN I	40	18.2%	NN I	32	22.1%
	NN II	40	15.5%	NN II	19	17.5%
	Lee-Carter	19	15.0%	Lee-Carter	25	20.0%
Liver	Naïve	314	28.1%	Naïve	62	39.9%
C22	VAR	784	36.2%	VAR	112	48.4%
	Sparse VAR	529	27.1%	Sparse VAR	97	42.3%
	NN I	784	39.5%	NN I	76	46.3%
	NN II	478	27.1%	NN II	92	41.0%
	Lee-Carter	453	58.1%	Lee-Carter	189	52.1%
Esophagus	Naïve	29	14.9%	Naïve	1.9	17.4%
C15	VAR	35	18.1%	VAR	4.3	24.3%
	Sparse VAR	26	16.3%	Sparse VAR	2.9	18.6%
	NN I	54	16.3%	NN I	2.5	21.1%
	NN II	75	16.4%	NN II	5.7	27.9%
	Lee-Carter	38	17.8%	Lee-Carter	3.7	22.4%

Table 19: Model Selection - Details about 5 Cancer sites (Japan)

Results for cancer all causes show strong results compared to the different cancer sites. All models have a MAPE below 10%. For male, the Sparse VAR has the smallest MSE and MAPE. For female, the VAR has the smallest MSE but it is the Sparse VAR which has the smallest MAPE. Predictions for both model show very close results. For skin cancer, the Lee-Carter model shows the best results. the selected model is Sparse VAR For pancreas cancer and Naïve for liver cancer. The results are the most imprecise for liver cancer. For esophagus cancer, the selected model is not the same for male (Sparse VAR) and female (Naïve). Sparse VAR has overall better results than VAR whereas results for the to different neural networks are more mixed. Detailed results for other cancer sites are available in the appendix.

Results summary The table below collects the selected model (with minimal MSE) for the different cancer sites and gender. For interpretability purposes, only the MAPE values are presented below:

Cancer Site	Male	MAPE	Female	MAPE
Cancer All causes	Sparse VAR	5.0%	VAR	6.4%
Stomach	Naïve	13.5%	Naïve	10.2%
Colon	Naïve	9.0%	Sparse VAR	8.0%
Rectum	Sparse VAR	8.5%	Lee-Carter	11.9%
Pancreas	Sparse VAR	17.0%	Sparse VAR	18.5%
Liver	Naïve	28.1%	Naïve	39.9%
Esophagus	Sparse VAR	16.3%	Naïve	17.4%
Lung	Naïve	9.5%	Sparse VAR	12.4%
Prostate	Lee-Carter	18.2%	.	.
Breast	.	.	Lee-Carter	6.5%
Uterus	.	.	Naïve	12.8%
Ovary	.	.	Sparse VAR	9.2%
Skin	Lee-Carter	16.8%	Lee-Carter	18.6%
Thyroid	Lee-Carter	16.2%	Sparse VAR	10.4%
Brain	NN I	18.1%	NN II	17.9%
Bladder	Naïve	23.1%	Lee-Carter	24.1%
Gallbladder	Naïve	17.8%	Lee-Carter	17.6%
Kidney	Lee-Carter	12.9%	Lee-Carter	14.8%
Leukemia	Naïve	13.4%	Lee-Carter	16.0%
Malignant Lymphoma	Sparse VAR	13.4%	Lee-Carter	13.0%

Table 20: Model Selection - Results Summary (Japan)

The quality of prediction is the test sample vary a lot depending cancer site. Cancer all causes have minimal MAPE among cancer site with respectively 5.0% for male and 6.4% for female. The MAPE is also small for breast cancer (6.5%) and colon cancer (9.0% for male, 8.0% for female). This stresses that small values are more sensible to variations, which can cause a greater MAPE.

Naïve, Sparse VAR and Lee-Carter are the most selected models throughout the nested cross-validation procedure. VAR is only the minimal MSE model for cancer all causes for female, while neural networks models are selected for brain cancer only. Detailed results (available in the appendix) show that Sparse VAR always performs better than VAR, except for cancer all causes for female.

While the indicators on cross-validation are used to compare the quality of forecasting between the different models, the forecasts must be evaluated to ensure consistency with the historical data.

3.4.8 Prediction

This section discusses the results of the predictions following all models and focuses on the selected models. The results of cancer all types for Japan are discussed in this section whereas other results are available in the appendix.

Methodology

The different models are first fitted in the whole historical data. Then, forecasts are performed for the next 55 years (from 2016 to 2070).

Naïve forecasts are necessarily constant. VAR and Sparse VAR forecasts are derived from iterated multi-period forecasting procedure. The one period forecast is used to derive the second period forecast, without model estimation. Neural networks forecasts are the average of 100 simulated future paths. Lee-Carter forecasts are done forecasting k_t modelled by an ARIMA process.

Model Comparisons (Male)

The following figure compare the predictions of the different models for cancer all causes for male following an estimation in the whole available historical data:

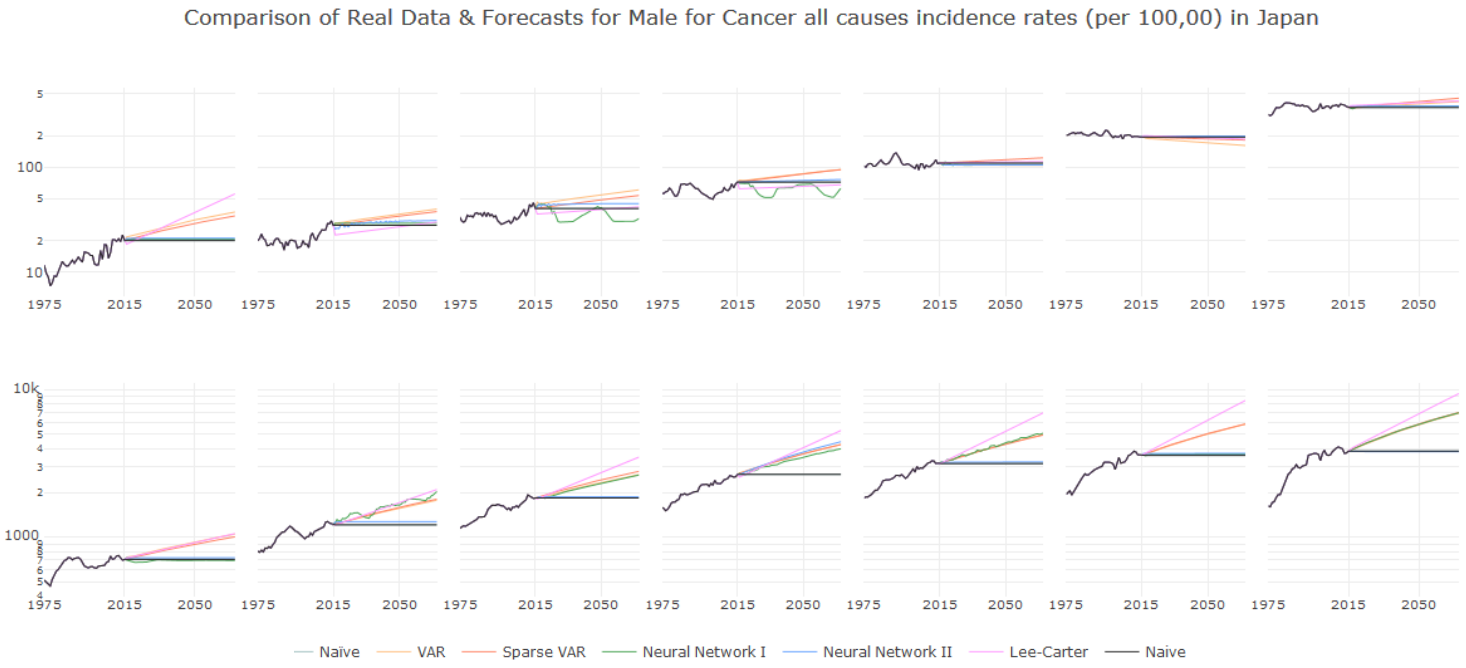


Figure 47: Model comparison (Japan data for male and all cancers)

For ages 20-39, most of forecasts show an upward trend whereas forecasts are all almost constant for ages 40-54. VAR and Sparse VAR models have similar predictions. Lee-Carter model gives the largest predictions. Neural Network I gives results with an upward trend but Neural Network II gives almost constant forecasts.

Model Comparisons (Female)

The following figure compare the predictions of the different models for cancer all causes for female following an estimation in the whole available historical data:

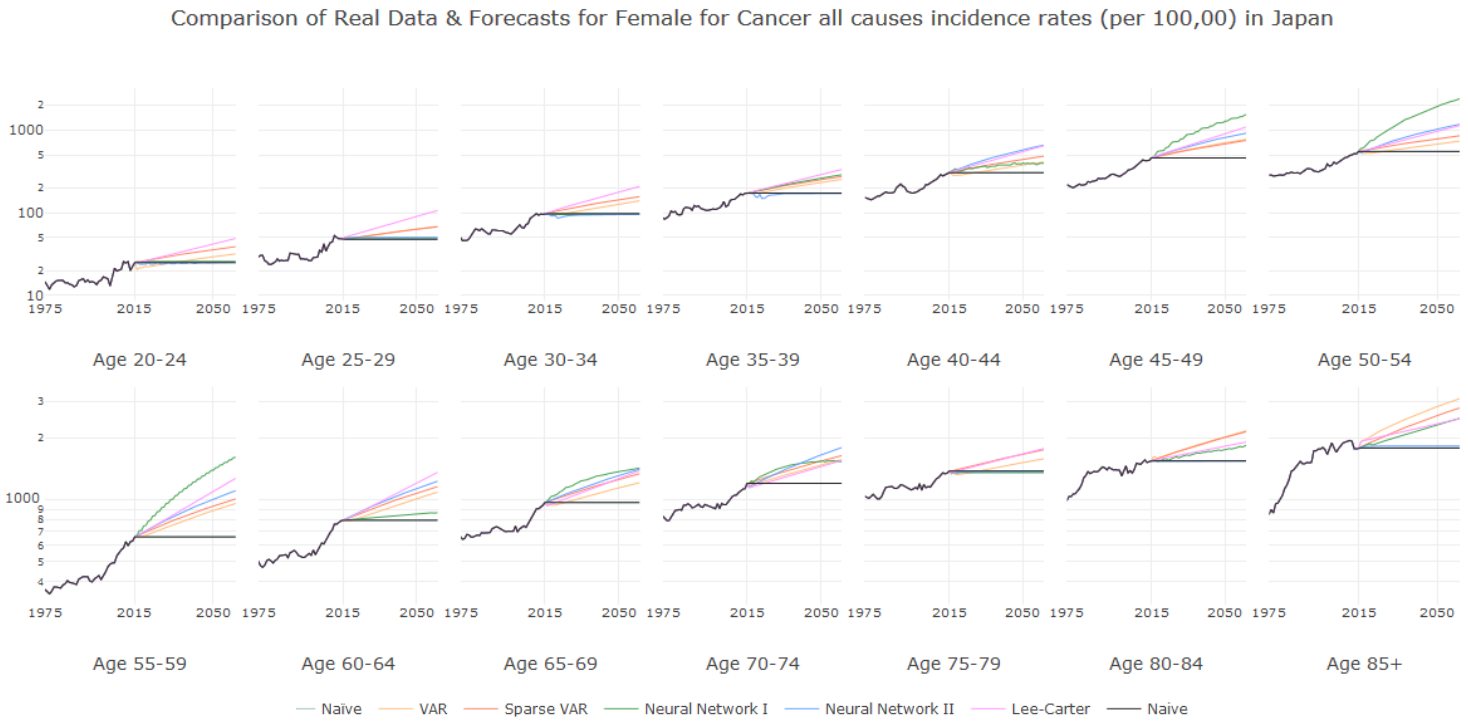


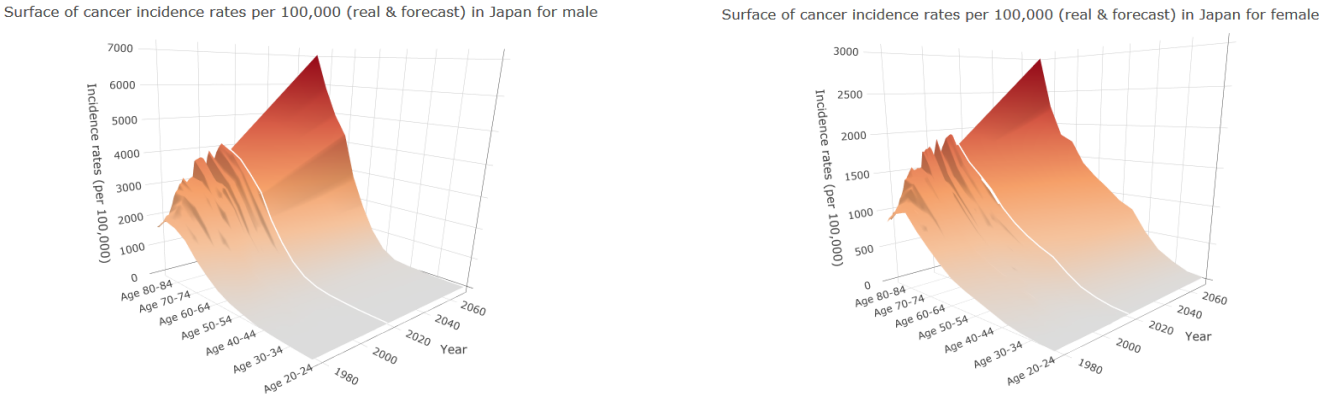
Figure 48: Model comparison (Japan data for female and all cancers)

Almost all forecasts (excluding Naïve model) have a long-term upward trend for female. The range in rates forecast can however strongly differ. For example for age group 55-59, the cancer all causes incidence rates in 2015 is 0.7 per 100 and the predictions in 2070 reach around 1 per 100 for VAR and Sparse VAR models to 1.8 per 100 (almost the double) for the Neural Network I model.

Forecasts for female using VAR model (model which minimized the MSE) show unusual values for age group 20-24, 40-44, 50-54, 65-69 and 75-79 for which predictions at short term (usually up to 10 years) are smaller than the 2015 incidence rates, whereas long-term predictions are higher. Then, the forecast trend is non-monotonic and there is a gap between last historical value and the one-year horizon predictions. This gap can reach -9.8%. The detailed table which compares the two predictions is available in the appendix. Recalling that, for female, MSE of VAR and Sparse VAR models are very close (respectively 2,598 and 2,730 for the MSE, and 6.4% and 6.1% for the MAPE - the MAPE of Sparse VAR being the smaller), and the overall quality is similar. Then, for forecasting consistencies purposes, the selected model is then Sparse VAR.

Focus on the Selected Models

The following surface figures collect the incidence rates per age group and incidence year for cancer all causes, both for real data (from 1975 to 2015) and forecasts (from 2016 to 2070) using the selected models. The limit between historical data and forecasts is denoted by the white line. The color range denotes the range of incidence rates, from 0 in grey to 7,000 per 100,000 (that is 7%) in red.



(a) Male

(b) Female

Figure 49: Surface of real and forecasted incidence rates for cancer all causes in Japan

The two figures summarize first the general level of cancer incidence at the different age groups, and emphasizes on the largest incidence rates at old ages.

Then, it illustrates the upward trend of forecasted incidence rates, visible for middle and old ages. This expected incidence increase is larger for old ages, such that the forecasted general level of cancer incidence at the different age groups show a larger gap in incidence between the age groups compared to the historical data. Also, the forecasts have less variability compared to the historical data.

The tabular values are presented in the next table.

The following table collects the evolution of cancer all causes rates to give more details about the results. A comparison for female between VAR and Sparse VAR models is available in the appendix.

Age	Male						Female					
Male	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
20-24	6.5%	13.1%	26.2%	39.3%	52.4%	65.5%	6.2%	12.4%	24.9%	37.3%	49.8%	62.2%
25-29	3.2%	6.5%	12.9%	19.4%	25.9%	32.3%	4.6%	9.3%	18.5%	27.8%	37.0%	46.3%
30-34	3.0%	6.0%	12.1%	18.1%	24.1%	30.2%	6.8%	13.6%	27.2%	40.7%	54.3%	67.9%
35-39	2.9%	5.7%	11.4%	17.2%	22.9%	28.6%	6.5%	13.0%	26.1%	39.2%	52.3%	65.4%
40-44	1.1%	2.2%	4.4%	6.6%	8.8%	11.0%	6.5%	13.0%	26.1%	39.1%	52.1%	65.1%
45-49	-0.6%	-1.1%	-2.3%	-3.4%	-4.6%	-5.7%	7.0%	14.0%	27.9%	41.9%	55.9%	69.9%
50-54	2.1%	4.2%	8.4%	12.6%	16.7%	20.9%	6.1%	12.2%	24.4%	36.7%	48.9%	61.2%
55-59	3.9%	7.7%	15.4%	23.1%	30.8%	38.5%	5.9%	11.8%	23.6%	35.4%	47.2%	59.0%
60-64	4.5%	9.0%	18.0%	27.0%	35.9%	44.9%	5.0%	10.1%	20.3%	30.4%	40.6%	50.8%
65-69	4.6%	9.3%	18.8%	28.2%	37.7%	47.2%	4.1%	8.4%	16.9%	25.4%	34.0%	42.5%
70-74	5.8%	11.2%	22.0%	32.8%	43.5%	54.3%	3.8%	7.9%	16.1%	24.3%	32.5%	40.7%
75-79	4.9%	10.1%	20.4%	30.7%	41.0%	51.3%	2.4%	5.4%	11.6%	17.8%	24.0%	30.1%
80-84	6.0%	11.8%	23.3%	34.8%	46.3%	57.8%	4.5%	8.8%	17.4%	25.9%	34.5%	43.0%
85+	7.9%	15.4%	30.5%	45.6%	60.6%	75.7%	6.0%	12.3%	24.9%	37.5%	50.1%	62.7%

Table 21: Evolution between last historical value & forecasts for cancer all causes

The upper table must be read as follows: for the selected model for male (Sparse VAR), the forecasts rates for age groups 20-24 are increased of +6.5% at a five-year horizon, +13.1% at a ten-year horizon compared to last historical data; for the selected model for female (Sparse VAR), the forecasts rates for age groups 20-24 decreases of 9.8% at a five-year horizon and 6.5% at a ten-year horizon compared to last historical data.

For male, results show monotonic forecasts compared to the last historical value. Forecasts have an upward trend, except for age group 45-49. For ages 20 to 44, the upward trend decreases with age. While the increase is 65.5% at 50-year horizon for age group 20-24, it is only 28.6% for age group 40-44. As mentioned, the forecasts for age group 45-49 are specific with negative evolution of forecasts. Still, the decrease remains low and reach only -5.7% after 50 years. For older ages, forecasts have an upward trend increasing with age. The results for age group 85+ differs from the other ages because the increase is much greater compared to other ages, reaching +75.7% after 50 years.

For female, all the forecasts increase at long term with a more homogeneous increase over the different age classes compared to male. Especially, age classes from 30 to 59 forecasts are around +25% at a 20-year time horizon and 59.0 to 67.9% at a 50-year time horizon. The increase is then small for ages up to 70. As for male, the results for age group 85+ show a strong increase in the forecasts compared to the historical data.

3.4.9 Results - Korea

Results summary The table below collects the selected model for the different cancer sites and genders.

An alternative estimation procedure for cancer all causes, named indirect method below, as been performed. Due to the importance of thyroid cancer in the historical data, the indirect estimation method for cancer all causes relies on the sum of (1) cancer all causes except thyroid cancer and (2) thyroid cancer in the nested cross-validation procedure. The direct estimation procedure is the standard used procedure used in the previous sections and also for other cancer sites hereafter.

Cancer Site	Male	MAPE	Female	MAPE
Cancer All causes (direct)	Naïve	5.7%	Naïve	6.1%
Cancer All causes (indirect)	Naïve	5.7%	Naïve	6.1%
Stomach	NN I	15.9%	Lee-Carter	7.6%
Colonrectum	Naïve	9.0%	Sparse VAR	8.9%
Pancreas	NN I	10.3%	Sparse VAR	13.8%
Liver	Sparse VAR	6.9%	Sparse VAR	11.5%
Esophagus	Lee-Carter	14.8%	Naïve	27.3%
Lung	Lee-Carter	5.7%	Naïve	12.2%
Prostate	Naïve	9.3%	.	.
Testis	Naïve	44.4%	.	.
Breast	.	.	Lee-Carter	4.2%
Cervix	.	.	Sparse VAR	8.2%
Corpus	.	.	Lee-Carter	8.8%
Ovary	.	.	Lee-Carter	8.4%
Thyroid	Naïve	21.9%	Naïve	26.1%
Brain	Naïve	14.6%	Lee-Carter	16.0%
Bladder	Lee-Carter	8.9%	Naïve	20.2%
Kidney	Naïve	9.6%	Lee-Carter	10.7%
Leukemia	Lee-Carter	11.9%	Lee-Carter	8.2%
Malignant Lymphoma	Sparse VAR	13.4%	Lee-Carter	13.0%
Gallbladder	Sparse VAR	17.8%	Lee-Carter	17.6%

Table 22: Model Selection - Results Summary (Korea)

The estimation with the smaller MAPE is the Lee-Carter model for breast cancer, with a MAPE of 4.2%. Then, the Naïve model for cancer all causes has a MAPE of respectively 5.7% for male and 6.1% for female, in the two direct and indirect estimations procedures. Therefore, related forecasts are constant incidence rates and no future trend are extrapolated from the historical data, which is a main difference compared to the data from Japan.

Colonrectum and Prostate cancer show quite small MAPE around 10%. Liver cancer, which has the worst MAPE using the Japan database, show relatively small MAPE (6.9% for male, 11.5% for female). Overall, the results per country strongly differs in terms of model selected and cross-validation estimation quality.

3.4.10 Limits

Data Quality

Both data used are public data from two countries with advanced medical system in which cancer registries have a long history. Still, such registries do not cover the whole health system and the exact method of incidence estimation is not disclosed. It results an uncertainty about rates used in the forecasting procedure. Available data per age instead of age group would be helpful, despite such data may be less robust through time.

Data Quantity

The historical data is relatively limited, especially comparing to similar studies about mortality rates. The Japan data has values for 41 years whereas data from Korea only 19. Similar studies about mortality forecasting usually rely on post-World War II data, and can include pre-WWII data for selected countries. In the cross-validation procedure, this available data is split in train and test samples and therefore train samples are small in the first steps of the procedure. The first steps of cross-validation procedure are the one with the larger test samples, which have much interest for long-term forecasts. Such estimation steps could then be weighted to have more importance in the estimation procedure. It can also be remarked that the forecast horizon is larger than the historical data range. Then, long-term forecasting should be handled carefully.

Data Availability

There is a time gap between the end of historical data (2015 for Japan, 2017 for Korea) and the time at which the data are used (2020). A smaller gap would reduce estimation uncertainty and increase the data quantity.

Data Adequacy

The data used for incidence rates are national statistics, for which an insured population is a sub-sample. Therefore, it is very likely that a sample bias exists. An insurer with cancer insurance history can then compare national statistics and in-house data to measure such bias. Eventually, the same estimation procedure can be applied on insurer data to build insurer specific incidence tables.

Future Uncertainty

Forecasting based on historical data assume that the future evolutions rely on past and present dynamics. Breakthroughs in cancer detection are then not considered, or at least with the same level of cancer detection improvements of the past decades. The same applies for other risk factors, including future behavior in terms of physical activities and food habits.

4 Profitability Testing

4.1 History

4.2 Methodology

Profitability testing (or profit testing) aims to evaluate the profits derived from insurance products. The profit calculation allows to assess the premium rates in central and stressed scenarios, and can be part of an iterative process of the ratemaking. Finally, it leads to operational decisions of a product launch, and gives profit target for the future profitability follow-up of the business.

The general method is to compute future cash-flows, derive future profits and to discount them using discount rates to obtain a different indicators, such as Profit Net Present Value, also named Present Value of Future Profit (PVFP) or Profit Margin. Additional indicators, such as Internal Rate On Equities (IROE) can also be considered but are not considered hereafter.

The profit can be calculated following different accounting norms, that is local norms or IFRS standards. The approach used hereafter is a classical and general framework.

4.3 Product

Following the product design and notations introduced in section 2.5.4, the possible benefits are:

- Sum Assured SA 10,000
- Sub-accelerated benefit for skin cancer: 10% SA
- Accelerated benefit for cancer all causes except skin cancer (named *general cancer* hereafter): 100% SA
- Benefit for esophagus, liver or pancreas cancer (named *special cancer* hereafter): 200% SA
- The policy stops at the benefit for special cancer

and the policy characteristics are as follows:

- Policy Term: Insured Age 80
- Payment Term: Insured Age 80
- Pricing per age and gender[59] with level annual premiums
- The premium reserve at death or lapse date is refunded

The product is assumed to be sold in the Japanese market, and priced using the most recent incidence rates of the national statistics. Then, incidence rates from diagnosis year 2015 presented in section 3.2.1 Datasource - Japan are used in the pricing.

The premium includes a 20% risk margin and loadings represent 10% of the gross premiums. The pure premiums and gross premiums are then available in the section 2.5.10 Numerical Illustrations.

[59] the pricing per gender is not authorized in the European Union but is a market practice in Asia

4.4 Sales Profile

We assume the sales of 100,000 policies following 10 different sales segments:

- (A) 10,000 insured male aged 30
- (B) 10,000 insured male aged 40
- (C) 10,000 insured male aged 50
- (D) 10,000 insured male aged 60
- (E) 10,000 insured male aged 70
- (F) 10,000 insured female aged 30
- (G) 10,000 insured female aged 40
- (H) 10,000 insured female aged 50
- (I) 10,000 insured female aged 60
- (J) 10,000 insured female aged 70

The aim is first to reproduce a portfolio with various insured profiles and also analyze profitability for each age & gender.

For the sake of simplicity and comparison purposes, the sales segment are uniform in terms of number of insured and sum assured. As an alternative, a numerical illustration using the Japanese population profile between 30 and 70 years old is provided in the appendix. Results are very similar between to the simplified sales profile above and the Japanese population profile.

4.5 Calculation Assumptions

The following assumptions are used in the base scenario:

- Annual lapse rate of 5%. This lapse rate does not depend on insured age, gender or policy seniority.
- Death incidence following Japan mortality table [52], which is a standard (static) life table per age and per gender.
- Claims are paid at claim occurrence. Therefore the incurred claims with the paid claims at all time periods, such that there is no claims reserve.
- Expenses supported by the insurer have two types:
  - Administration costs: 5 unit per policy per year, due to the required management and follow-up of the in-force policies.
  - Claim costs: 5% of the paid claims, due to the management of the claim process by the insurer.
- Start date January 1st, 2020
- Projection horizon: 50 years (until portfolio extinction). Due to policy term being insured age 80, the two sales segments (A) and (F) with insureds aged 30 at inception have both a 50-year guaranty term.
- Calculation step: annual. The number of policies, the amount of premiums, claims and other items are calculated annually.

4.6 Calculations

This section details the required calculations for the projection.

Populations

4 populations must be assessed throughout the projection: *healthy* population denotes insured with no cancer diagnosis, *skin* population denotes insured with skin cancer diagnosis (but not other cancer diagnosis), *general* population denotes insured with a cancer diagnosis (except skin or special cancer); all three populations denote a sub-population of insured. Finally, the *payment* population denote insured paying premiums, that is the sum of *healthy* & *skin* populations.

Denoting by l the diverse populations, $d^{a>b}$ the transition from state a to state b and la the lapse, the populations evolve as follows:

- $l_t^{healthy} = l_{t-1}^{healthy} - d_t^{healthy>skin} - d_t^{healthy>general} - d_t^{healthy>special} - d_t^{healthy>death} - la_t^{healthy}$
- $l_t^{skin} = l_{t-1}^{skin} + d_t^{healthy>skin} - d_t^{skin>general} - d_t^{skin>special} - d_t^{skin>death} - la_t^{skin}$
- $l_t^{general} = l_{t-1}^{general} + d_t^{healthy>general} + d_t^{skin>general} - d_t^{general>death}$
- $l_t^{payment} = l_t^{healthy} + l_t^{skin}$

Moreover the number of special cancer claims at time t using the previous expressions is $d_t^{special} = d_t^{healthy>special} + d_t^{skin>special} + d_t^{general>special}$.

For each population, the number of deaths is the product between population and the annual death rate. We assume the lapse occur at the end of the year (in line with the premium frequency), and that lapses occur only for the payment population (premium waived population do not lapse). Then, the number of lapses is the product between the population beginning of period and the annual lapse rate. No lapses are assumed for insured following a cancer all causes diagnosis except skin cancer, because no future premiums are expected due to the premium waiver and the coverage for special cancer remains.

The number of claims is calculated from populations and the probabilities to claim, such as defined in the section 2.5.4.

Premium Reserves

The premium reserve at the end of the year t depend on the remaining insureds and the corresponding premium reserves per policy V_t. This amount depends on the state of the in-force insured and therefore 3 values must be distinguish: ${}^{healthy}_{t}V$, ${}^{skin}_{t}V$ and ${}^{general}_{t}V$ (abbreviated ${}^{h}_{t}V$, ${}^{s}_{t}V$ and ${}^{g}_{t}V$ hereafter). They can be computed recursively using Fackler formula [53] as follows:

- ${}^{h}_{t}V = \frac{((q^{(0)}_{t+1} - q^{(1)}_{t+1})\cdot\alpha_1 + q^{(1)}_{t+1}\cdot\alpha_1 + q^{(2)}_{t+1}\alpha_2)\cdot SA}{(1+i)^{0.5}} - P + \frac{(1-q^{(0)}_{t+1})\cdot {}^{h}_{t+1}V + q^{(1)}_{t+1}\cdot {}^{s}_{t+1}V + (q^{(0)}_{t+1} - q^{(1)}_{t+1} + q^{(2)}_{t+1})\cdot {}^{g}_{t+1}V}{(1+i)}$
- ${}^{s}_{t}V = \frac{((q^{(0)}_{t+1} - q^{(1)}_{t+1})\cdot\alpha_1 + q^{(2)}_{t+1}\alpha_2)\cdot SA}{(1+i)^{0.5}} - P + \frac{(1-q^{(0)}_{t+1} + q^{(1)}_{t+1})\cdot {}^{s}_{t+1}V + (q^{(0)}_{t+1} - q^{(1)}_{t+1} + q^{(2)}_{t+1})\cdot {}^{g}_{t+1}V}{(1+i)}$
- ${}^{g}_{t}V = \frac{q^{(2)}_{t+1}\cdot SA}{(1+i)^{0.5}} + \frac{(1-q^{(2)}_{t+1})\cdot {}^{g}_{t+1}V}{(1+i)}$

Technical Result

The technical result is the result arising from the risk-based component of the premium. It is defined as the difference between the earned risk premium P^E and the incurred claims S. The earned risk premium P^E is defined as follows:

$$P^E = P - L + {}^{tot}_{t-1}V + {}^{tot}_{t}V + {}^{term}_{t}V + T/I \tag{18}$$

With T/I denoting technical interests. Recall that premiums are paid beginning of period and claims are paid middle of period. Denoting by $i^{(1/2)}$ the semi-annual equivalent technical interest rate defined as $((1+i)^{0.5} - 1)$:

$$T/I = P \cdot i + S \cdot i^{(1/2)} + {}^{tot}_{t-1}V \cdot i \tag{19}$$

The pure technical loss ratio (also named loss ratio hereafter) is defined as the ratio of claims on earned risk premium, that is $\frac{S}{P^E}$. It measures profitability from the risk component of the insurance product. A pure technical loss ratio greater than 1 means that the technical result is negative.

Operational Result

The operational result is the result arising from the management of the policy. It is defined as the difference between the earned loadings L and the expenses E. Written loadings are entirely earned therefore $L = l \cdot GP$. Expenses are the sum of administration costs (using a unit cost u) and claims costs (using a claims cost rate c). Then, $E = u \cdot l^{insured} + c \cdot S$.

Financial Result

The financial result is the result arising from the financial considerations of the policy. Is is defined here as the difference between the financial revenue L and the technical interests T/I. In order to focus on technical items, no financial revenue and no technical interests are considered. Therefore the financial result is always null.

Profit

The profit is the sum of the technical, operational and financial result.

Expected Present Value

The Expected Present Value (EPV) of a cash-flow or an accounting result is the sum of the future values discounted by discount rates to measure the present amount equivalent to the future amounts. Especially, the PVFP is the EPV of the profit. Using a flat discount rate d and a discounting at the end of the year, the PVFP at calculation date is $\sum_{t>0} \frac{profit_t}{(1+d)^t}$.

Customer Value

The customer value is defined as the ratio of benefits on paid premium. Benefits are claim payments in case of cancer diagnosis and premium refunds in case of lapse or death. It gives the average amount pays back to the insured for 1 unit of paid premium.

4.7 Scenarios

The three main following scenarios are discussed in this section.

1. Scenario I: base scenario, base assumptions and cancer incidence from 2015 historical data
2. Scenario I*bis*: base scenario, base assumptions and cancer incidence from 2020 forecasted data
3. Scenario II: base scenario, base assumptions and cancer incidence from generation tables, which uses the forecast incidence rates from the previous section using the selected model for each cancer site required for the product design.

The aim to compare the two scenarios is to assess the impact on indicators of considering best-estimate incidence rates including rates forecasting. Assume a product priced with national tables with fixed incidence rates throughout the policy term, what is the impact of not considering the probable derive of incidence rates?

Apart the derive of incidence rates, multiple deviations on assumptions can impact the profitability of a product [54]. Sensibilities are performed using 3 types in assumptions changes, with both one at a time changes and multiples assumptions variations. The considered assumptions variations are:

- Claims: 3 alternative scenarios with a change of claims incidence to evaluate the impacts of changes in future cancer incidence in the insured population
  - Level of claims: Alternative calculations in a favourable scenario with 80% of the base scenario and a unfavourable one with 120%
  - Presence of anti-selection: Alternative scenario in which incidence rates are increased by 50% in year 1 and 25% in year 2.
- Lapse: two alternative scenarios, with a decrease and an increase of lapse rates (2.5% and 7.5%), to evaluate the impacts of changes in insured behavior
- Discount rate: two alternative scenarios, with a decrease and an increase of discount rates (4% and 6%), to evaluate the impacts of variations in the insurer preference for the present

A variation in death incidence rates is not considered, because death occurrence has the same effect as lapse on the product. A variation on expenses assumptions (unit cost, claims cost) are not considered to focus the discussion on technical items.

A key assumption on product profitability is the structure of current portfolio (or sales, if assumed), that is the number of sales and the structure in terms of gender, insured age at inception and sum assured. Variations on the initial population are not considered but the projection will be analyzed by segment.

4.8 Results

4.8.1 Projection Dynamics

We first present below the results in terms of Gross Written Premiums (GWP) and Mathematical Reserves in the Scenario I.

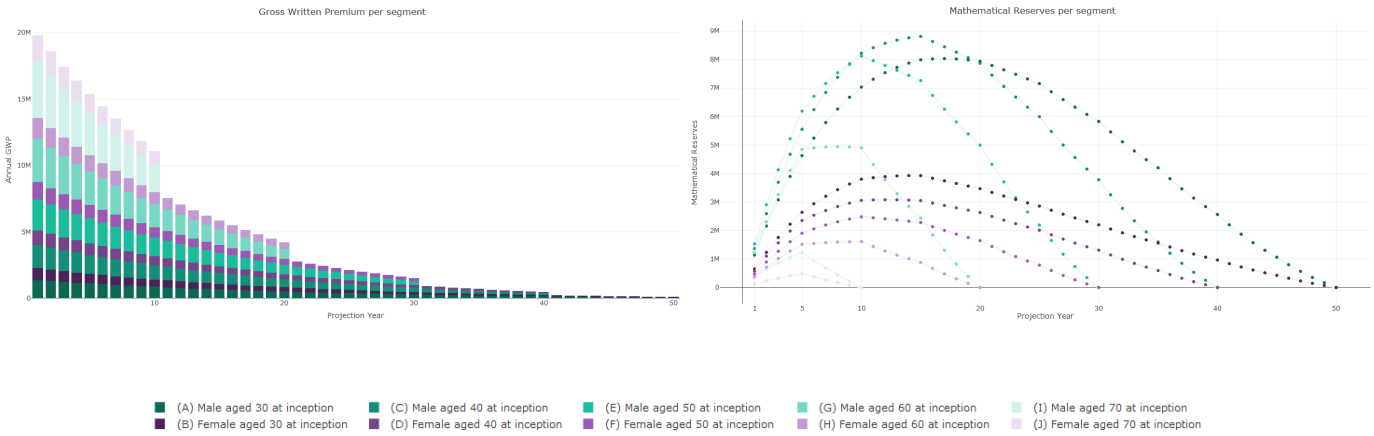


Figure 51: Annual Premiums (left) & Premium Reserves (right) per segment

The left figure represents the GWP through time from the 10 different insured segments. GWP decreases among time, mostly because of lapses and policy maturities, and marginally from general cancer claims which includes a premium waiver. From year 1 to year 10, all the ten segments are in-force. The premiums per policy being bigger for segments (I) and (J) aged 70 at policy inception, those segments represent about a third of the total GWP until year 10 at which the policy term is reached. Then, the GWP jumps down due to the reduced number of policies. Each jump down means that the maturity has been reached for some sales segments.

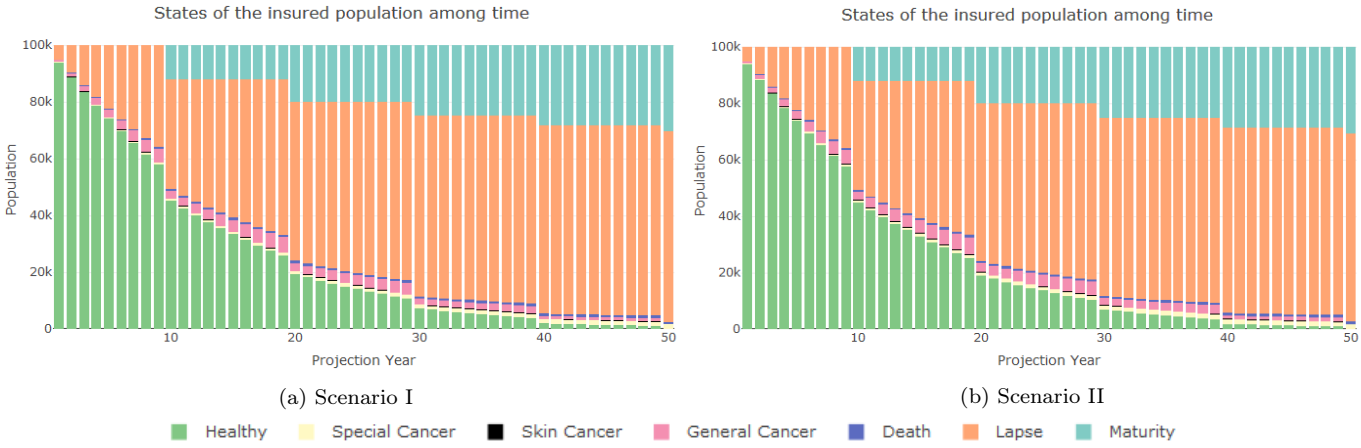
The right figure represents the mathematical reserves per segment. Mathematical reserves are calculated with incidence rates from pricing, not the forecasted incidence rates used to challenge the profitability assessment. For each segment, the mathematical reserves increase during the first years of the coverage and then decrease to 0. This is due to the risk which is concentrated at old ages and then require to put a part of the premium into reserves at the beginning of the policy. Reserves for male are more important compared to female due to two reasons: premiums for male are higher, and the risk for male is more concentrated at old ages.

4.8.2 Comparison between Scenario I & II

This section focuses on comparisons between the scenarios. Detailed results about population dynamics, technical result, profit and premium decomposition are presented. Indicators are presented in the next section.

Population

The following graphs give an overview of the results in the base scenario.



(a) Scenario I

(b) Scenario II

Figure 52: Population states in scenarios I & II

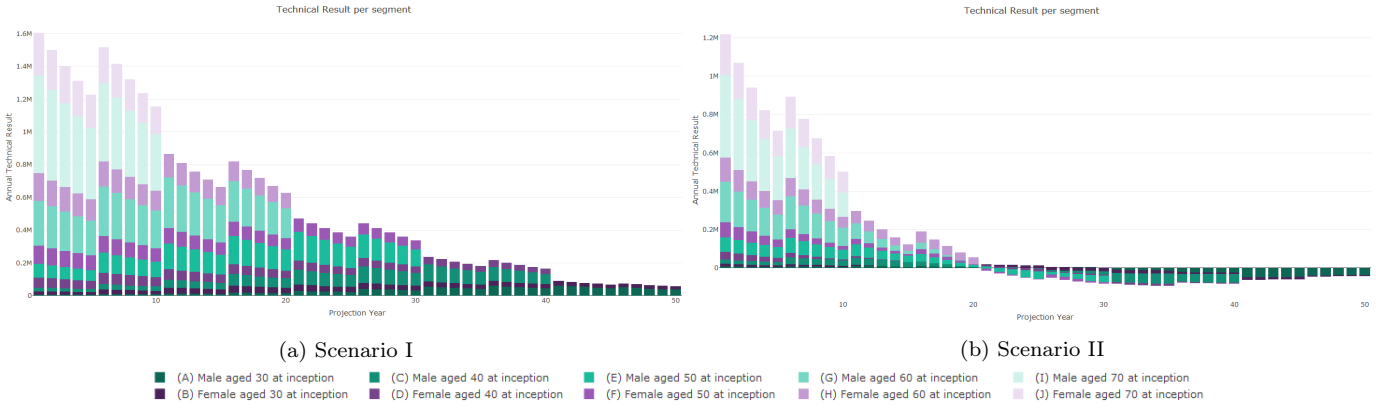
The upper left figure represents the populations of the different segments among time at the end of each projection year for the Scenario I. From inception at which all insured are healthy (in green), the two main drivers of populations evolutions are lapse (in orange) and maturities (in teal). The number of deaths (in blue), which implies a premium reserves refund, is a small driver of the population pattern. Then, the number of insured affected by cancer evolves among time, with the number of insured affected by skin cancer (in black) remain low, so as the number of insured affected by special cancer (in yellow). The main population related to claims is general cancer (in pink), that is the unlapsed insureds diagnosed with cancer all causes except skin cancer. The dynamics are very similar between scenario I & II.

The pattern of the payment population (healthy insureds and skin cancer diagnosed insureds) recall the pattern of the GWP detailed in the previous section. Due to annual lapse rate of 5%, the number of in-force policies at the last year of the projection reaches only 2,115 in scenario I and 2,239 in scenario II (slightly higher in scenario II due to a higher number of general cancer diagnosed, which are premium waived and do not lapse).

The main difference between two scenarios is that claimants are more important in scenario II because higher incidence rates, but incidences rates remains low compare to lapse rates. Therefore the difference is difficult to evaluate graphically.

4.8.3 Technical Result

The following graphs give an overview of the results in the base scenario.



(a) Scenario I

(b) Scenario II

Figure 53: Technical Result in scenarios I & II

The technical result has a decreasing pattern due to the decrease of in-force policies. Older ages at the beginning of the projection contribute largely to the technical result, because corresponding earned risk premiums are important. For young ages at policy inception, technical result is low at the beginning due to low earned risk premium, and remains low throughout the projection in the scenario I due to lapses. In the scenario II, due to the increase of incidence rates, the technical result turns negative from 2041.

The following figure presents the loss ratio (defined as the ratio between claims incurred and earned risk premium) in the different scenarios:

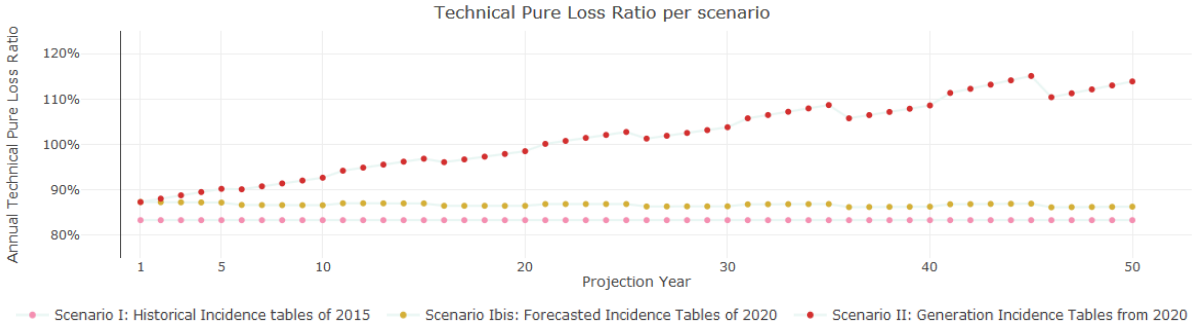
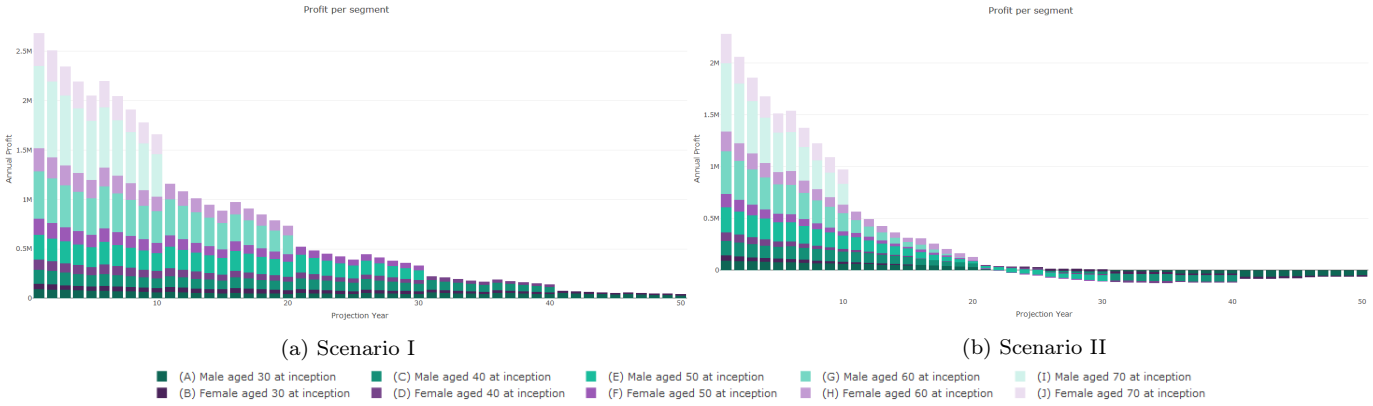


Figure 54: Loss Ratio in scenarios I & II

The pure technical loss ratio in scenario I is constant at 83.3 due to the 20% risk margin. It rises up to 87.3% in the scenario I*bis* with the forecasted incidence rates of 2020. Then, in scenario II, the pure technical loss ratio gradually increases from 87.3% up to 115.1%. The increase of incidence rates has a direct consequence on the increase of claims and then the increase of the loss ratio. While the gap between the two scenarios is very thin in the beginning of the projection, the situation get worse over time in the scenario II. The pure technical loss ratio exceeds 100% in 2041 (turning the technical result negative.

4.8.4 Profit

The following graphs detail the Profit in the two scenarios.



(a) Scenario I

(b) Scenario II

Figure 55: Profit in scenarios I & II

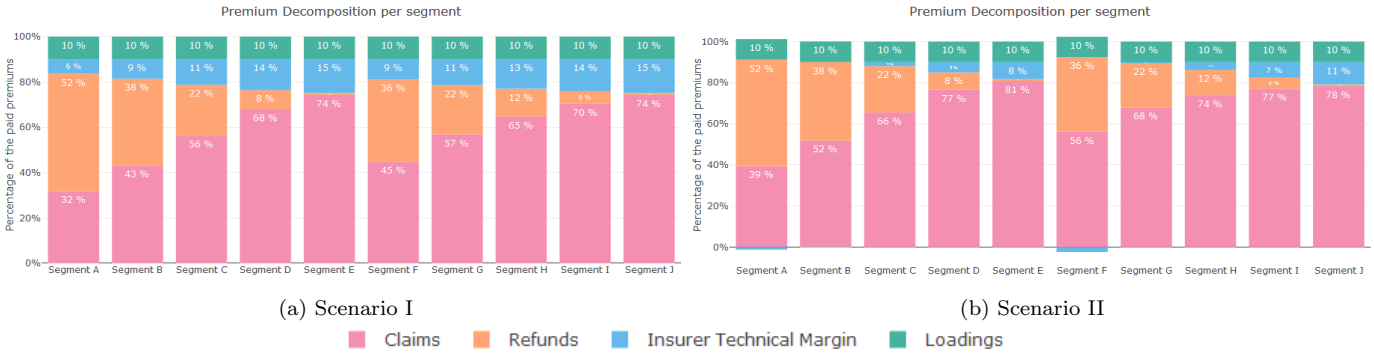
The pattern of the profit is very similar to the technical result. In scenario I, the profit is positive for each year of the projection. The high level of profit then decreases due to the decrease of in-force policies. Recall that the profit is the sum of technical, operational and financial margin. The operational margin is quite low compared to the technical margin. In both scenario, the financial margin is null (no asset return and no technical interest).

While the profit in year 1 is almost equal in the two scenarios, then the profit in scenario II decreases more sharply and eventually turns negative in 2043. It is to years later than the technical result because the operational result remain positive and compensate the technical losses in 2041 and 2042.

The profit remains negative until projection end, with first an increase of losses due to incidence rates increase, and then a decrease due to lapses and maturities. Despite the amounts of negative profits seem small, some lower lapse rates in the beginning of the projection would create worse situation later in the projection. The diverse sensibilities presented in stress test section give alternative scenarios with deep loss due to lower lapse rates.

4.8.5 Premium Decomposition & Customer Value

The following graphs give the premium decomposition per segment, that is the destination of paid premiums:



(a) Scenario I

(b) Scenario II

Figure 56: Premium decomposition in scenarios I & II

The two main items of premium decomposition are claims (in pink) and refunds (in orange). The part of claims in the decomposition increases with insured age at inception, because claims are more likely to occur (due to high incidence rates at old ages) and lapses less important (smaller time horizon to lapse and small mathematical reserves). Due to the assumption on lapses, young insureds mostly lapse before getting diagnosed cancers at old ages. The part of refunds is more important for males, because high incidence rates occur at higher ages and therefore mathematical reserves are higher.

The incidence rates increase in scenario II imply more claims and therefore the part of claims in the premium decomposition increases. The part of the refunds is almost equal. The only difference, not visible in the figure is that skin and general claimants do not lapse, whereas the number of special cancers slightly differs between the two scenarios.

The consequence of the increase of claims is the decrease of the insurer technical margin, which turns below 0 in scenario II for segments A and F aged 30 at inception and about 0 of segments B and G aged 40 at inception. Young ages have a longer guaranty term and are therefore more affected by the increase of cancer incidence until portfolio extinction.

Indicators for Scenario I: Constant Incidence Rates

The key main indicators are as follows:

Indicators	Value
Number of Claims	12,756 (10,894 General Cancer, 212 Skin Cancer, 1,650 Special Cancer)
Customer Value	78.3%
Pure Technical Loss Ratio	83.3%
EPV Technical Result	15.8 million
EPV Operational Result	7.1 million
EPV Profit (PVFP)	22.9 million

Most of claims are general cancer claims, and about 10% of the initial population will eventually be a claimant for such claims. The number of skin cancer is very limited. Recall that if a skin cancer is diagnosed after a general cancer, no additional benefit are paid. Due to 20% margin on earned risk premium, the pure technical loss ratio is 83.3%. Due to loadings, the customer value is smaller to the loss ratio. The product is profitable, with positive both technical and operational result.

Indicators for Scenario Ibis: Forecasted 2020 Incidence Rates

Generation incidence rates includes forecasted incidence rates following the methodology and results presented in the previous sections.

Indicators	Value
Number of Claims	13,282 (11,330 General Cancer, 246 Skin Cancer, 1,706 Special Cancer)
Customer Value	80.8%
Pure Technical Loss Ratio	86.8%
EPV Technical Result	12.4 million
EPV Operational Result	6.9 million
EPV Profit (PVFP)	19.3 million

Indicators for Scenario II: Forecasted 2020 and after Incidence Rates

Generation incidence rates includes forecasted incidence rates following the methodology and results presented in the previous sections.

Indicators	Value
Number of Claims	14,612 (12,366 General Cancer, 383 Skin Cancer, 1,863 Special Cancer)
Customer Value	86.6%
Pure Technical Loss Ratio	95.1%
EPV Technical Result	7.0 million
EPV Operational Result	6.6 million
EPV Profit (PVFP)	13.6 million

The number of claims logically increased. General cancer & special cancer claims increase by +10%, while skin cancer increases by 60% and therefore customer value, pure technical loss ratio increase and the technical result decreases. Because claims are more likely in scenario II despite the same amount of premium per insured, the situation of scenario II is more favourable to the insured. The operational result slightly decreases due to more premium waivers and more claims cost. However the main driver of the decrease of the profit is clearly the technical result.

4.8.6 Analysis of Change from Historical Incidence to Generation Tables

This section details the different steps from scenario I to II with base assumptions, introduction sub-steps to measure the impact of each change in the cancer incidence tables used in the profitability assessment.

The analysis of change aims to evaluate the two main effects:

- The time gap between estimation time period and projection start date
- The drift of incidence rates from static to generation incidence tables. Because multiple generation incidence tables are used (general, skin and special cancer), this effect is sub-divided in 3 steps.

In the initial step, the scenario I, the present value of future profit is 22.9 million. The step 2, in which incidence rates used for the projection are no longer 2015 historical rates but 2020 forecasted rates corresponds to scenario I*bis*. Because forecasted rates are higher than historical rates, the number of claims increases leading to 3.6 million profit decrease.

Then, in step 3, generation tables are used for general cancer, while skin and special cancer rates are still 2020 forecasted rates. It leads to 5.8 million profit decrease, and it is the major impact in the analysis of change. In step 4, skin cancer rates move to generation tables. The impact on profit is positive, because skin cancer increases mean general cancer decreases. Finally, in step 5, special cancer rates use generation tables and the impact is a 0.5 million profit decrease and the final profit level of 13.6 million is reached.

Graphically,

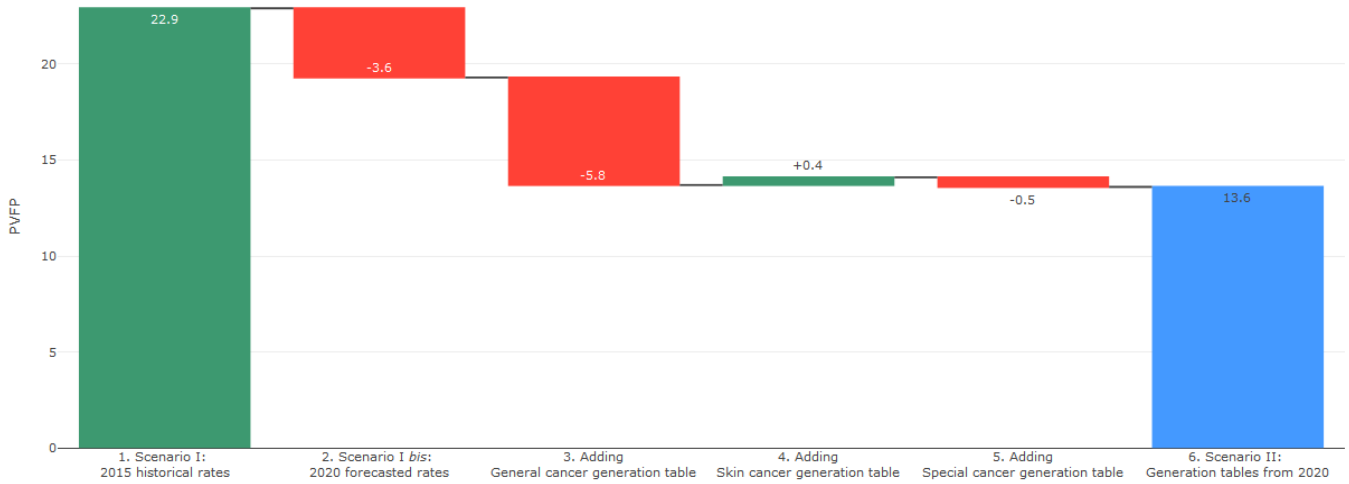


Figure 57: Analysis of change of the PVFP depending incidence rates table

4.9 Sensibilities

The following tables present the profit for each sensibility in the base assumptions for scenarios I and II. The profits for scenarios I and II previously presented (respectively 22.9 and 13.6 million) are highlighted in green:

4.9.1 Scenario I: Constant Incidence Rates

Scenario I - Sensibilities	No Anti-Selection			Anti-Selection		
Lapse\Claims	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
2.5%	50.2	28.5	7.5	45.2	22.3	0.3
5.0%	36.2	22.9	6.5	35.1	16.9	-0.7
7.5%	33.1	19.2	5.7	28.2	13.2	-1.4

The profit decreases with lapse rate, because the loss-ratio being constant at 83.3%, the more policies, the more profit. For example, the profit is 19.2 million changing the annual lapse from 5.0 to 7.5%. The variations on the general level of claims strongly impact the profit. A variation from a claim level of 100% of the pricing assumptions to 120% leads the profit to 6.5 million. In that particular scenario and following the 20% risk margin pricing assumption, the technical result is null.

In absence of anti-selection, the profit is positive for all sensibilities; in presence of anti-selection, the profit turns negative in the scenarios with the higher level of claims (-0.7 million in case of a standard lapse rate to -1.4 million in case of a high lapse rate). In that cases, the losses supported the t first years are not compensated by the future profits.

4.9.2 Scenario II: Generation Incidence Rates

Scenario II - Sensibilities	No Anti-Selection			Anti-Selection		
Lapse\Claims	**80%**	**100%**	**120%**	**80%**	**100%**	**120%**
2.5%	38.8	15.0	-7.9	33.6	8.6	-15.3
5.0%	32.2	13.6	-4.2	27.0	7.3	-11.7
7.5%	27.4	12.3	-2.3	25.6	6.0	-9.7

The general level of profit is smaller in all sensibilities of scenario II compared to corresponding sensibilities of scenario I. The profit turns negative with a high level of claims (from -2.3 million in case of a high lapse rate to -9 million in case of a low lapse rate in absence of anti-selection, and from -9.7 million to -15.3 million in presence of anti-selection). In the case of an unfavourable level of claims, the business is not profitable (at least from a future point in the projection horizon), and then it is the losses which increase with lapse, oppositely to the scenario I.

4.10 Discussion

The use of forecasted incidence rates has a strong impact on results. Due to the expected increase in the future incidence rates, the profitability deteriorates. The general level on claims is obviously a key component for profitability, and can turn a supposed profitable product to a non-profitable one. This impact is stronger in case of forecasting incidence rates.

The presence of anti-selection, modelled here by an increase of the level of claims in the first two years, is an additional factor which deteriorates the profit. The impact is particularly important because it occurs at a time when most of the policies are in-force, and only a few premiums have been paid when such claims occur. While it can be partially mitigated by a waiting period or limited benefits at the beginning of the coverage period, it remains an important challenge for such products.

Lapse rates are a central element of product profitability, especially due to the long-term coverage. A constant annual lapse rate has been assumed previously. However, lapse rates usually depends on policy seniority. Moreover, in case of an increase of incidence rates at national level, one may assume that lapse rates may decrease, because insured would have less interest to stop the policy. A market innovation (in terms of product design, or cancer detection) can have a strong impact on lapses.

The presented product is voluntary simple in terms of loadings. Market practices includes multiple types of loadings, including acquisition costs loadings, administration costs loadings and commissions loadings. Commissions schemes can involve complex mechanisms, with multiple payments and claw-back commissions in case of early lapse. Such elements may have an impact on profitability.

Some traditional items, such as financial revenue have been omitted in the calculation. While the general levels of financial rates and expected financial revenue are very low in Europe or in Japan, it may not be true for other parts of the world. Due to the option to lapse and the need to pay benefits, a part of assets must be liquid. Still, financial result (or a part of this result in case of a profit sharing mechanism) can be an additional revenue for the insurer to improve a product profitability.

Different assumptions of the projection or the sensibilities can be challenged. In case of the evaluation of a future sales plan, the sales assumptions (in terms of number of sales per age and gender) are by nature very hypothetical and sensibilities must be performed. Unit cost and claims costs assumptions can be challenged to perform sensibilities on expenses, which may deteriorate the operational result. Historical experience may be used to make assumptions and sensibilities levels.

Conclusion

Once a so-called incurable disease, Cancer detection and cancer treatments have dramatically progressed in the past decades. Survival rates have increased for most cancer sites. However, cancer cost has also gradually increased, and despite public or private health coverages, out-of-pocket payments remain a burden for some patients.

Cancer insurance, as products including lump-sum in case of cancer diagnosis, is unequally developed in the different markets. While it is a rare coverage in France, it has been widely spread in Asia, and especially in Korea and Japan with long-term guarantees. A pricing model is exposed in which multiple guaranties can be nested and is illustrated with a 3 insured events: one accelerated benefit for a small cancer, one main benefit for a general cancer and one benefit for a special cancer, which terminates the policy. It is then possible to determine different guarantees to personalize the product design.

Multiple databases about historical incidence data are available. A database from Korea (with year range 1999-2017, data from KOSIS) and one from Japan (year range 1975-2015, data from NCC) are evaluated. The analysis show strong differences in the composition of cancer incidence between the two countries, especially with the high incidence of Thyroid Cancer in Korea. While the data is relatively rich in terms of Sites covered and the historical depth for Japan, it remains relatively small databases in terms of available historic. Also, some inconsistencies have been spotted in the data.

VAR models, Neural Networks (with and without external regressors) and Lee-Carter models are used in order to forecast incidence rates per site. This variety of models allows to compare different approaches in order to select the optimal model through a nested cross-validation procedure. A common challenge is the ability to provide long-term forecasts despite small tests samples. Model selection on Korean data for cancer all causes leads to constant cancer incidence rates in the future. Model selection on Japanese data for cancer all causes leads to an expected increase on future cancer incidence rates following forecasting with Sparse VAR model for both male and female.

Data could be enriched or replaced by insurer in-house data, in order to have a better fit to the current or target population of an insurer. Ideally, the data would include incidence rates per age, dependencies between the difference cancer site diagnosis, and more historical data. However, the data is hard to find, and the used data provides respectable quality for forecasting cancer incidence rates.

The numerical illustrations show a significant impact on profitability of the incidence rates forecast. A decision-making without such information may overly optimistic about product profitability. The discussion evaluates the uncertainty on claim level, due to adverse selection and the general level of claims. Such parameters have a also strong impacts on profitability and variations can turn expected profits into veritable losses.

Overall, cancer insurance aims to reduce financial burden after a cancer diagnosis. It is also a tool to raise cancer awareness and then accompany the insureds throughout their journey. The development of cancer insurance products is then an useful additional tool in fighting cancer.

References

[1] F. Baillet (2015), *Cancérologie*, Polycopié de l'Université Pierre et Marie Curie, 2, 631-633, http://www.chups.jussieu.fr/polys/cancero/

[2] J. H. Breasted (1930), The Edwin Smith Surgical Papyrus, Volume 1: Hieroglyphic Transliteration, Translation, and Commentary, Chicago: The University of Chicago Press, ISBN:978-0-918986-73-3

[3] B. Weinstein and K. Case (2008), The History of Cancer Research: Introducing an AACR Centennial Series I.,*Cancer Res September*, 1 (68) (17) 6861-6862, doi:10.1158/0008-5472.CAN-08-2827

[4] A. S. Ahmad, N. Ormiston-Smith and P. D. Sasieni (2018), Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960, *British Journal of Cancer*, Mar 3; 112(5): 943–947, doi:10.1038/bjc.2014.606

[5] IARC (2018), Les cancers attribuables au mode de vie et à l'environnement en France métropolitaine. Lyon: International Agency for Research on Cancer.

[6] Inserm (2008), Cancer et environnement. Rapport. Paris: Les éditions Inserm, XVII, http://hdl.handle.net/10608/102

[7] C. A. Thompson, S. L. Gomez, K. G. Hastings and al. (2016), The burden of cancer in Asian Americans: a report of national mortality trends by Asian ethnicity. *Cancer Epidemiology, Biomarkers & Prevention*, 25(10), 1371–1382, doi:10.1158/1055-9965.EPI-16-0167

[8] E. M. Grindedal, C. Heramb, I. Karsrud and al. (2017), Current guidelines for BRCA testing of breast cancer patients are insufficient to detect all mutation carriers, *BMC Cancer*, 17:438, doi:10.1186/s12885-017-3422-2

[9] C. Fiala and E. P. Diamandis (2020), A multi-cancer detection test: focus on the positive predictive value, *Annals of Oncology*, Sep;31(9):1267-1268, doi:10.1016/j.annonc.2020.04.013

[10] J. D. Cohen, L. Li, Y. Wang and al. (2020), Detection and localization of surgically resectable cancers with a multi-analyte blood test, *Science*, Vol. 359, Issue 6378, pp. 926-930, doi:10.1126/science.aar3247

[11] J. M. G. Wilson and G. Jungner (1968), Principles and practice of screening for disease, World Health Organization, https://apps.who.int/iris/handle/10665/37650

[12] Y. Kim, J. K. Jun, K. S. Choi and al. (2011), Overview of the National Cancer Screening Programme and the Cancer Screening Status in Korea, *Asian Pacific Journal of Cancer Prevention*, 12, 725-730, PMID:21627372

[13] M. Suh, S. Song, H. N. Cho and al. (2016), Trends in Participation Rates for the National Cancer Screening Program in Korea, 2002-2012, *Cancer Res Treat.* 2017;49 (3): 798-806., doi:10.4143/crt.2016.186

[14] C. Hamashima (2018), Cancer screening guidelines and policy making: 15 years of experience in cancer screening guideline development in Japan, *Japanese Journal of Clinical Oncology*, Volume 48, Issue 3, March 2018, 278–286, doi:10.1093/jjco/hyx190

[15] Foundation of Promotion of Cancer Research (2019), *Cancer Statistics in Japan - 2018* https://ganjoho.jp/en/professional/statistics/brochure/

[16] M.S. Sabel, K. Diehl and A. Chang (2006), Principles of Surgical Therapy in Oncology, *Oncology: An Evidence-Based Approach*, Springer, doi:10.1007/0-387-31056-8_4

[17] R. Baskar, K. A. Lee, R. Yeo and al. (2012), Cancer and Radiation Therapy: Current Advances and Future Directions, *International Journal of Medical Science*, 9(3): 193-199, doi:10.7150/ijms.3635

[18] American Cancer Society (2019), Cancer Facts & Figures 2019. Atlanta: American Cancer Society

[19] M. Arnold, M. J. Ruthrford, A. Bardot and al. (2019), Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study, *The Lancet Oncology*, Volume 20, Issue 11, 1493-1505, doi:10.1016/S1470-2045(19)30456-5

[20] W. Jones, G. Allardice, I. Scott and al. (2017), Cancers of unknown primary diagnosed during hospitalization: a population-based study, *BMC Cancer*, doi:10.1186/s12885-017-3083-1

[21] S. Kim, D. W. Shin, H. K. Yang and al. (2016), Public Perceptions on Cancer Incidence and Survival: A Nation-wide Survey in Korea, *Cancer Res Treat.*, 48(2):775-788, doi:10.4143/crt.2014.369

[22] A. B. Mariotto, K. R. Yabroff, Y. Shao and al. (2011), Projections of the Cost of Cancer Care in the United States: 2010–2020. *JNCI Journal of the National Cancer Institute*, 103(2), 117–128, doi:10.1093/jnci/djq495

[23] Caisse nationale d'assurance maladie (2019), Cartographie médicalisée des dépenses de santé, https://www.ameli.fr/fileadmin/user_upload/documents/DP_Cartographie_des_depenses_et_des_pathologies.pdf

[24] Ligue nationale contre le cancer (2008), Enquête sur les frais restant à charge des malades atteints de cancer : constats, ressenti et propositions, https://www.ligue-cancer.net/article/download/3907

[25] A. C. Dash and D. L. Grimshaw (1993), Dread Disease Cover - An Actuarial Perspective, *Journal of the Staple Inn Actuarial Society*, 33(1), 149–193, doi:10.1017/S2049929900010564

[26] S. Golberg and Y. Ren (2015), *Critical Illness: At a critical stage in Asia*, SOA Asia-Pacific Symposium

[27] C. Cypris and K. de Braaf (2013), Critical Illness in Asia – A success story, Presentation to the Institute and Faculty of Actuaries

[28] C. L. Bennett CL, P. D. Weinberg and J. J. Lieberman (1998), Cancer insurance policies in Japan and the United States, *West J Med.*, 168(1): 17–22.

[29] Asia Pacific Observatory on Public Health Systems and Policies (2015), *Republic of Korea - Health System Review*

[30] Asia Pacific Observatory on Public Health Systems and Policies (2018), Japan Health System Review, *Health Systems in Transition*, Vol. 8, No. 1

[31] The Life Insurance Association of Japan (2018), *Life Insurance Business in Japan 2017-2018*

[32] Japan Institute of Life Insurance (2019), 生活保障にする調査《速報版》(Survey on Life Guarantees «Flash report ») https://www.jili.or.jp/research/report/pdf/r1hosho.pdf

[33] Aflac (2019), *Aflac's Financial Analysts Briefing*

[34] D. Dickson, M. Hardy and H. Waters (2009), *Actuarial Mathematics for Life Contingent Risks*, Cambridge, U.K.: Cambridge University Press

[35] H. Gerber (1997), *Life Insurance Mathematics*, Springer, Third Edition, ISBN: 3-540-62242-X

[36] A. S. MacDonald (2006), Euler–MacLaurin Expansion and Woolhouse's Formula, *Encyclopedia of Actuarial Science*, 2, 631-633

[37] K. Wang, J. L. Peng, Y. Y. Sun and al. (2011), The Asymmetric Information Problem in Taiwan's Cancer Insurance Market, *The Geneva Risk and Insurance Review*, 36: 202, doi:10.1057/grir.2010.12

[38] H. S. Ahn, H. J. Kim and H. G. Welch (2000), Korea's Thyroid-Cancer "Epidemic" — Screening and Overdiagnosis, *N. Engl. J. Med.*, 371:1765-1767, doi:10.1056/NEJMp1409841

[39] H. Li and Y. Lu (2016), Coherent Forecasting of Mortality Rates: a Sparser Vector-Autoregression Approach, *ASTIN Bulletin, Cambridge University Press (CUP)*, 47 (2), pp.563-600, doi:10.1017/asb.2016.37

[40] Q. Guibert, O. Lopez and P. Piette (2019), Forecasting mortality rate improvements with a high-dimensional VAR, *Insurance. Mathematics and Economics*, vol. 88, p. 255-272, doi:10.1016/j.insmatheco.2019.07.004

[41] D. Hainault (2018), A Neural-Network Analyzer for Mortality Forecast, *ASTIN Bulletin*, Volume 48, Issue 2, pp. 481-508, doi:10.1017/asb.2017.45

[42] L. Rahib, B. D. Smith, R. Aizenberg and al. (2014), Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States, *Cancer Research*, CAN-14-0155, doi:10.1158/0008-5472

[43] Australian Institute of Health and Welfare (2012), Cancer incidence projections: Australia 2011 to 2020, ISBN:978-1-74249-282-7

[44] C. R. Smittenaar, K. A. Petersen, K. Stewart and al. (2016), Cancer incidence and mortality projections in the UK until 2035, *British Journal of Cancer*, 115,9 (2016): 1147-1155, doi:10.1038/bjc.2016.304

[45] X. Q. Yu, Q. Luo, S. Hughes and al. (2019), Statistical projection methods for lung cancer incidence and mortality: a systematic review, *BMJ Open*, doi:10.1136/bmjopen-2018-028497

[46] J. Stock and M. Watson (2007), Introduction to Econometrics, 2nd edition, *Pearson*, ISBN-13:9780321278876

[47] E. Zivot and J. Wang (2003), Vector Autoregressive Models for Multivariate Time Series, in Modeling Financial Time Series with S-Plus®, *Springer*, New York, NY

[48] W. Nicholson and D. Matteson (2017), BigVAR: Tools for Modeling Sparse High-Dimensional Multivariate Time Series, arXiv preprint, arXiv:1702.07094

[49] S. F. Crone and N. Kourentzes (2010), Feature selection for time series prediction – A combined filter and wrapper approach for neural networks, *Neurocomputing*, 73(10–12), 1923–1936, doi:10.1016/j.neucom.2010.01.017

[50] F. Günther and S. Fritsch (2010), neuralnet: Training of Neural Networks, *The R Journal*, Vol. 2/1, doi:10.32614/RJ-2010-006

[51] R. Lee and L. Carter (1992), Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87(419), 659-671, doi:10.2307/2290201

[52] The Institute of Actuaries in Japan (2018), Development Process of Standard Mortality Table 2018, http://www.actuaries.jp/english/lib/

[53] D. S. Promislow (2011), *Fundamentals of Actuarial Mathematics*, 2nd Edition, John Wiley & Sons, ISBN:978-0-470-97807-8

[54] I. C. Smart (1977), Pricing and profitability in a life office, *Journal of the Institute of Actuaries*, 104, pp. 125-172, doi:10.1017/S0020268100018187

A Appendix A: Age-Standardised Rate Definition and Data

A.1 Age-Standardised Rate (ASR)

The IARC definition of ASR is as follow [60]:

An age-standardised rate (ASR) is a summary measure of the rate that a population would have if it had a standard age structure. Standardization is necessary when comparing several populations that differ with respect to age because age has a powerful influence on the risk of dying from cancer. The ASR is a weighted mean of the age-specific rates; the weights are taken from population distribution of the standard population. The most frequently used standard population is the World Standard Population. The world standard population used within the application is as proposed by Segi (1960) and modified by Doll and al. (1966). The ASR is also expressed per 100,000.

The distribution is done applying weights defined as follows:

Age group	**Weight**
0-4	12,000
5-9	10,000
10-14	9,000
15-19	9,000
20-24	8,000
25-29	8,000
30-34	6,000
35-39	6,000
40-44	6,000
45-49	6,000
50-54	5,000
55-59	4,000
60-64	4,000
65-69	3,000
70-74	2,000
75-79	1,000
80-84	500
85+	500
Total	100,000

Table 23: Age distribution of the world standard population

Denoting by w_i the weight and q_i the cancer incidence rate for age group i, the age-standardised rate is then $ASR = \frac{\sum_i w_i \cdot q_i}{\sum_i w_i}$.

[60] http://www-dep.iarc.fr/WHOdb/glossary.htm

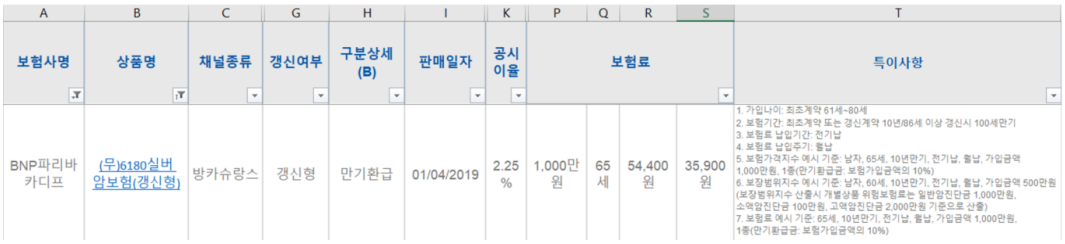
A.2 World Incidence Rates per Gender using ASR

The world incidence rates from the WHO per gender and per site using ASR in 2018 are compiled in the table below. The figures have been used to generate the figure in section 1.2.2 Cancer classification per site.

ICD Code	Cancer Type	Incidence rate for male	Incidence rate for female
C00-97	**All cancers**	**245.0**	**228.0**
C00-06	Lip & oral cavity	6.4	2.9
C07-08	Salivary glands	0.8	0.6
C09-10	Oropharynx	1.9	0.49
C11	Nasopharynx	2.4	0.9
C12-13	Hypopharynx	1.8	-
C15	Esophagus	10.4	4.6
C16	**Stomach**	**17.8**	**9.3**
C18-21	**Colorectum**	**26.6**	**21.8**
C22	**Liver**	**15.5**	**6.5**
C23-24	Gallbladder	2.5	3.2
C25	Pancreas	6.3	5.7
C32	Larynx	4	0.6
C33-34	**Lung**	**35.5**	**19.2**
C43	Melanoma of skin	3.9	3.6
C45	Mesothelioma	0.6	-
C46	Kaposi sarcoma	0.7	-
C50	**Breast**	**-**	**55.2**
C51	Vulva	-	1.2
C52	Vagina	-	0.5
C53	**Cervix uteri**	**-**	**15.1**
C54	Corpus uteri	-	10.1
C56	Ovary	-	7.8
C60	Penis	0.9	-
C61	**Prostate**	**33.1**	**-**
C62	Testis	1.8	-
C64-65	Kidney	6.6	3.9
C67	Bladder	11.0	3.3
C70-72	Brain & nervous system	4.2	3.6
C73	**Thyroid**	**3.4**	**11.5**
C81	Hodgkin lymphoma	1.2	0.9
C82-86, C96	Non-Hodgkin lymphoma	7.4	5.9
C88+C90	Multiple myeloma	2.3	1.9
C91-95	Leukaemia	6.5	5.0

B Appendix B: Korean Product Disclosure Data

Disclosure data from Korean national insurance associations are available online[61]. Data for all products, including cancer products, are available in Excel workbook file format. One row denotes one insurance product, but similar products with different distribution channel will lead to multiple rows. An extract of this database is presented below:



A	B	C	G	H	I	K	P	Q	R	S	T
보험사명	상품명	채널종류	갱신여부	구분상세 (B)	판매일자	공시 이율	보험료				특이사항
BNP파리바 카디프	(무)6180실버 암보험(갱신형)	방카슈랑스	갱신형	만기환급	01/04/2019	2.25 %	1,000만 원	65 세	54,400 원	35,900 원	1. 가입나이: 최초계약 61세~80세 2. 보험기간: 최초계약 또는 갱신계약 10년/86세 이상 갱신시 100세만기 3. 보험료 납입기간: 전기납 4. 보험료 납입주기: 월납 5. 보험가격지수 예시 기준: 남자, 65세, 10년만기, 전기납, 월납, 가입금액 1,000만원, 1종(만기환급금: 보험가입금액의 10%) 6. 보장범위지수 예시 기준: 남자, 60세, 10년만기, 전기납, 월납, 가입금액 500만원 (포장범위지수 산출시 가별상품 위험보험료는 일반암진단금 1,000만원, 소액암진단금 100만원, 고액암진단금 2,000만원 기준으로 산출) 7. 보험료 예시 기준: 65세, 10년만기, 전기납, 월납, 가입금액 1,000만원, 1종(만기환급금: 보험가입금액의 10%)

Figure 58: Extract of disclosure data for cancer products from life insurers

with:

* (A) 보험사명: Company name (in the example: BNP Paribas Cardif)
* (B) 상품명: Product name (in the example: 61-80 Silver Insurance Cancer)
* (C) 채널종류: Distribution channel (in the example: Bancassurance)
* (G) 개신여: Possibility to renew (in the example: renewable)
* (H) 구분상세(B): Classification (in the example: maturity benefit)
* (I) 판매일자: Date of sale (in the example: April 1st, 2019)
* (K) 공시이률: Technical rate (in the example: 2.25%)
* (P-S) 보험료: Premium amount, and the four columns denote the sum insured, the insured age at policy inception and the premium for male and female
* (T) 특이사항: Particular items: provides a summary of the product design

This data has been used to derive the figures in the Korean market study about products per insurer, age range at inception and possible policy terms, as well as premium amount illustrations. The different inception age ranges have been derived from column (T) Particular items. A scrapping using the age character 세 (age) was first used following by a one by one check. More details about product characteristics can be found in the different insurers' websites.

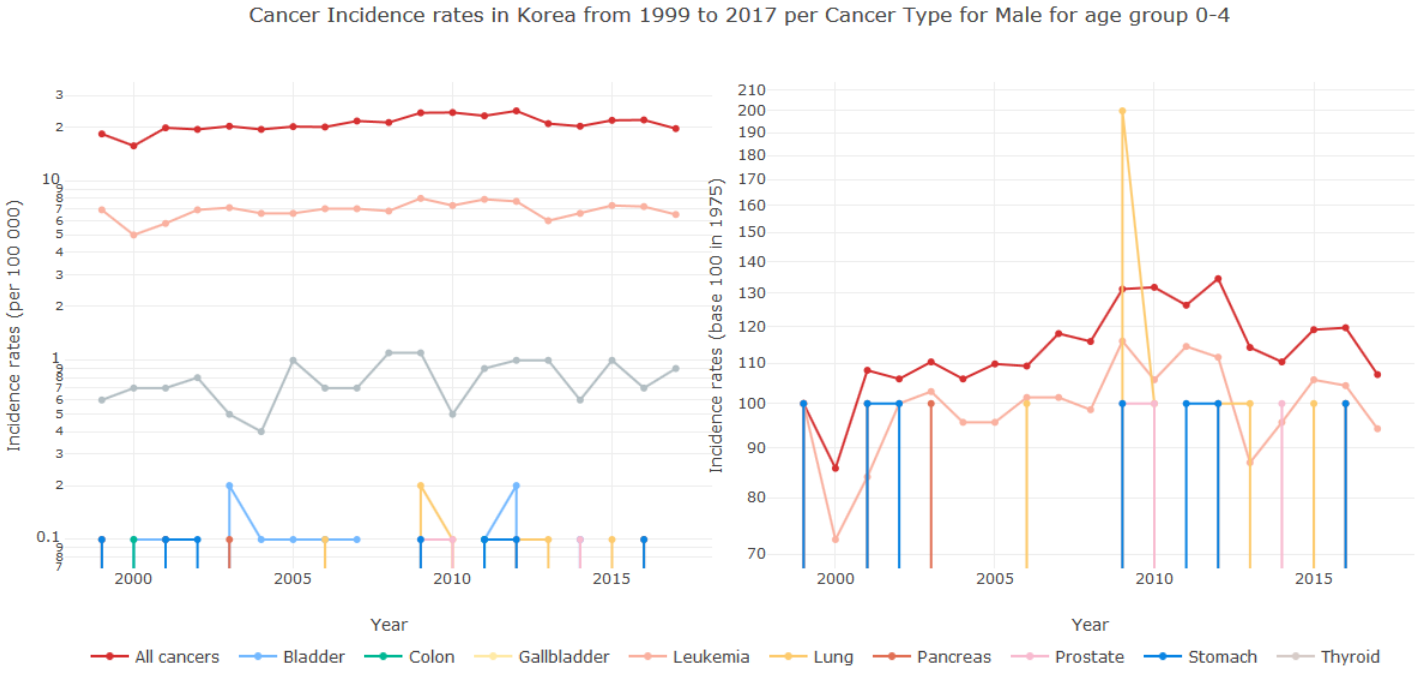
[61] `https://pub.insure.or.kr` for Life insurers
`http://kpub.knia.or.kr` for Non-Life insurers

C Appendix C: Korea Cancer Incidence Data

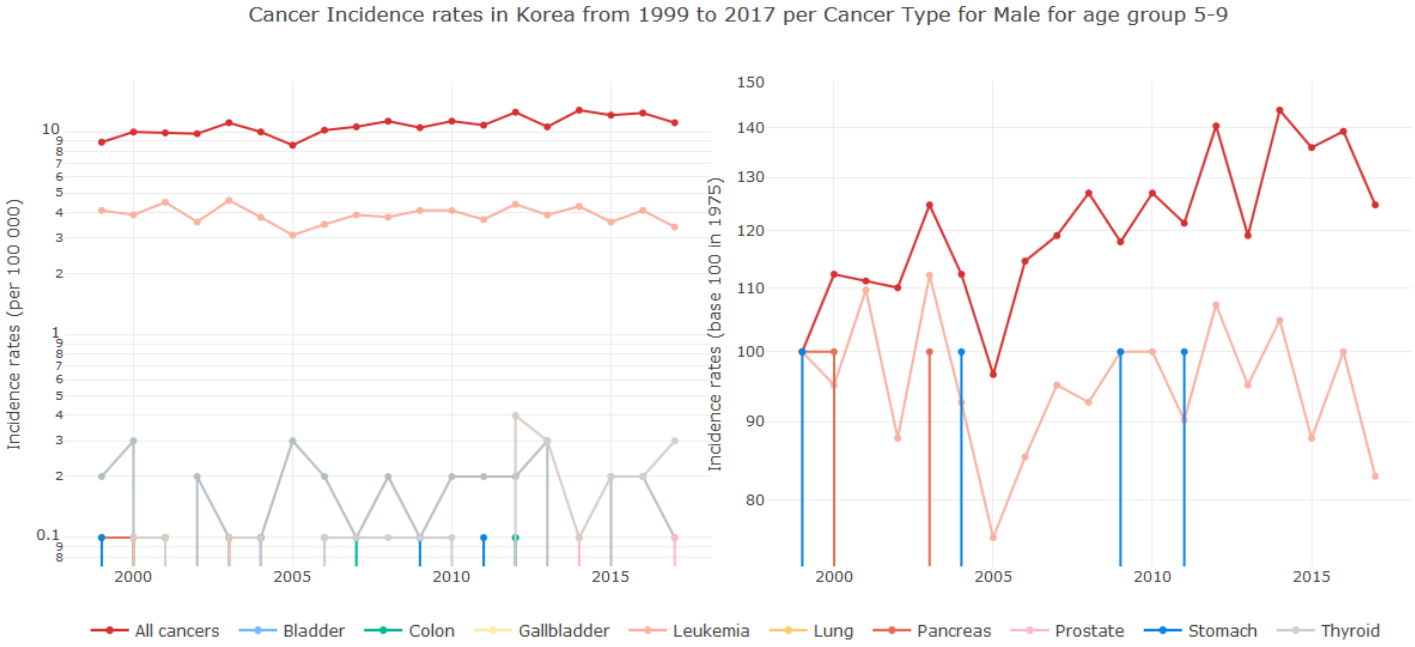
This section illustrates the database with cancer incidence rates per gender and age group. For each age group, two figures have been generated: on left, the crude incidence rates and on right, the index rates assuming an index 100 in 1999. For both figures, total incidence rates as well as 11 cancer site are presented. Due to the huge differences in incidence rates, the y-axis is using a logarithmic scale.

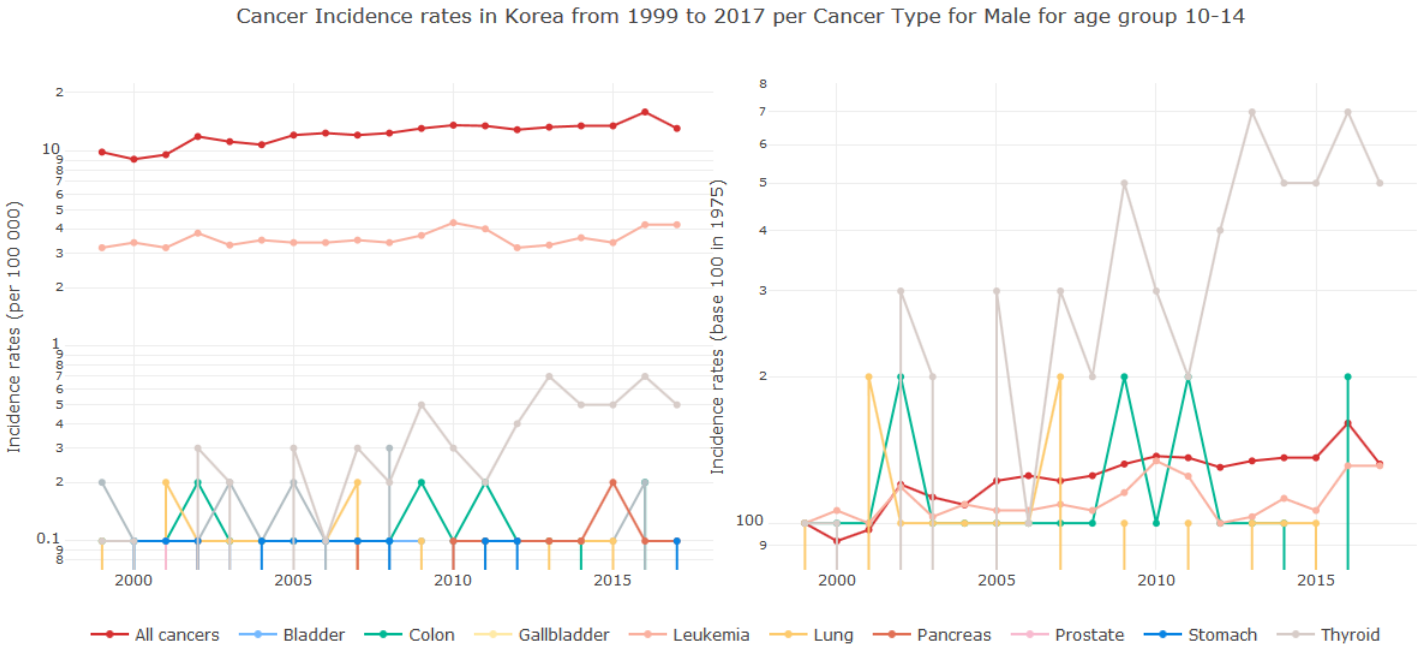
For the age group 85+, it should be analyzed remembering that the average age may not be stable along time, while it would be an acceptable assumption for other age classes. Due to the life expectancy rise, the average age of people aged more than 85 may not be constant over time.

C.1 Korean Data - data per age group for male

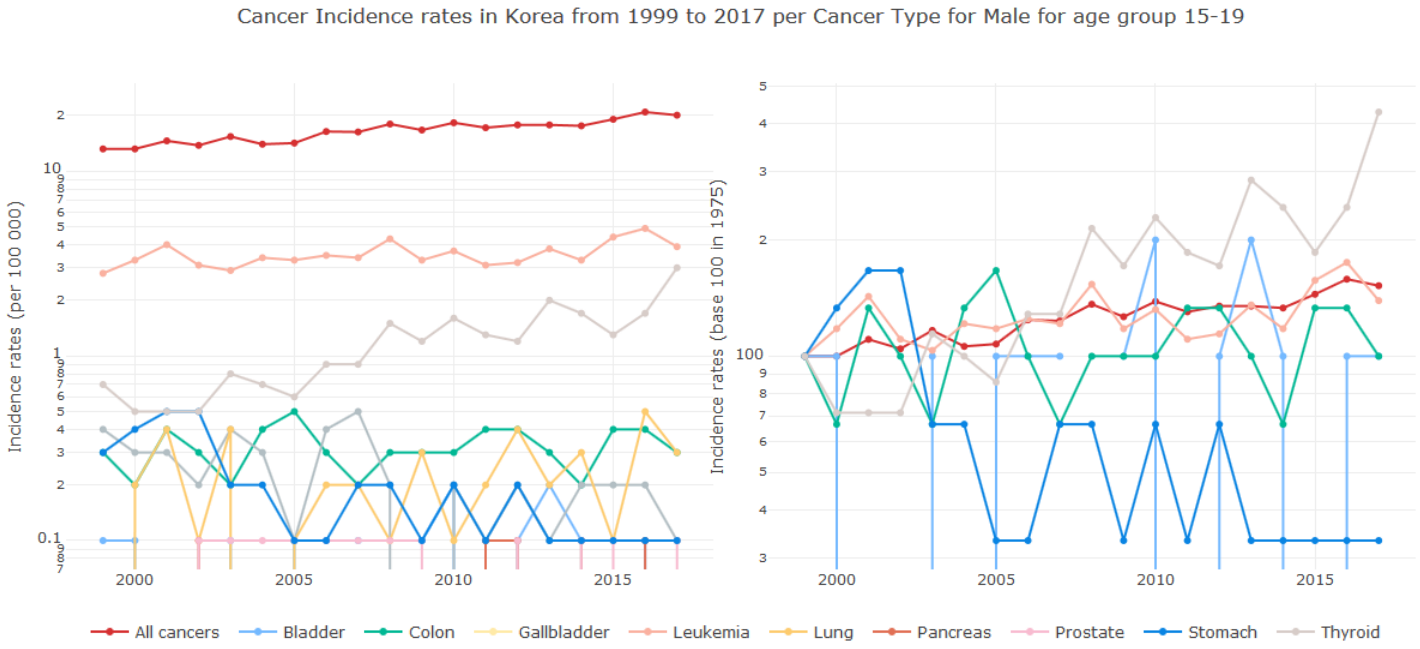


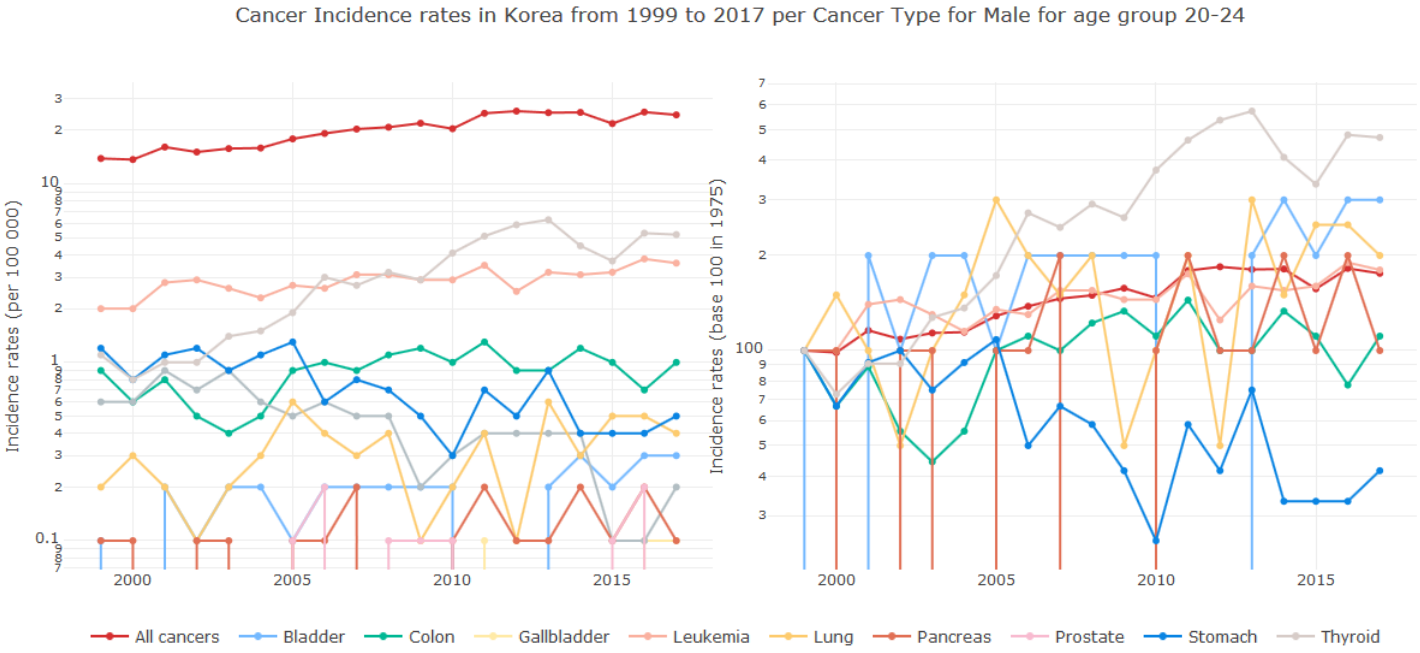
For male children aged 0 to 4, Cancer all causes annual incidence rate has been around 2 for 100,000. The most common cancer is leukemia and the second one thyroid cancer. Other types are very rare while unknown cancer (not represented in the figures but can be deduced from existing values) is quite important.



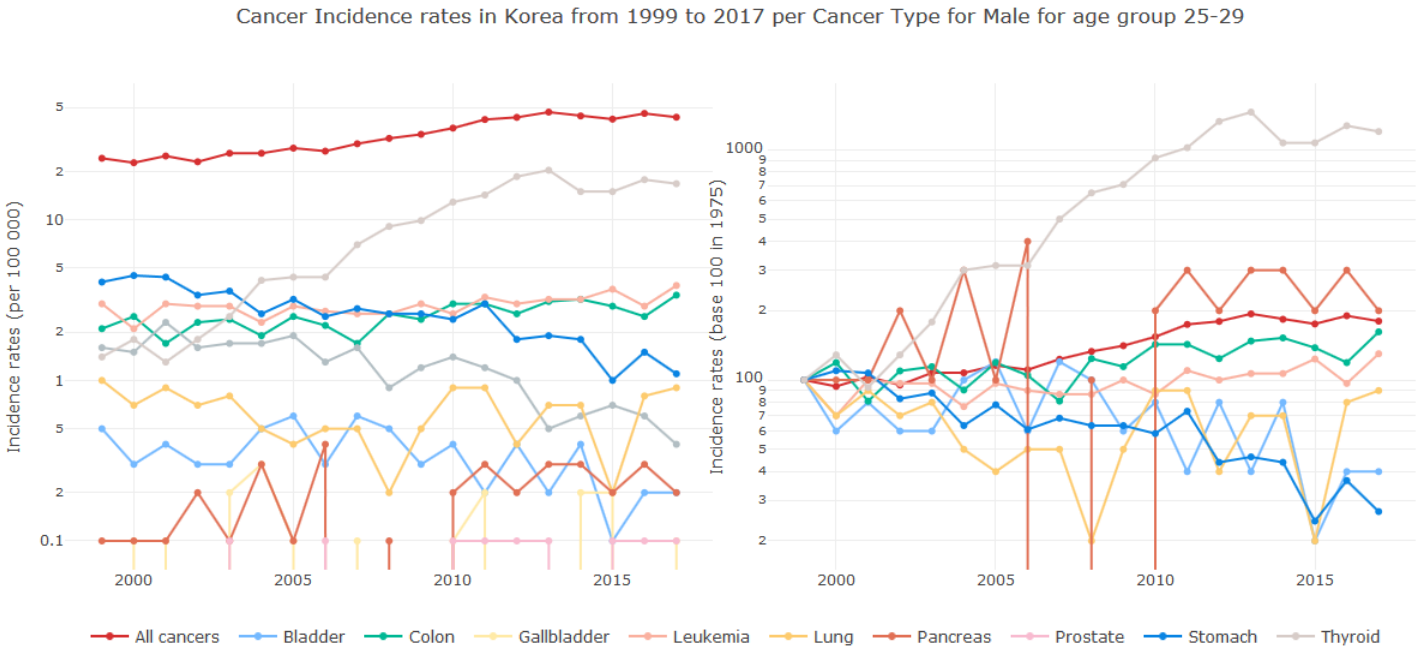


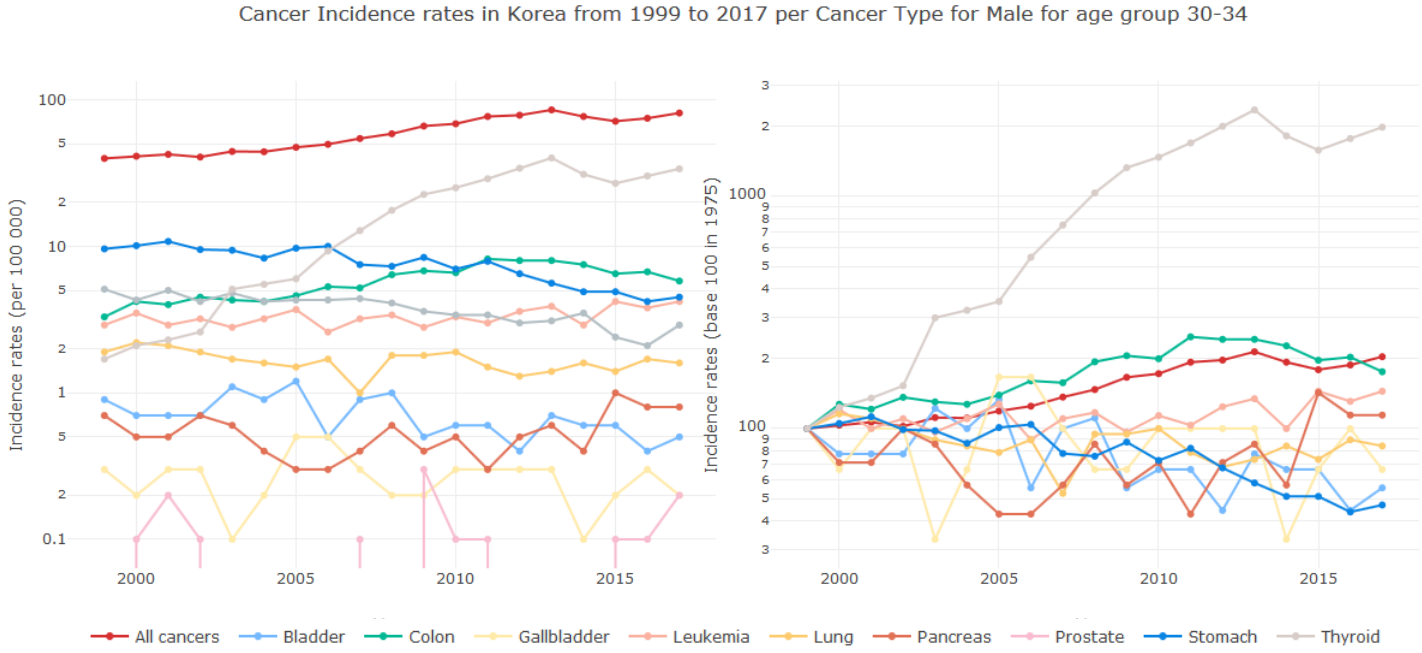
Cancer all causes annual incidence rate has been respectively around 10 for 100,000 and between 10 and 13 for 100,000 for male children aged 5 to 9 and 10 to 14. Both show an upward trend. Again, the most common cancer is leukemia and the second one thyroid cancer.



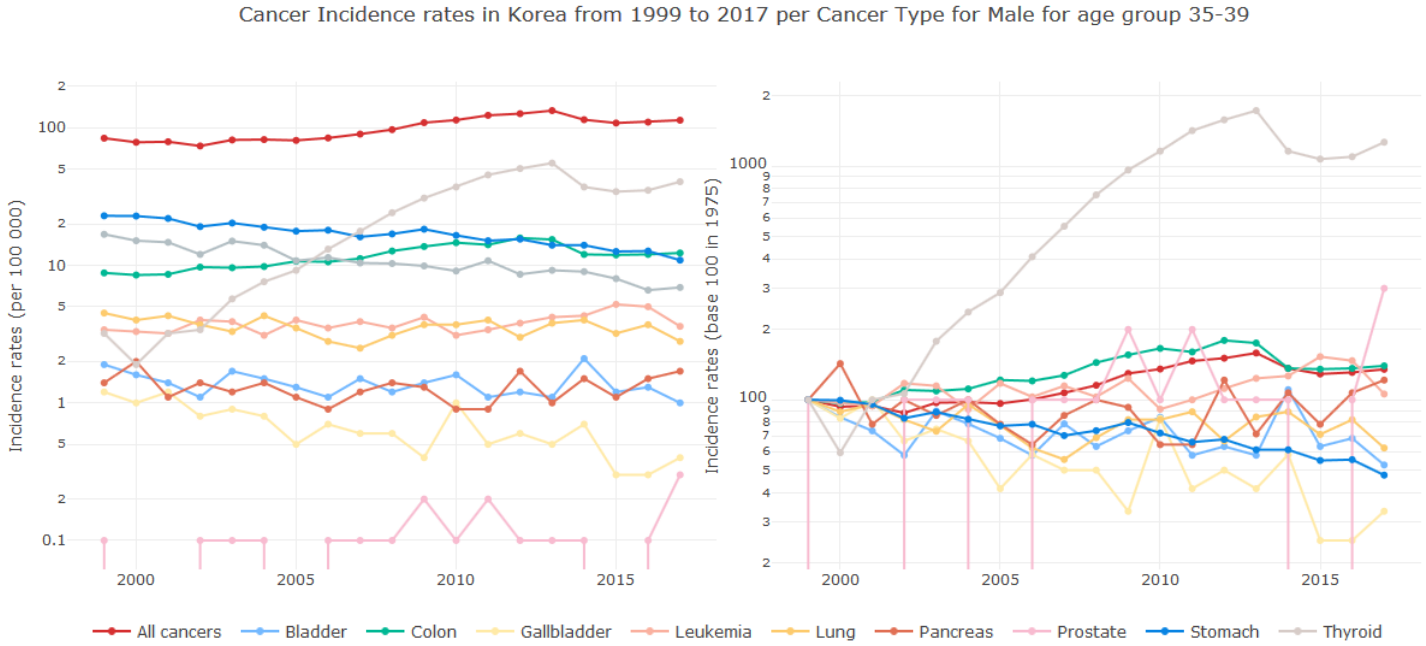


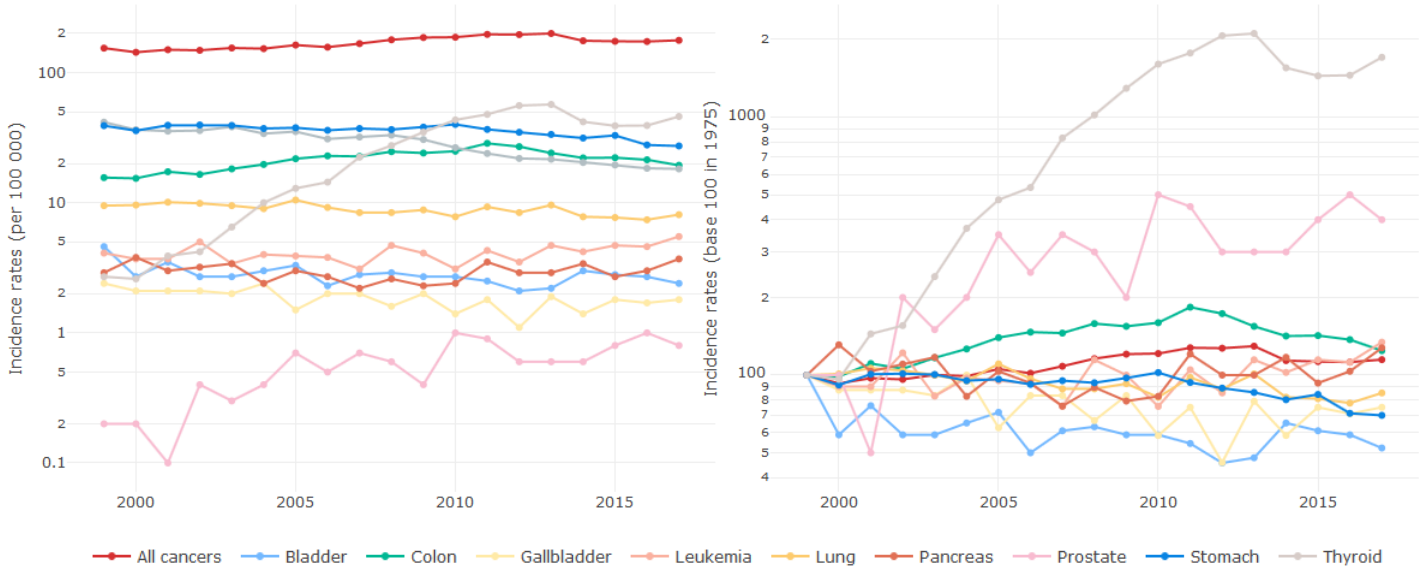
Cancer all causes annual incidence rate reaches 20 for 100,000 and 24 for 100,000 for male children aged 15 to 19 and 20 to 24 in 2017. Both also show an upward trend. The most common cancers are leukemia and thyroid cancer, with similar levels in 2017.



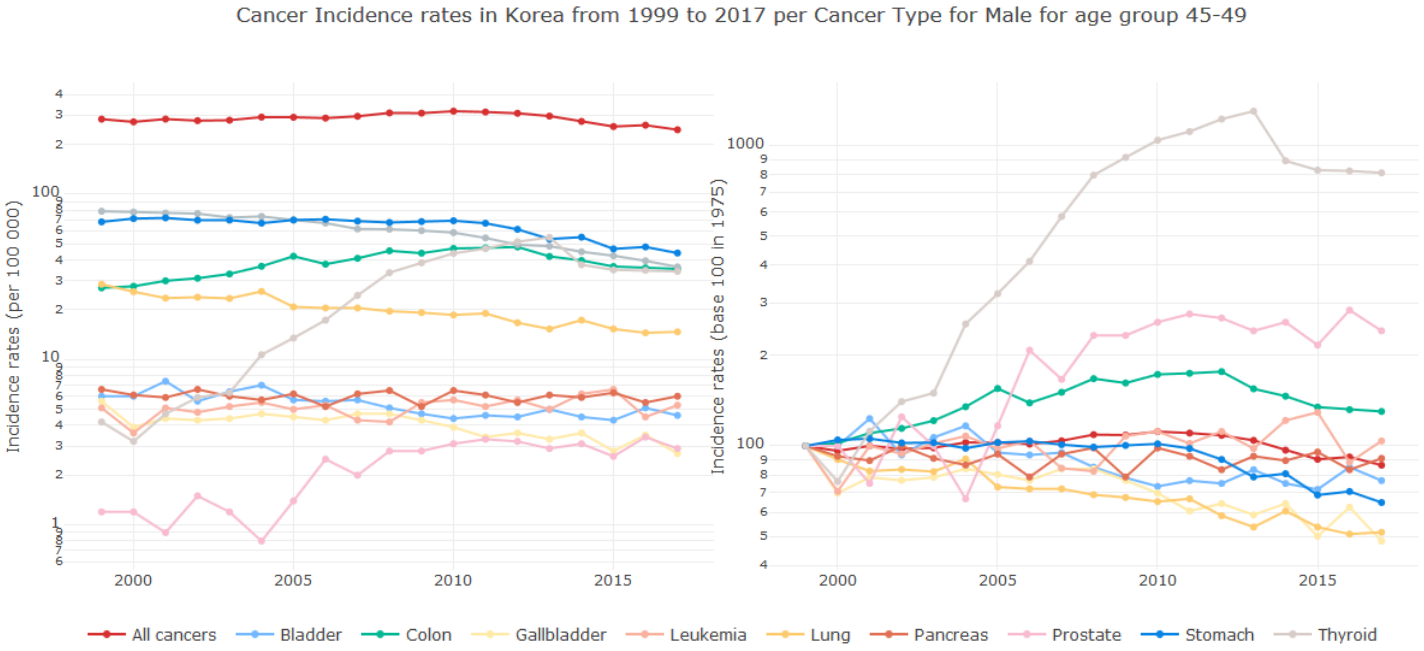


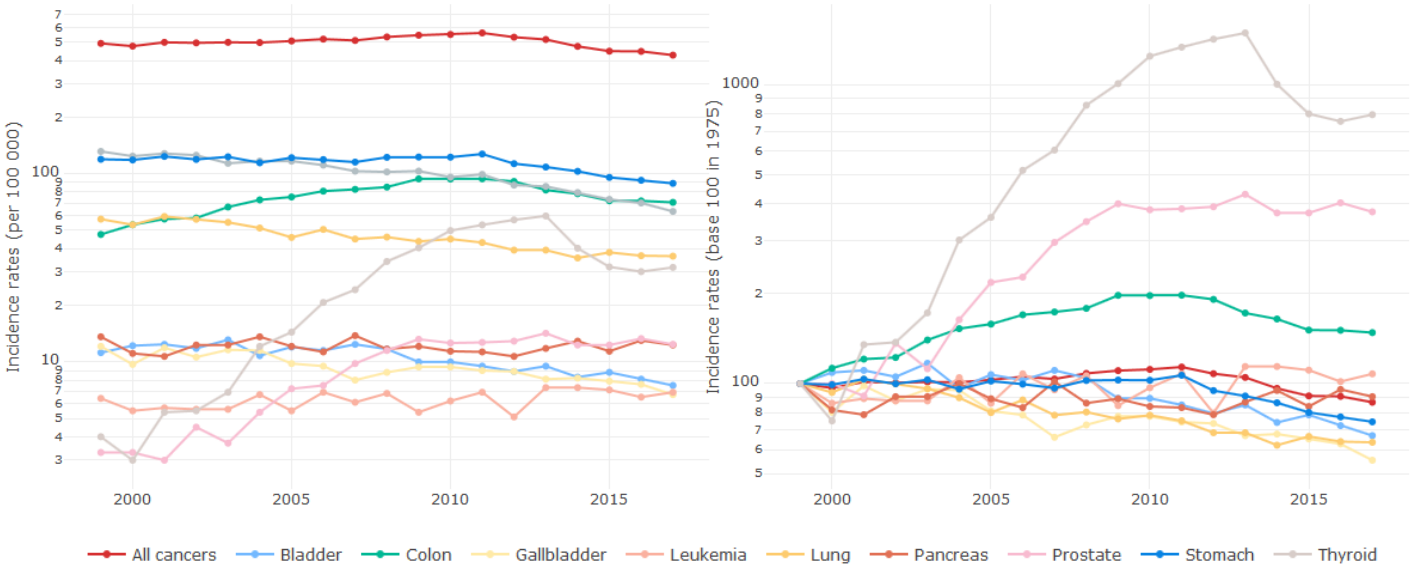
Cancer all causes annual incidence rate reaches respectively 44 and 81 for 100,000 for male aged 25 to 29 and 30 to 34. Both also show an upward trend, with a small inflexion in the most recent years. Common cancer sites are thyroid, with a strong increase through time and then leukemia (relatively stable), stomach (downward trend) and colon (upward trend).



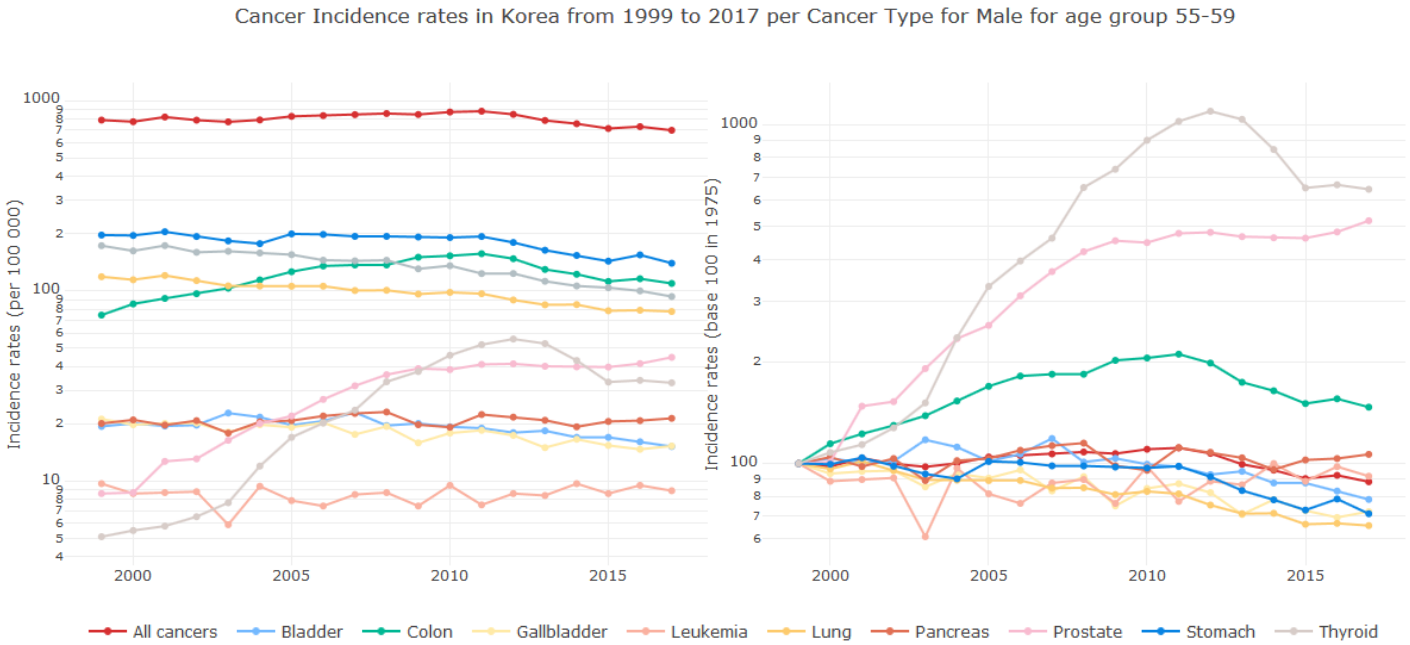


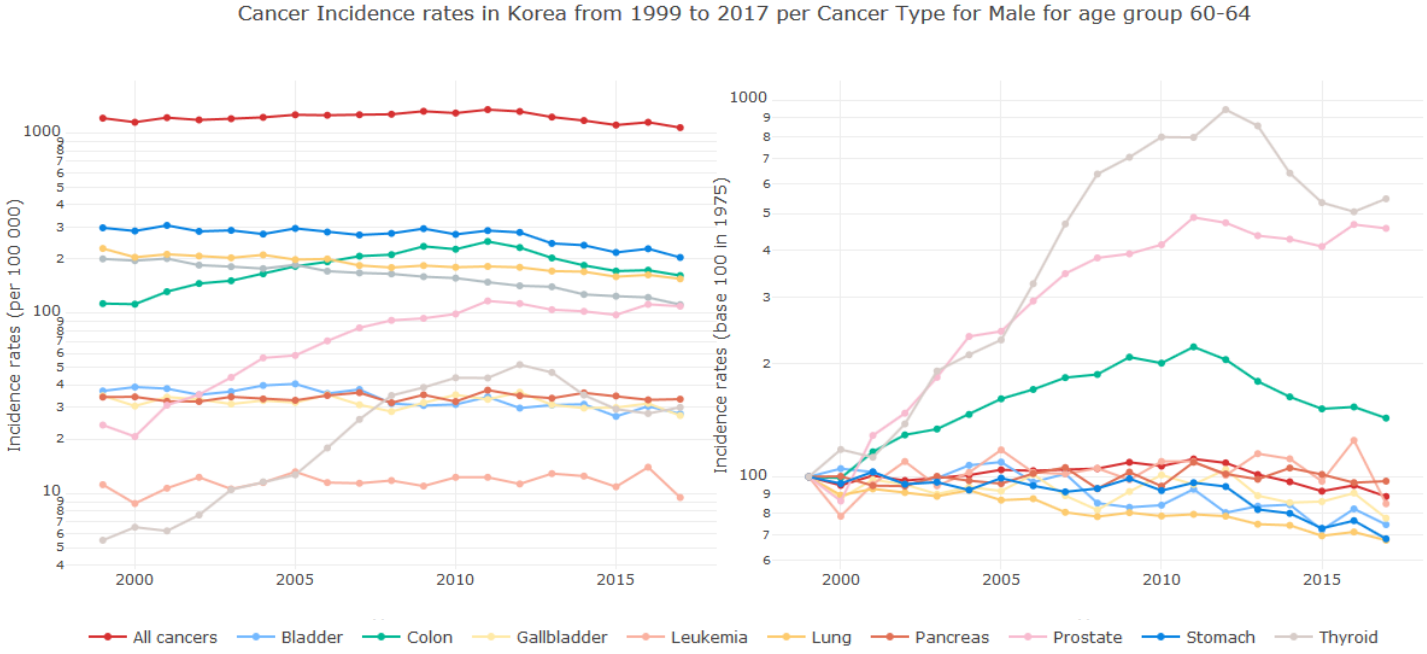
Cancer all causes annual incidence rate reaches respectively 113 and 177 for 100,000 for male aged 35 to 39 and 40 to 44. The incidence is increasing for age class 35-39 while being quite stable for age class 40-44. Common cancer sites are thyroid (strong increase), stomach and liver (downward trend) and colon (upward and then downward trend).



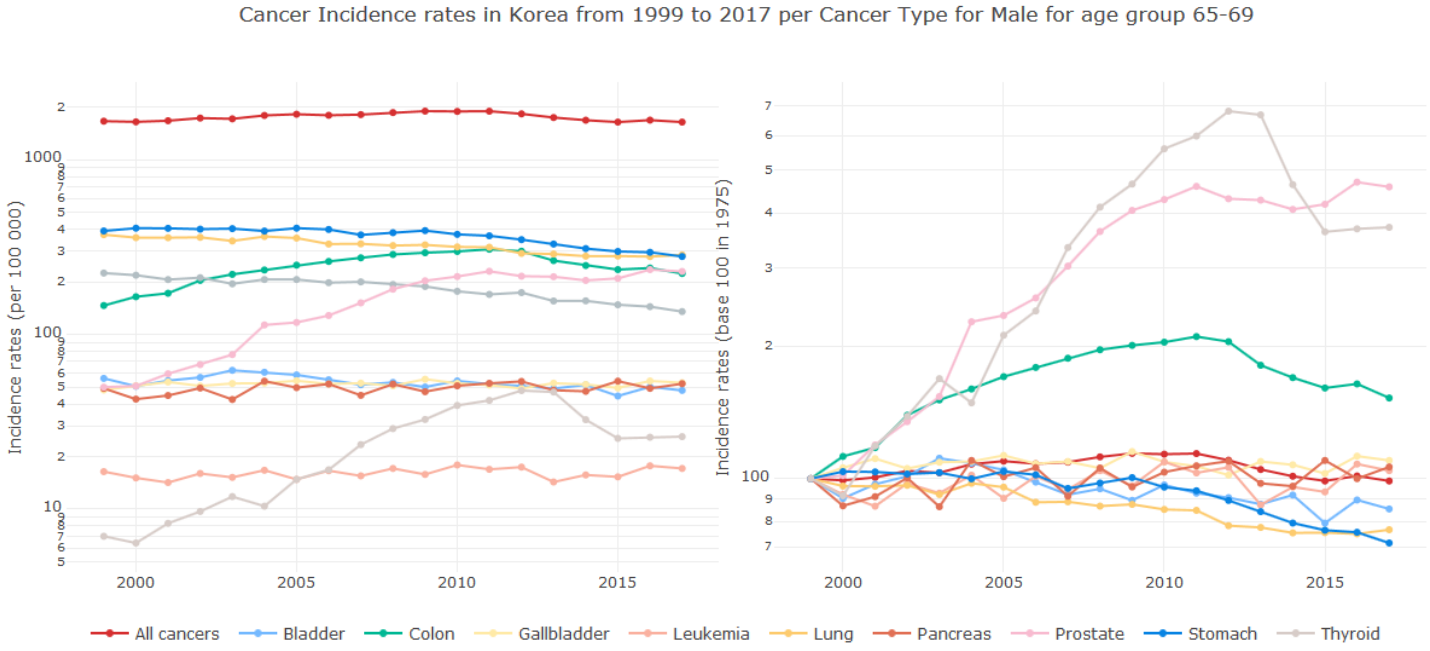


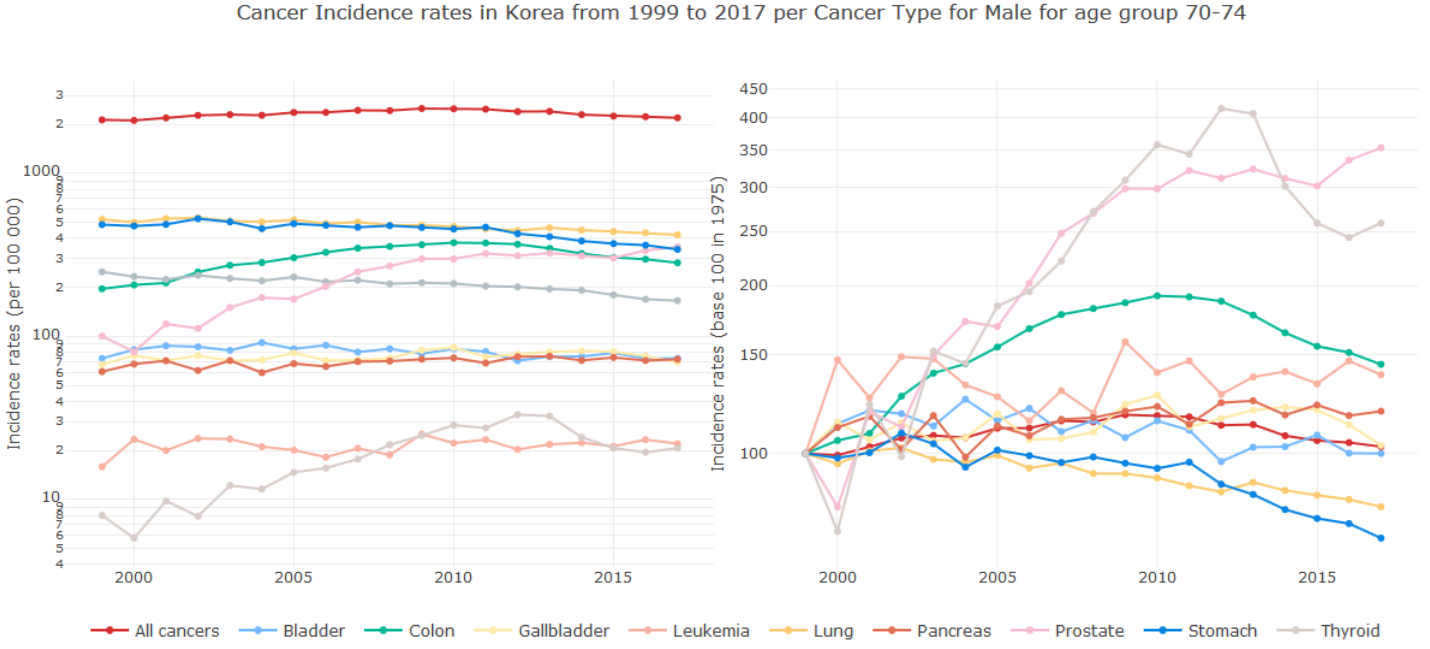
Cancer all causes annual incidence rate reaches respectively 2.5 and 4.3 for 1,000 for male aged 45 to 49 and 50 to 54. The incidence has been quite stable and then decreased. Common cancer sites are stomach and liver (downward trend), thyroid (strong increase), and colon (upward and then downward trend). As other age classes, thyroid cancer incidence slightly decreases from 2013.



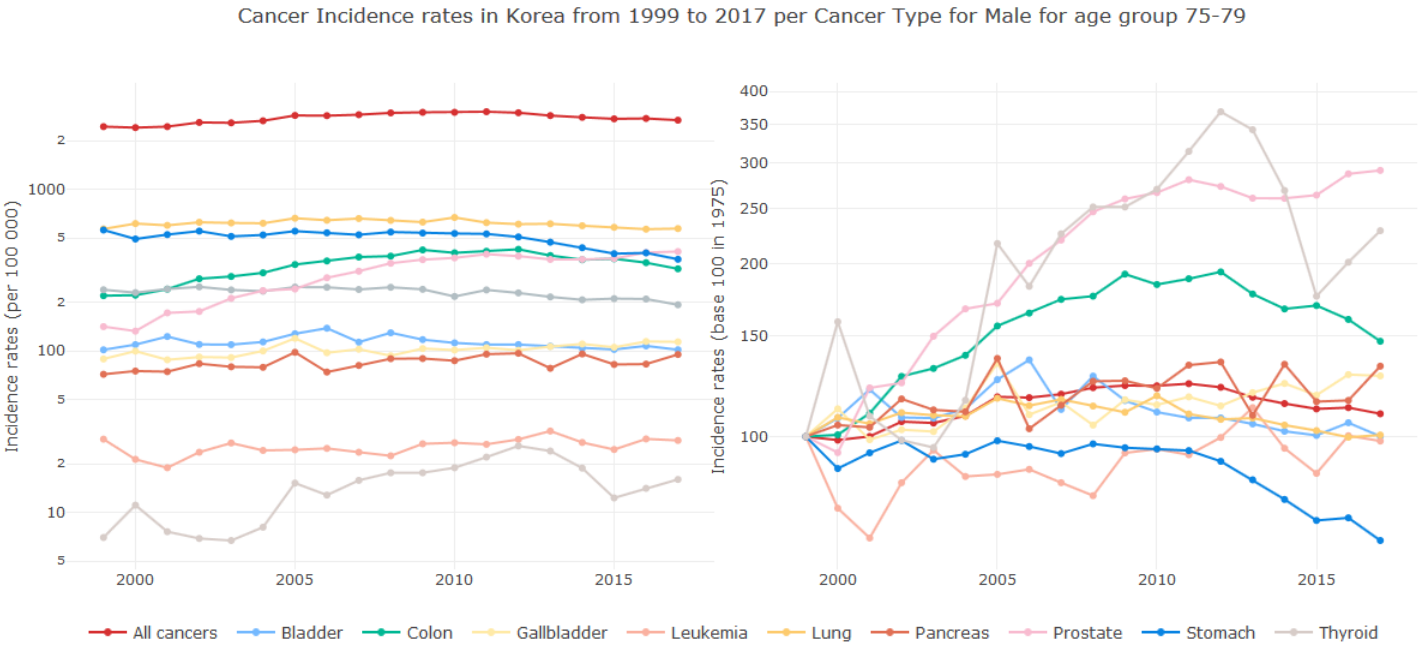


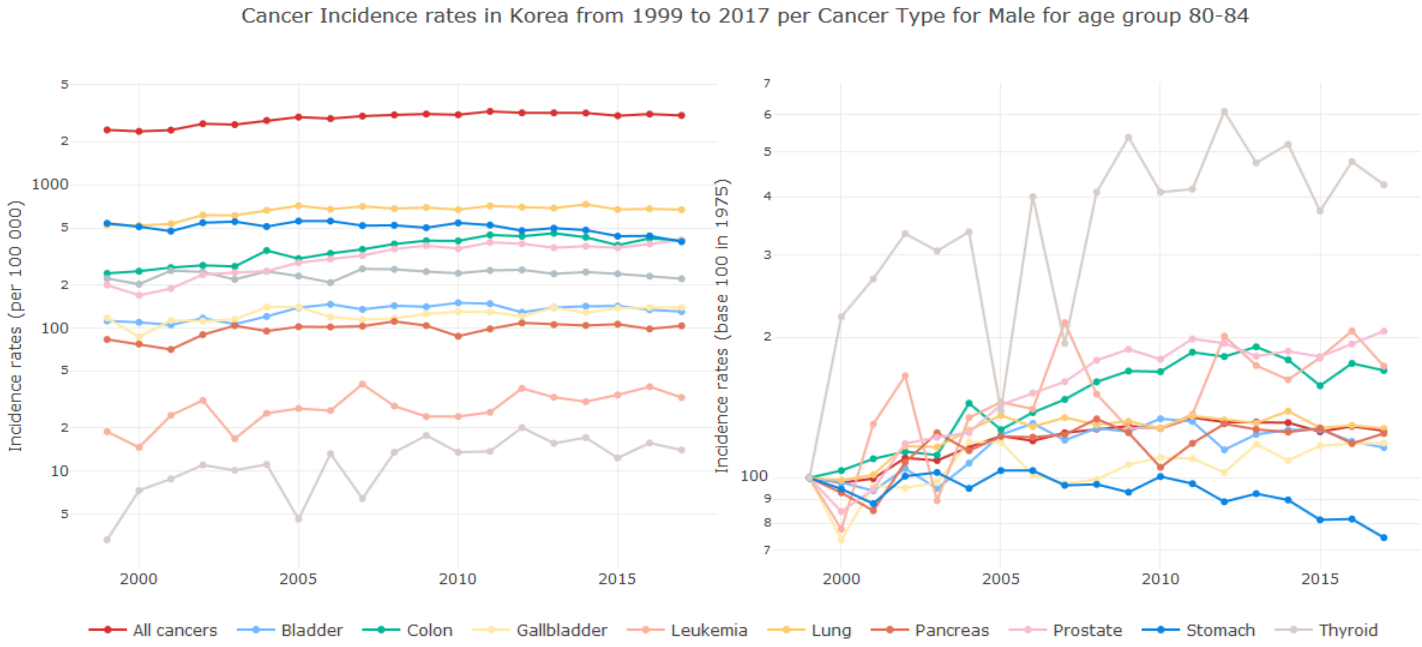
Cancer all causes annual incidence rate reaches respectively 7 for 1,000 and 1% for male aged 55 to 59 and 60 to 64. The incidence has been quite stable and then decreased. Common cancer sites are stomach, liver and lung (downward trend), thyroid and prostate (strong increase), and colon (upward and then downward trend).



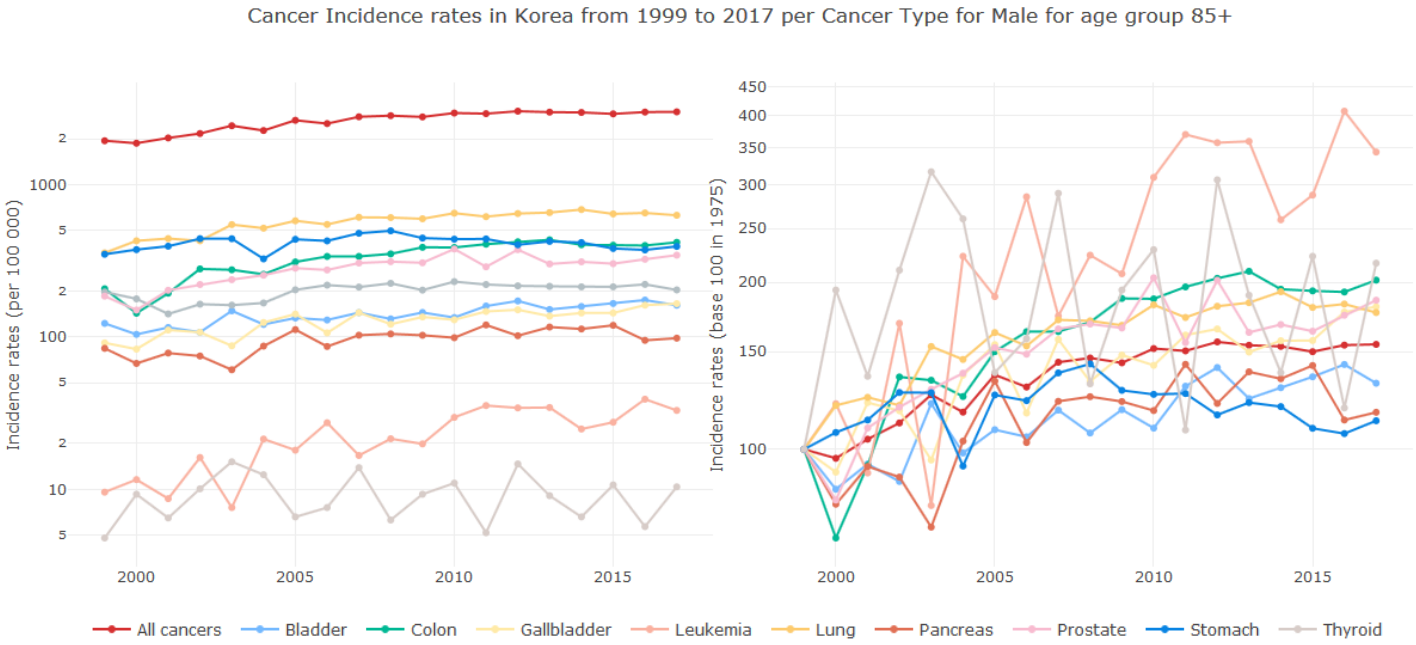


Cancer all causes annual incidence rate reaches respectively 1.6% and 2.2% for male aged 65 to 69 and 70 to 74. The incidence all causes is relatively stable and dynamics per cancer sites are similar to the previous age classes. Lung, stomach, prostate and colon have almost the incidence in 2017. Bladder, gallbladder and pancreas cancer share similar levels and trends.





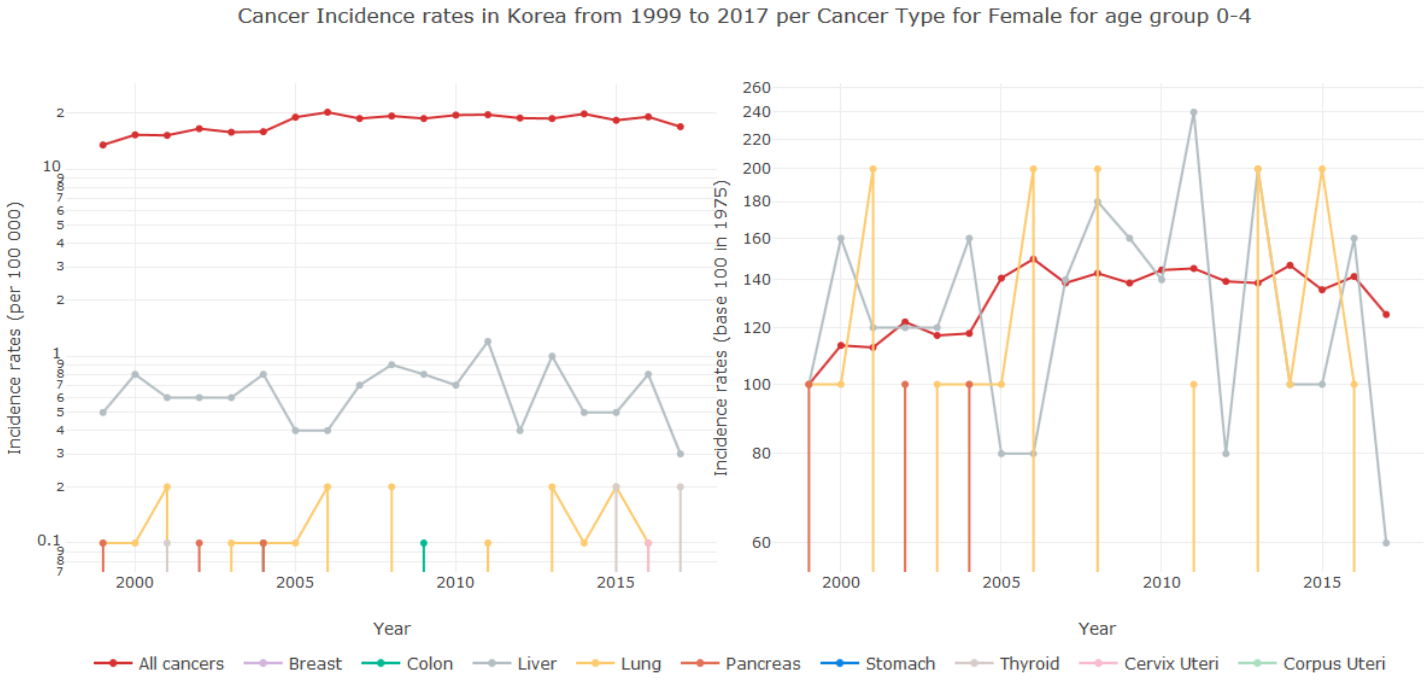
Cancer all causes annual incidence rate reaches respectively 2.2% and 2.7% for male aged 75 to 79 and 80 to 84. The incidence all causes has increased until 2010 and then decreases for age class 75-79 while being stable for age class 80-84. Prostate cancer incidence increases while other cancer sites incidence tends to decrease.

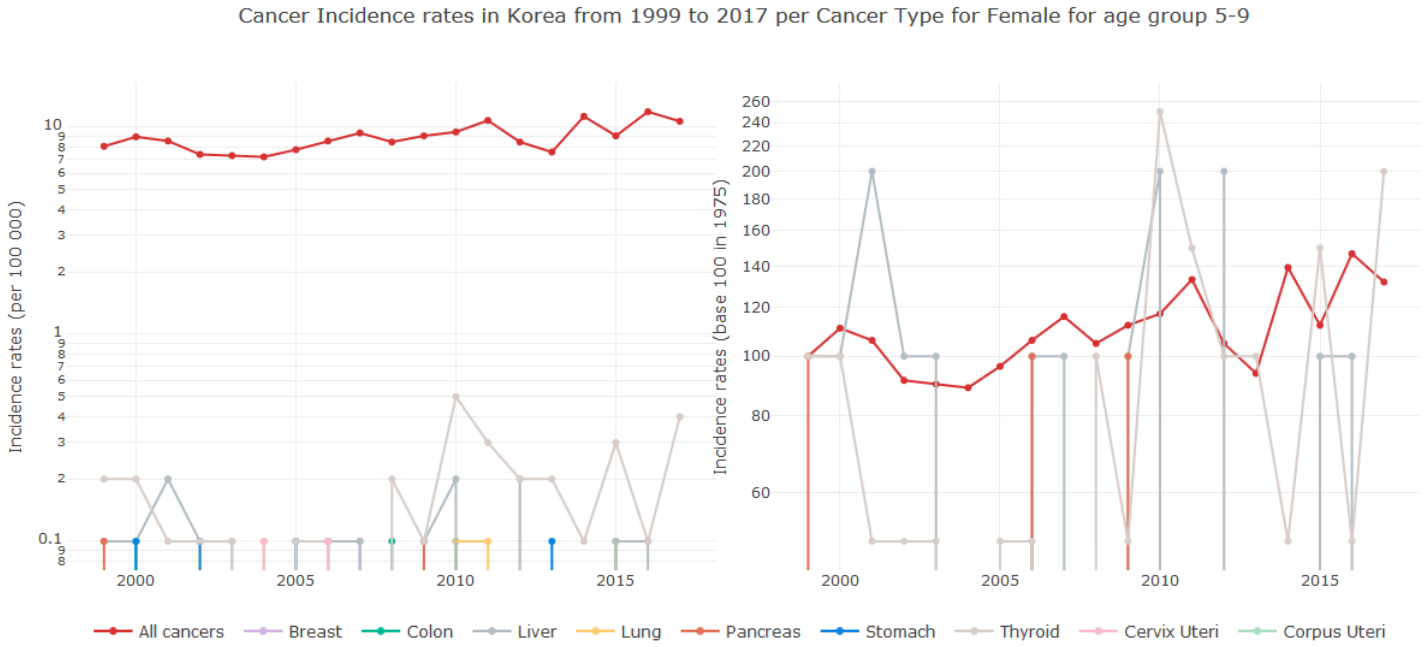


Cancer all causes annual incidence rate reaches 3% for male older than 85. The incidence all causes has increased and has been multiplied by 50% from 1999 to 2017. Cancer incidence per site have strongly increased until 2010 and then remain relatively stable. Low-incidence liver and thyroid cancer have an erratic evolution through time.

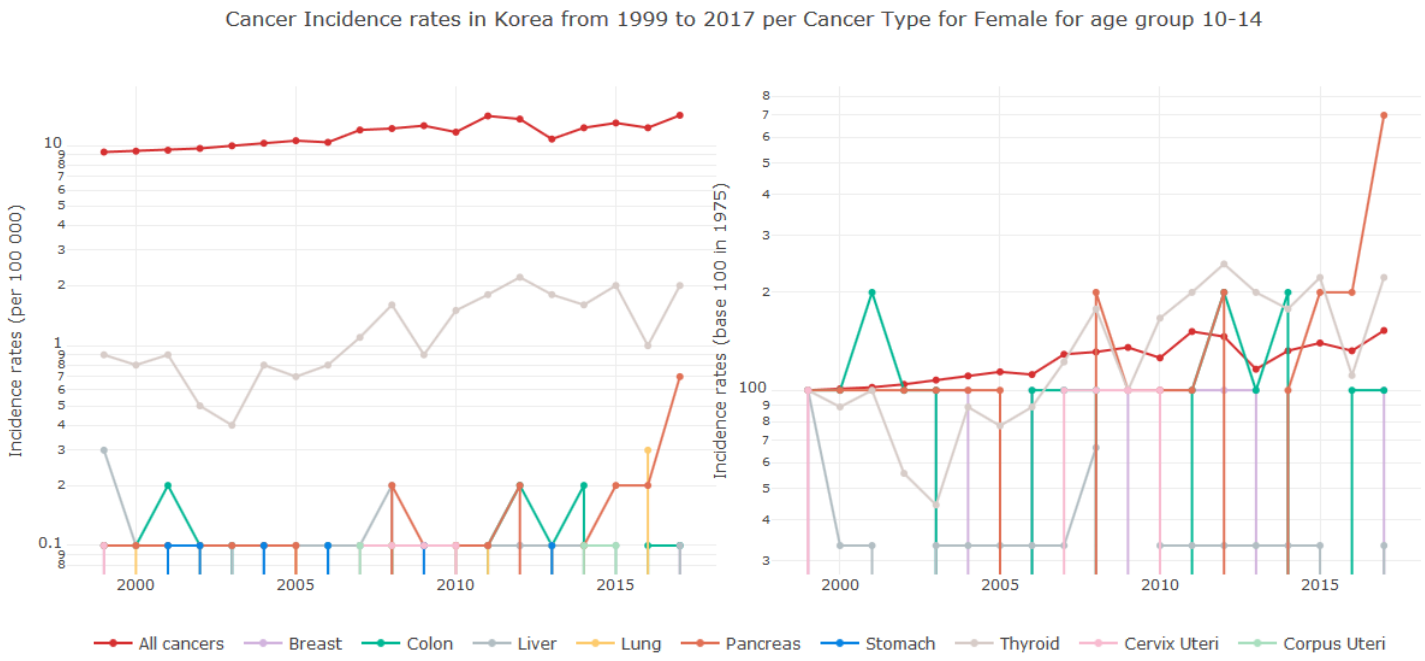
Due to life expectancy increase, the average age through time for male older than 85 may not be stable (and should has increased), which may cause cancer incidence increase. Further studies on this age class would be required to determine the role of both time and average age in cancer incidence evolution in the study period.

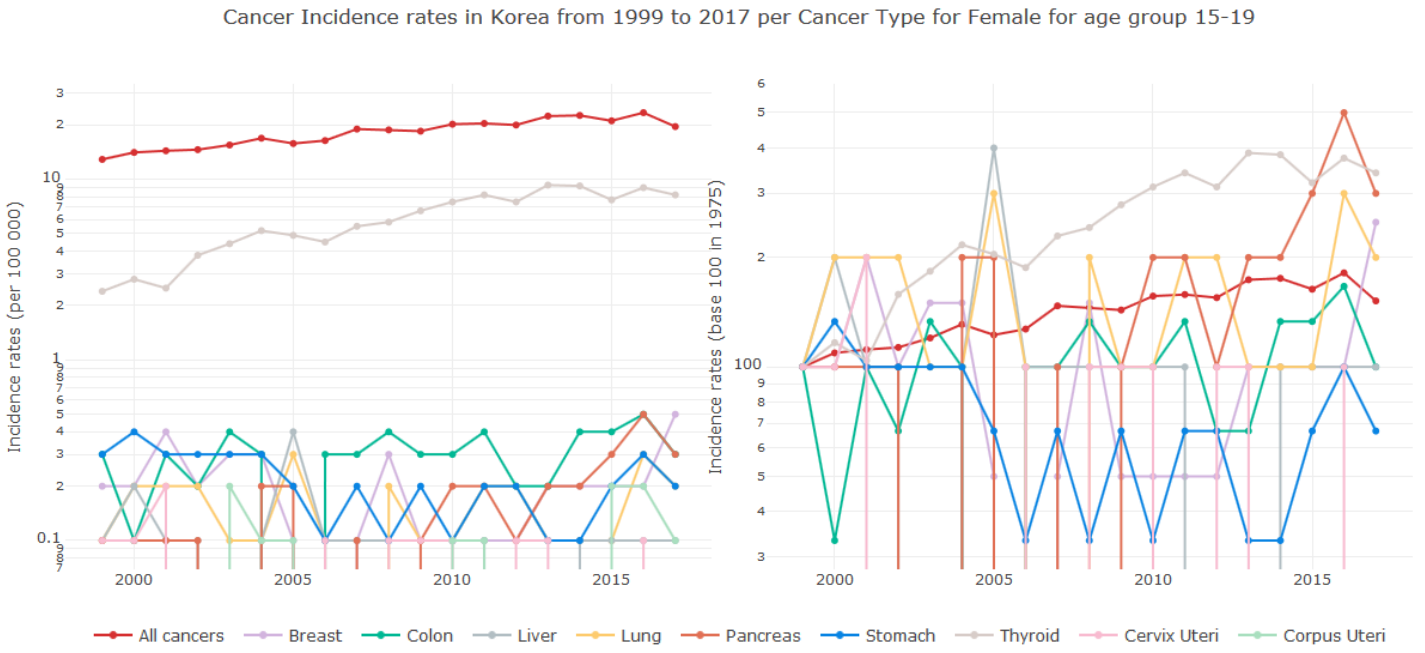
C.2 Korean Data - data per age group for female



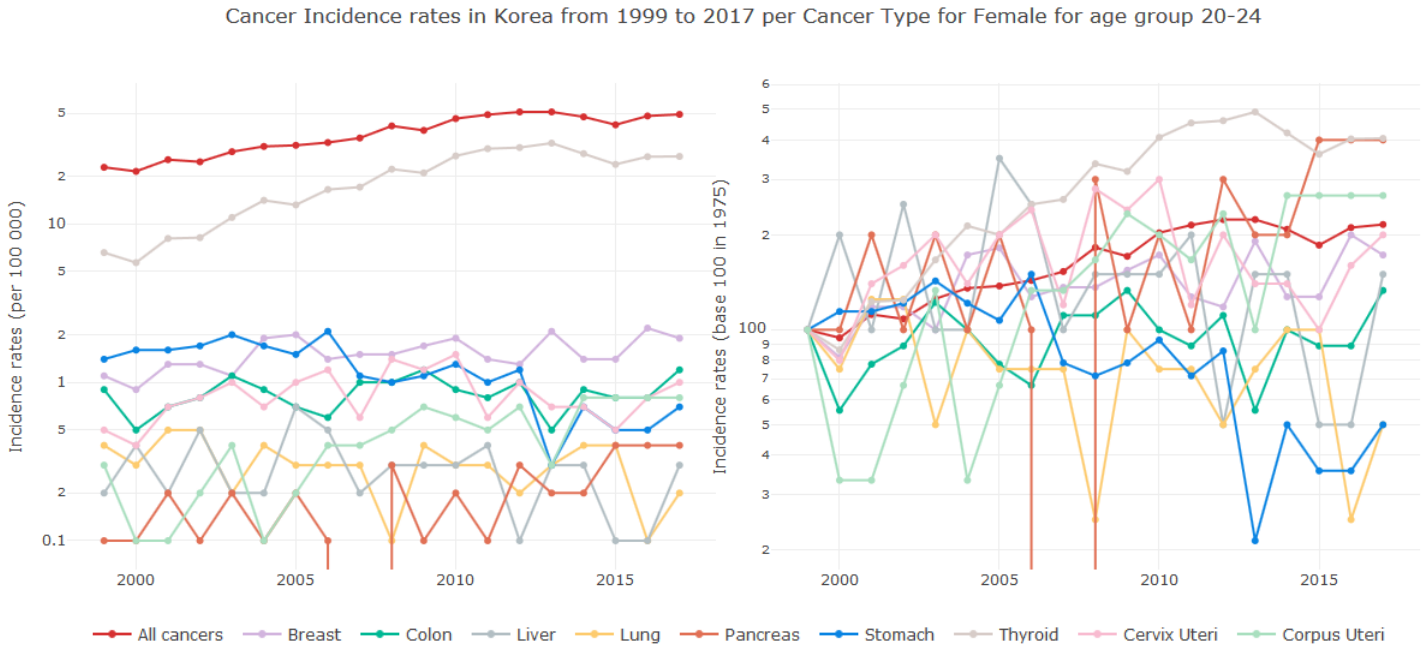


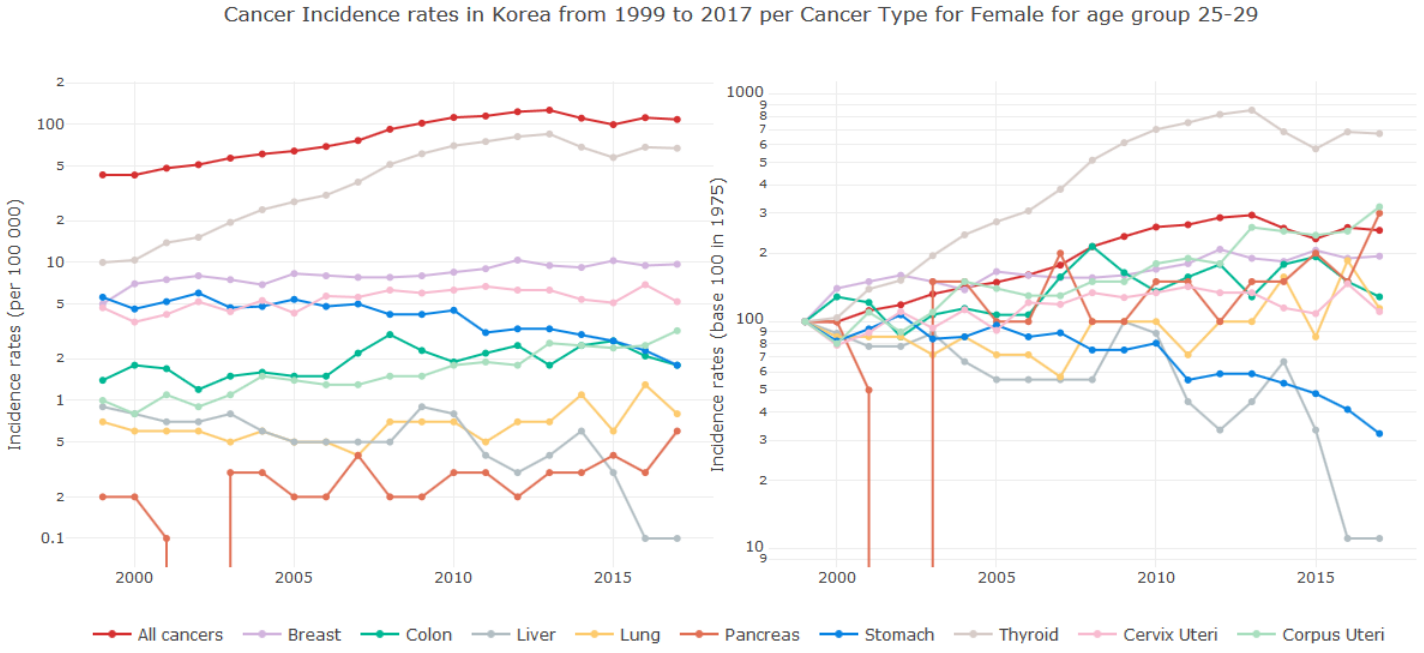
Cancer all causes annual incidence rate reaches respectively 17 and 11 for 100,000 for age classes 0 to 4 and 5 to 9. The most common cancer is leukemia (not represented in the figures).



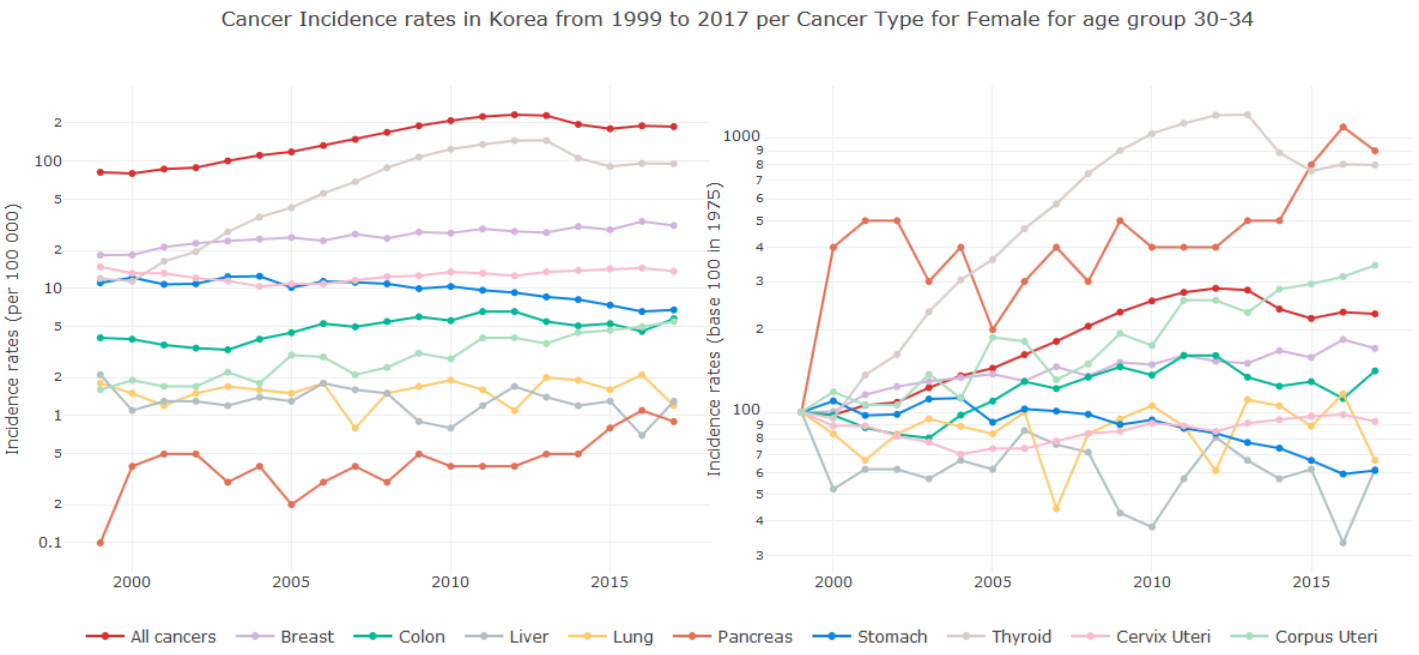


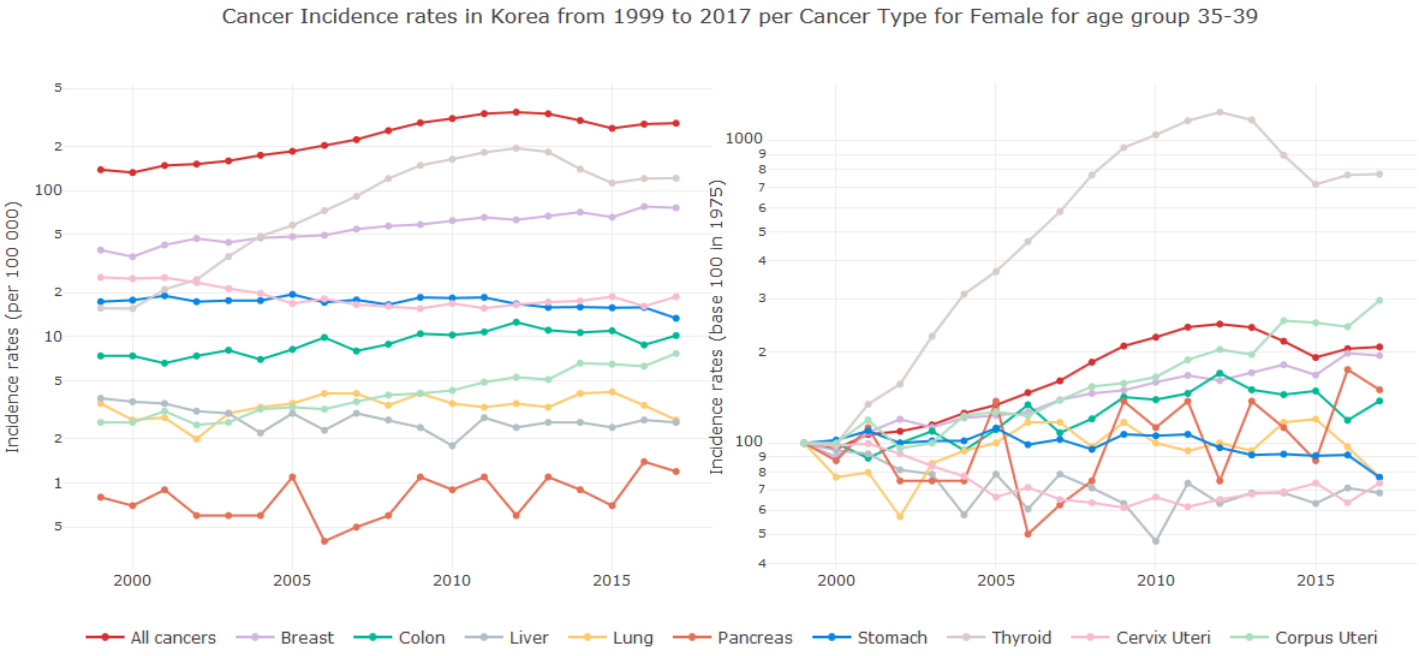
Cancer all causes annual incidence rate reaches 14 for 100,000 and 20 for 100,000 for female children aged 10 to 14 and 15 to 19. Both also show an upward trend. The most common cancers are leukemia (not represented in the figures) and thyroid cancer.



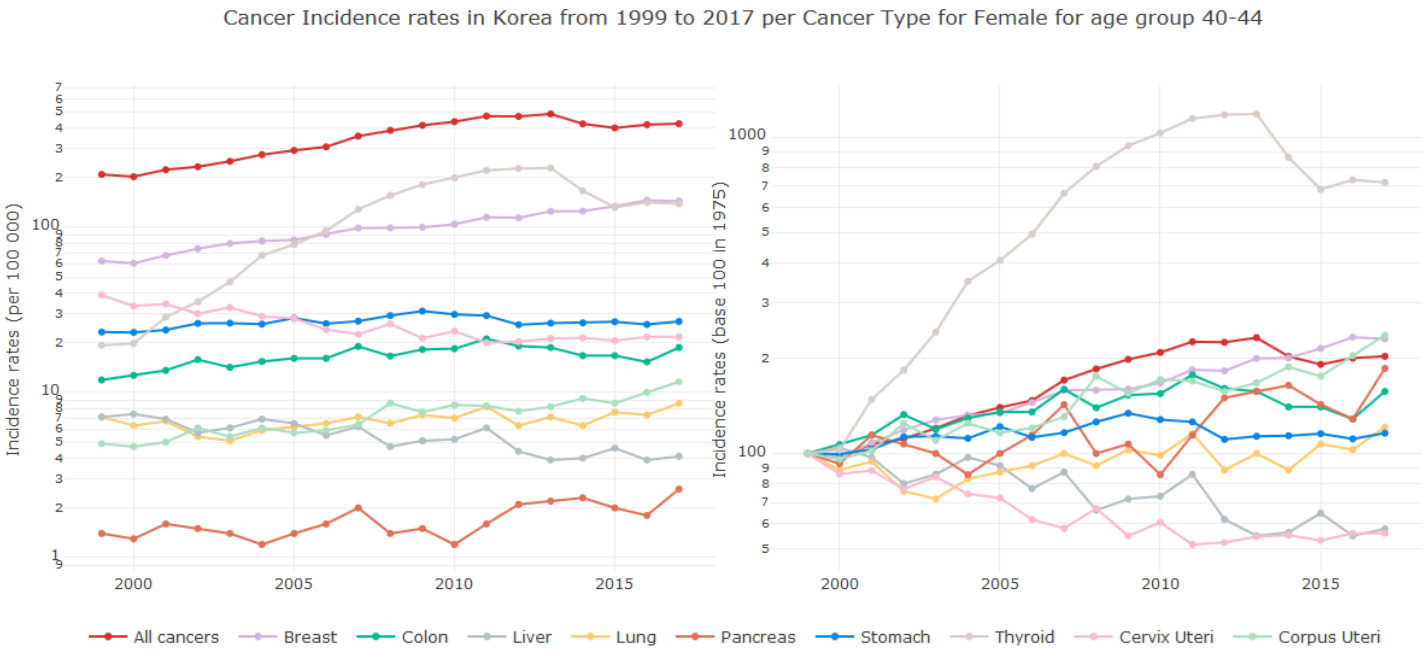


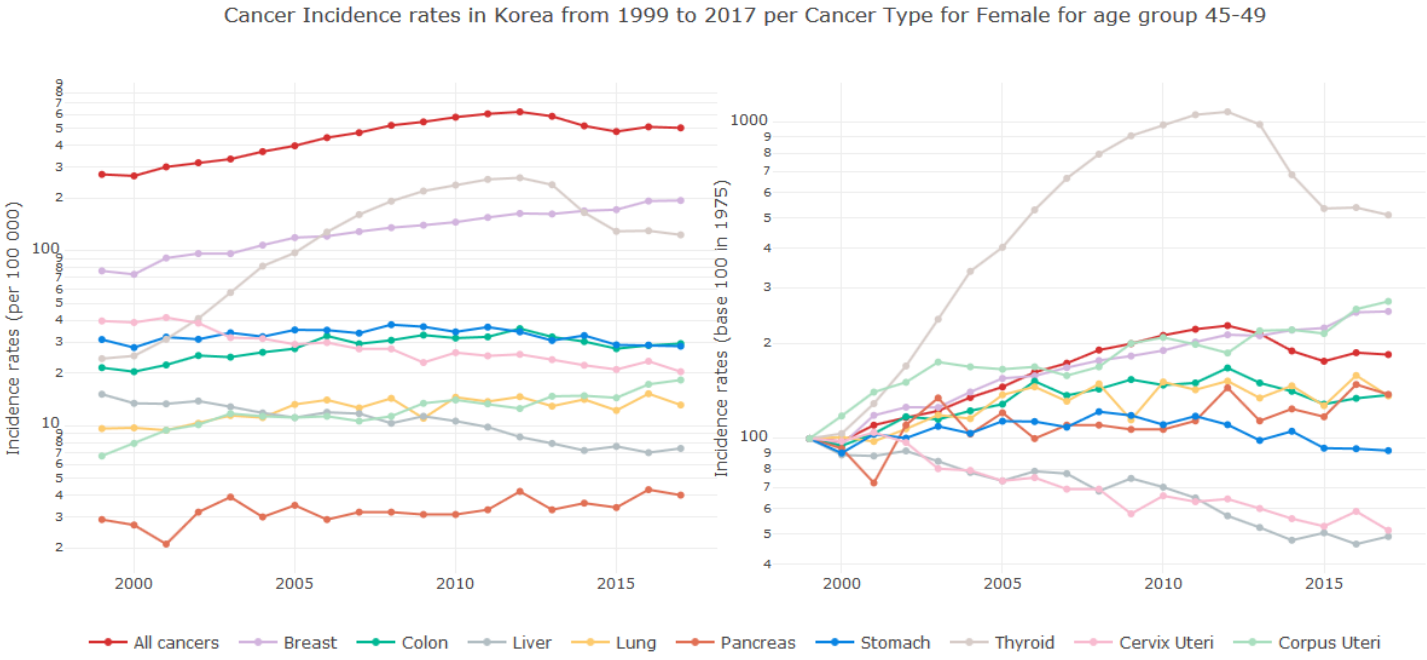
Cancer all causes annual incidence rate reaches 49 and 109 for 100,000 for female aged 20 to 24 and 25 to 29. Both also show an upward trend. Thyroid cancer represents the most frequent cancer for both age classes.



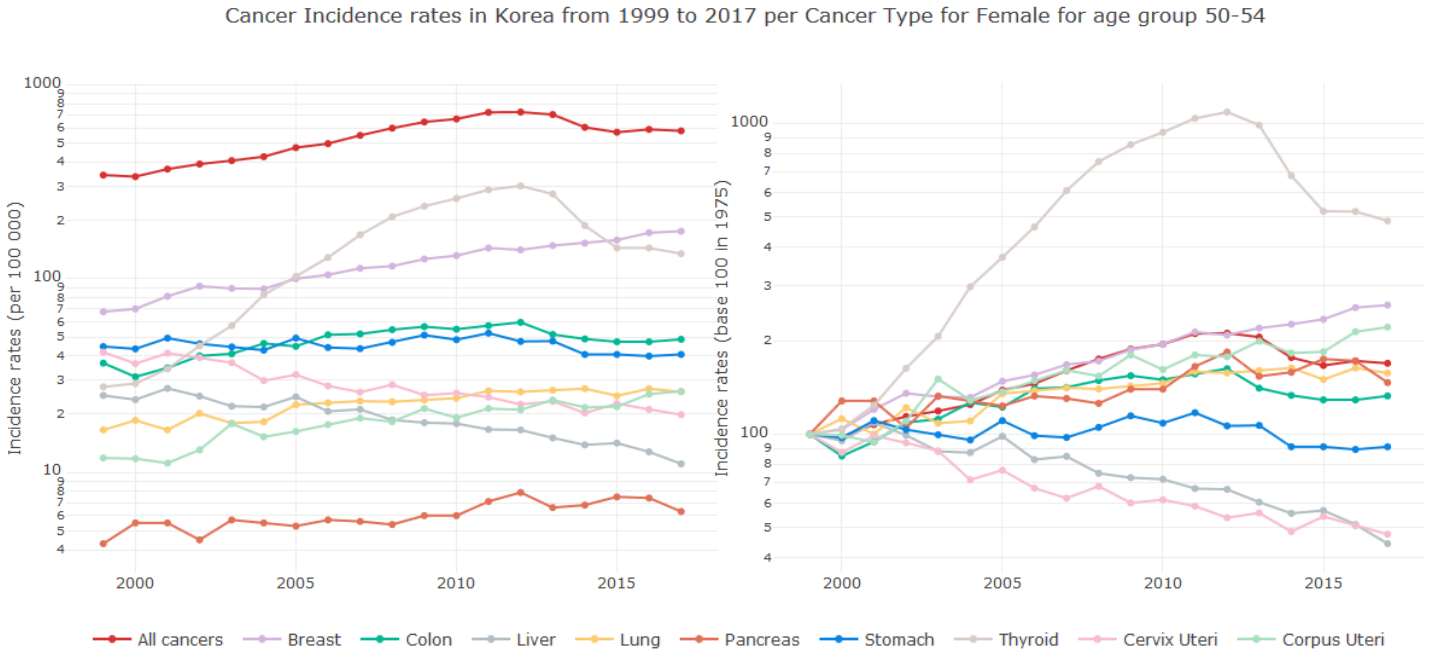


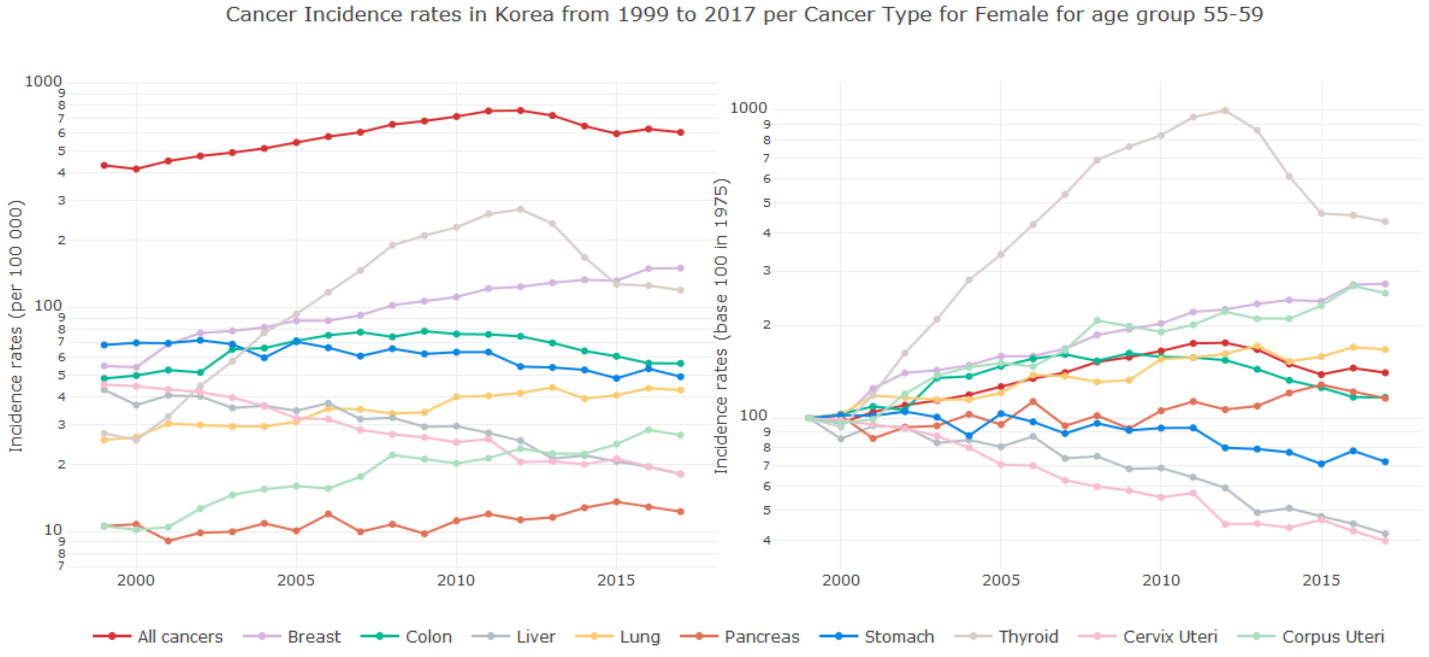
Cancer all causes annual incidence rate reaches 188 and 289 for 100,000 for female aged 30 to 34 and 35 to 39. Both also show an upward trend. Thyroid cancer and breast cancer represent most cancer for both age classes. Breast cancer regularly increase while thyroid cancer has decreased from 2013.



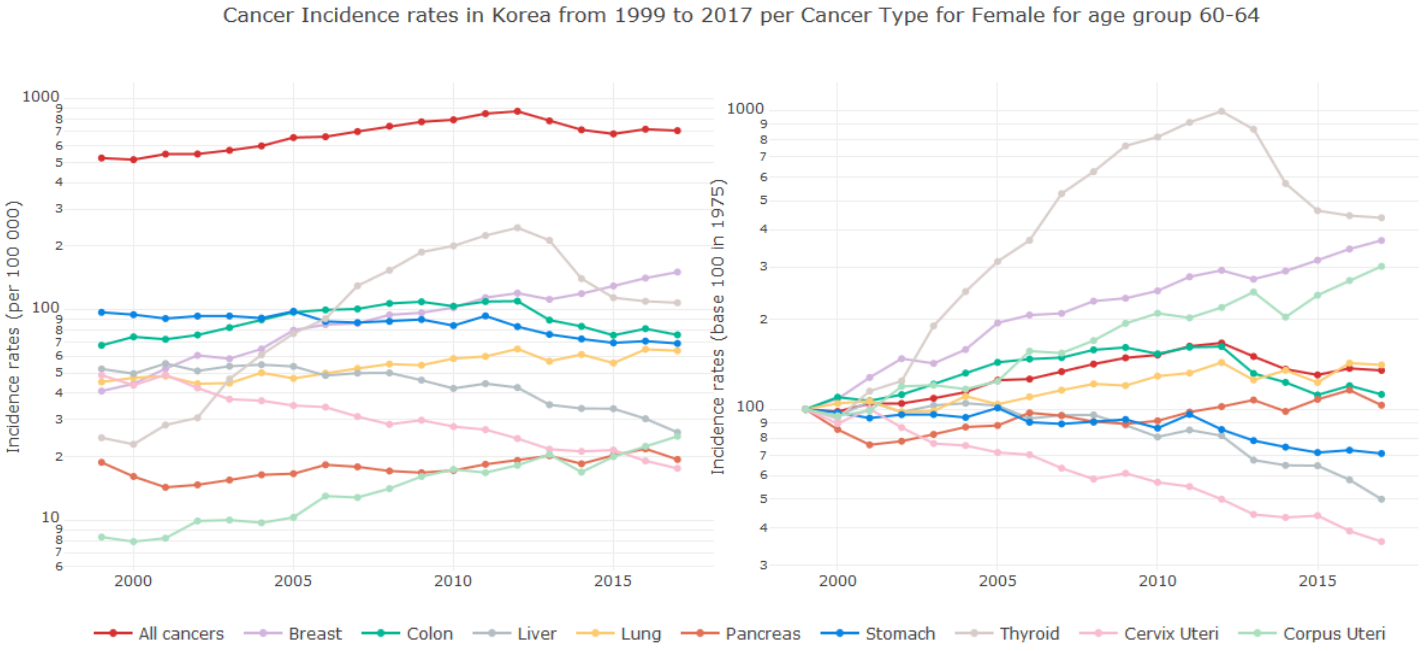


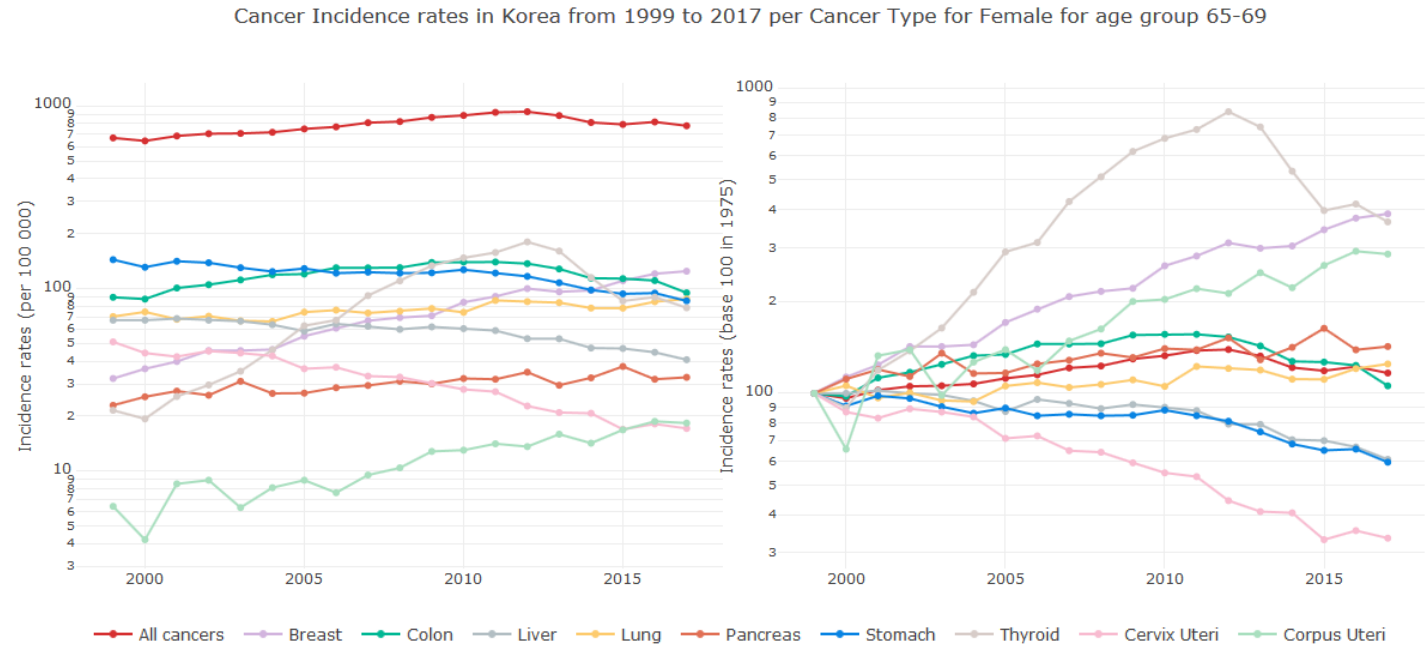
Cancer all causes annual incidence rate reaches 4 for 1,000 and 5 for 1,000 for female aged 40 to 44 and 45 to 49. Incidence trends for breast and thyroid cancer are similar to the previous age class. Stomach, liver and cervix uteri cancer incidence have decreased while corpus uteri cancer incidence has increased.



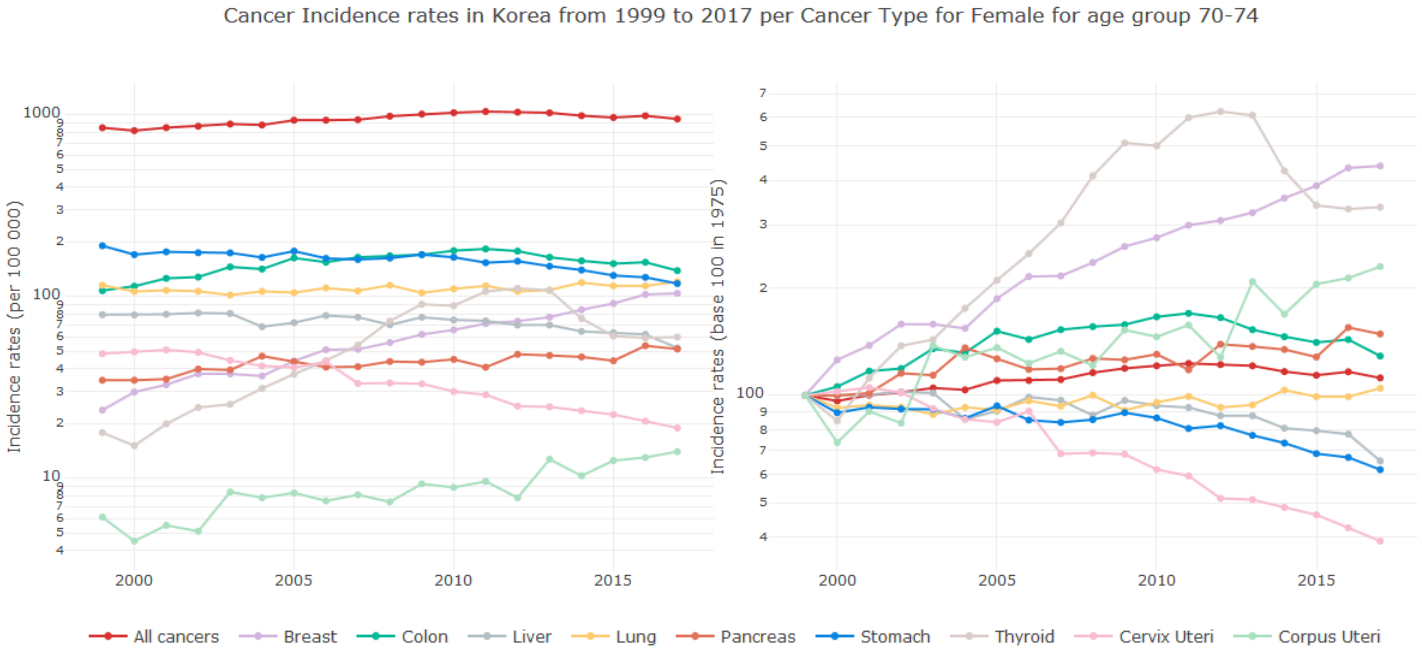


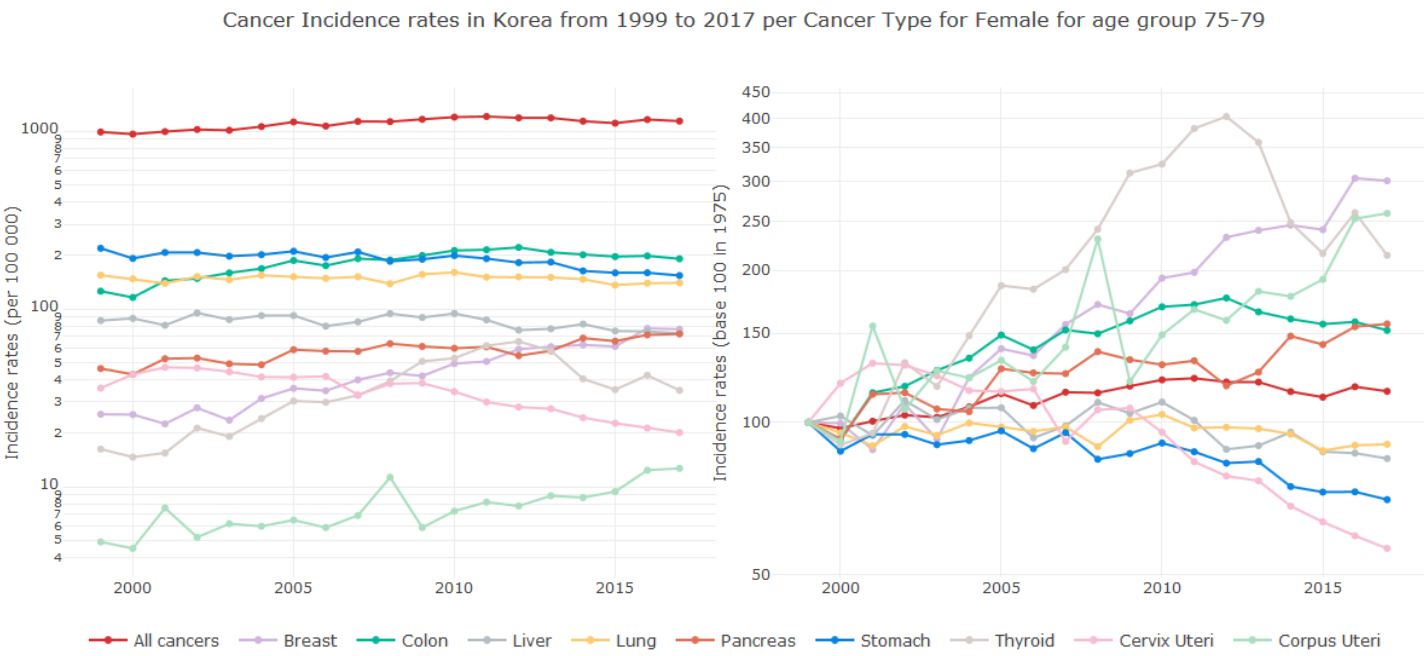
Cancer all causes annual incidence rate reaches 6 for 1,000 for both female aged 50 to 54 and 55 to 59. Again, the incidence increase is due to both breast and thyroid cancer incidence increase. These age classes demonstrate the thyroid-cancer epidemic in Korea[38]. The increase of breast cancer is steady and overall very strong from 1999.



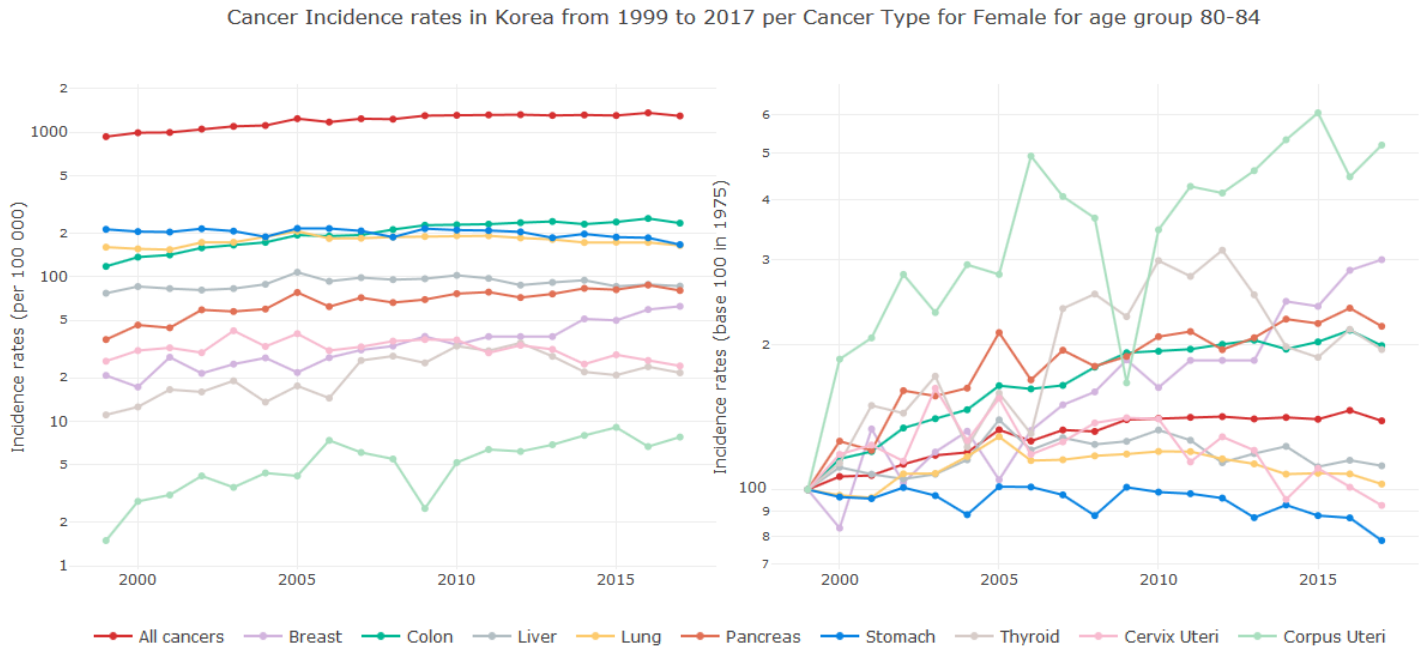


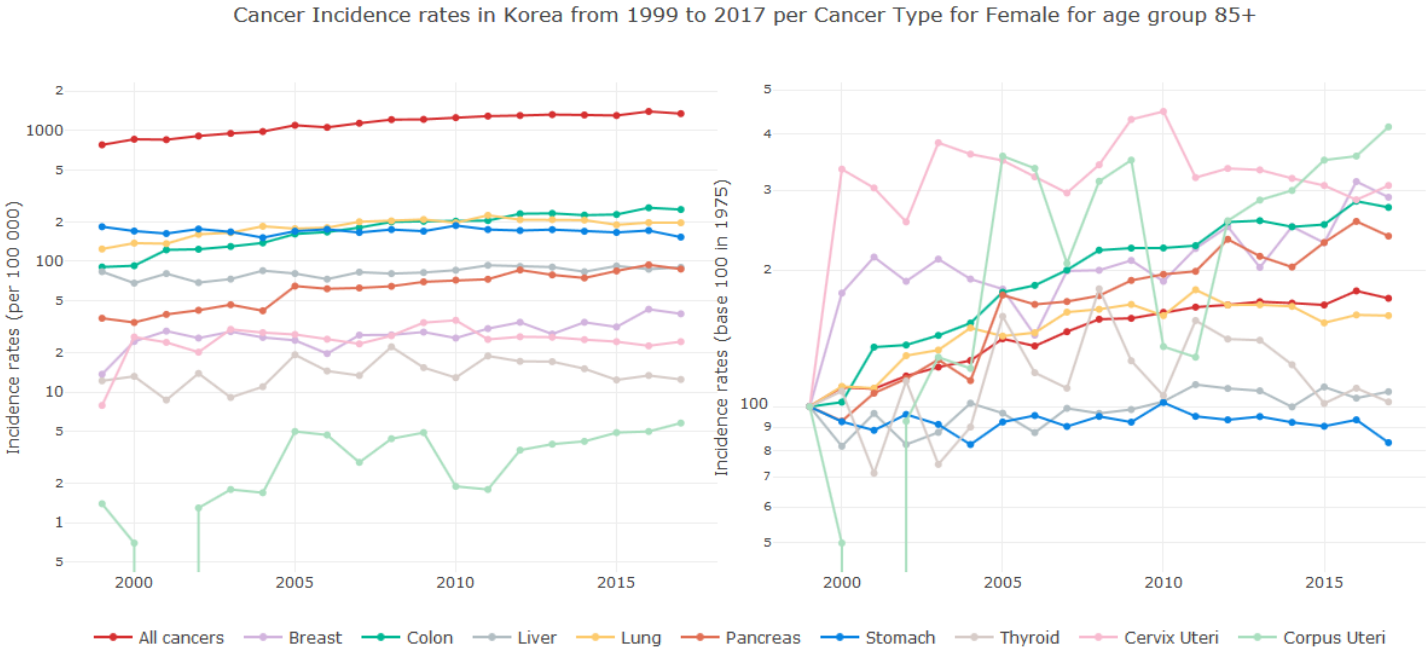
Cancer all causes annual incidence rate reaches respectively 7 and 8 for 1,000 for female aged 60 to 64 and 65 to 69. Both breast and thyroid cancer incidence have increased while stomach and colon cancer incidence have decreased. Then, lung cancer incidence has increased in the study period.





Cancer all causes annual incidence rate reaches respectively 0.9 and 1.1% for female aged 70 to 74 and 75 to 79. Most frequent cancer sites are colon, stomach and lung, while breast and thyroid are sites with the largest incidence increase. While less frequent, cervix uteri, corpus uteri and pancreas cancer have globally increased through time.





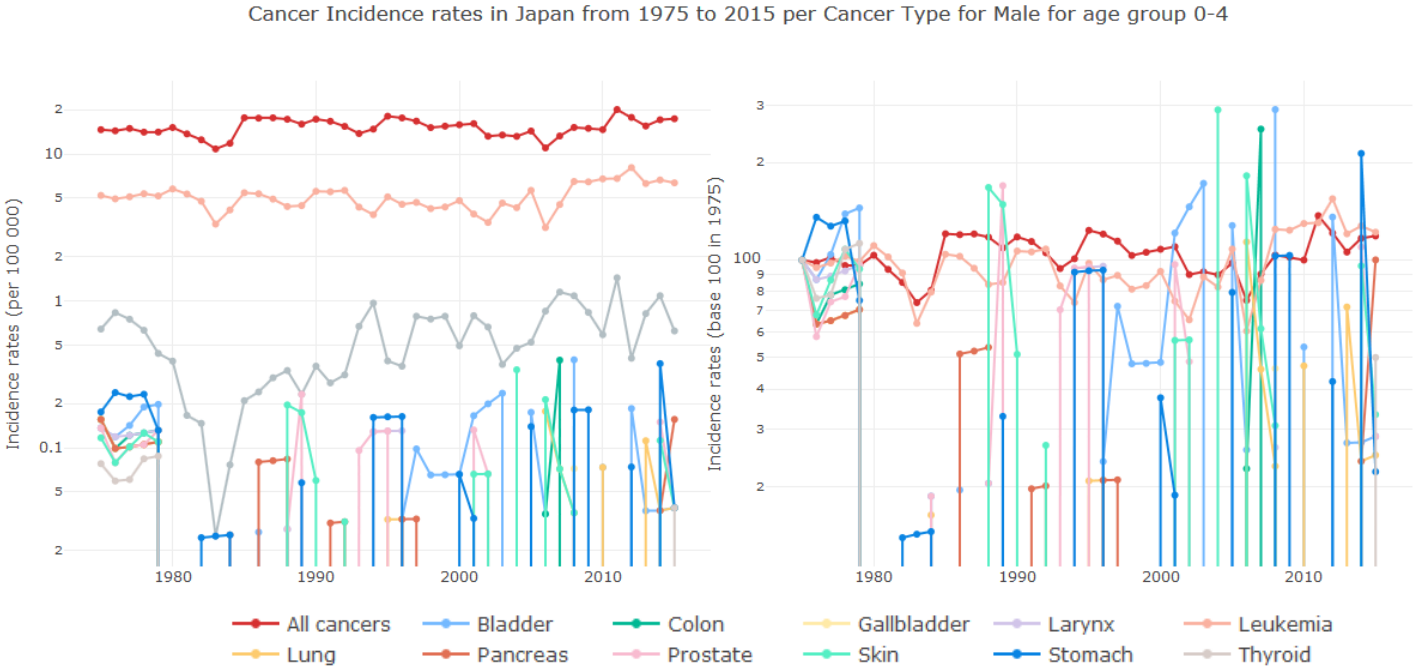
Cancer all causes annual incidence rate reaches respectively 1.3 and 1.35% for female aged 80 to 84 and 85 and more. Most frequent cancer sites are colon, stomach and lung, and then liver and pancreas. As mentioned for male, the age class 85 and more must be analyzed with care due to possible changes in average age.

D Appendix D: Japan Cancer Incidence Data

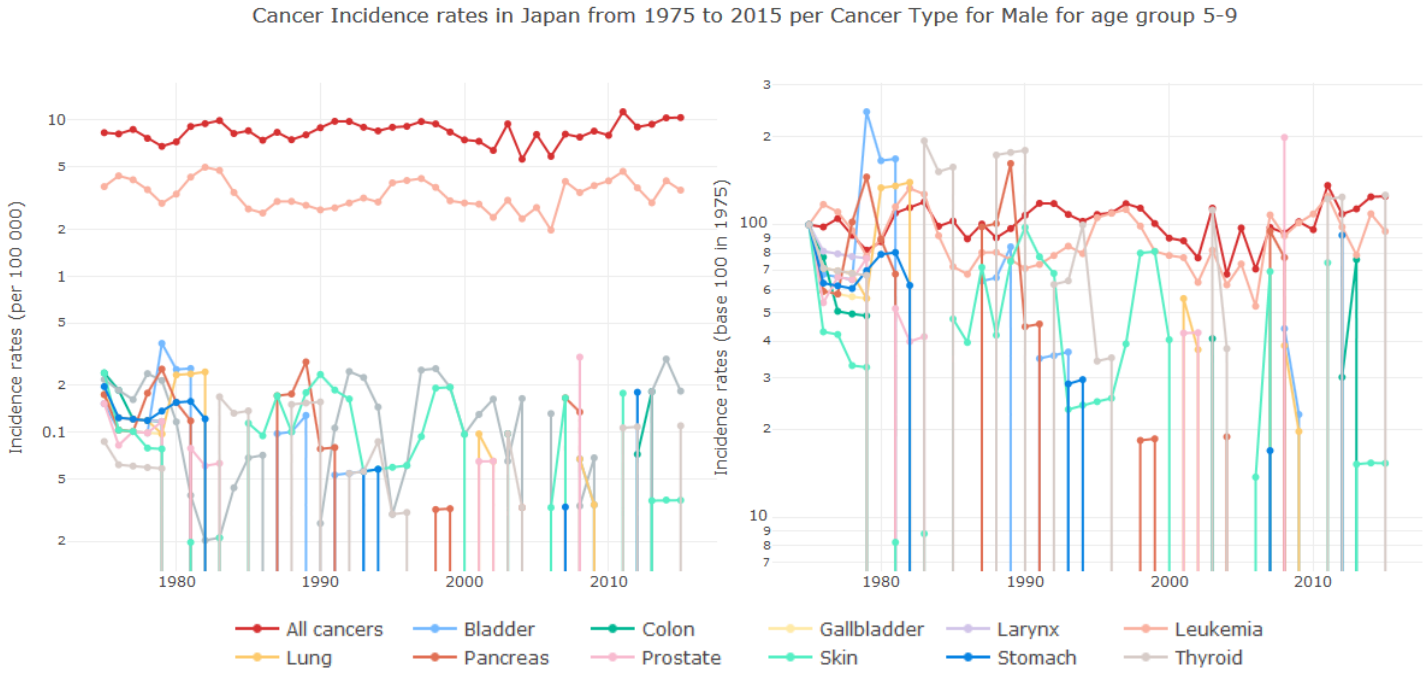
This section illustrates the database with cancer incidence rates per gender and age group. For each age group, two figures have been generated: on left, the crude incidence rates and on right, the index rates assuming an index 100 in 1975. For both figures, total incidence rates as well as 11 cancer site are presented. Due to the huge differences in incidence rates, the y-axis is using a logarithmic scale.

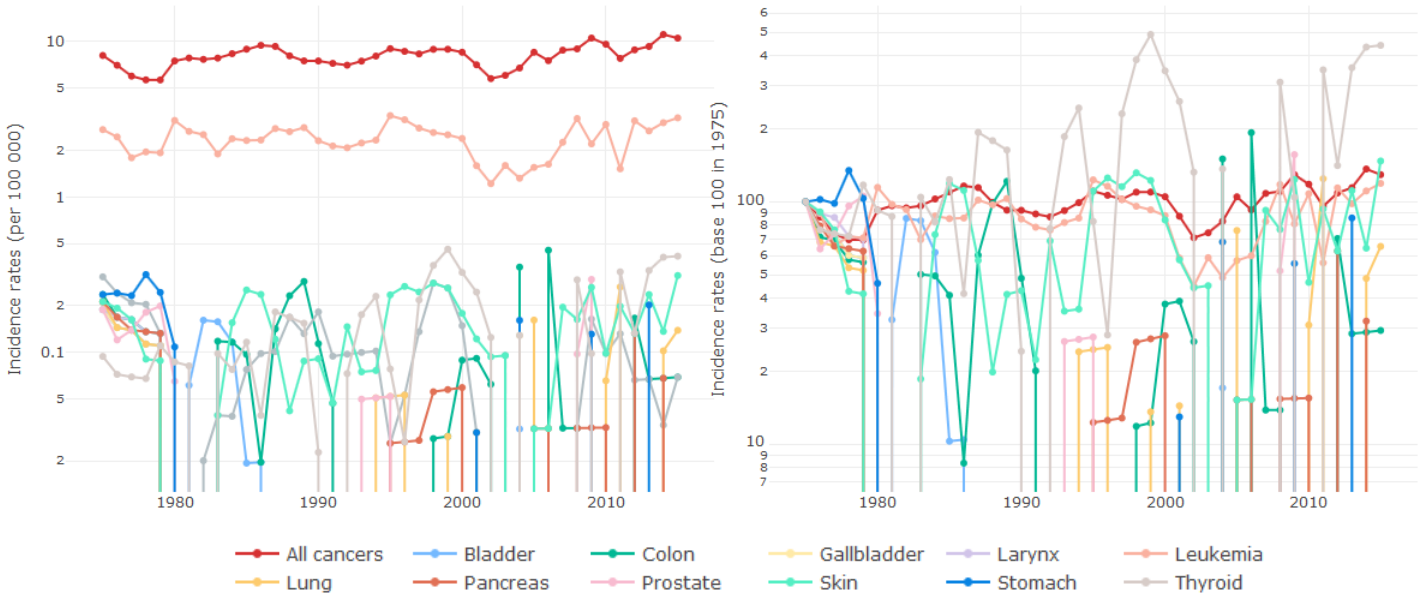
For the age group 85+, it should be analyzed remembering that the average age may not be stable along time. Due to the life expectancy rise, the average age of people aged more than 85 may not be constant over time.

D.1 Japanese Data - data per age group for male

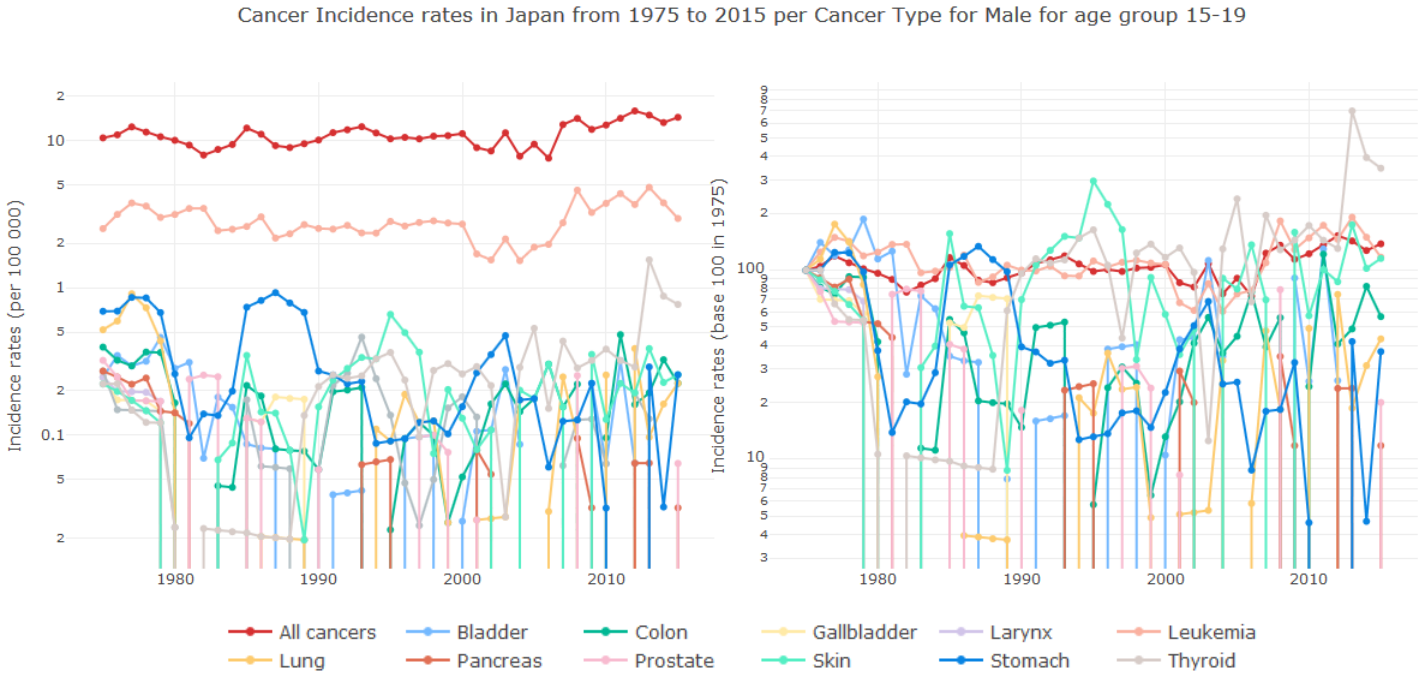


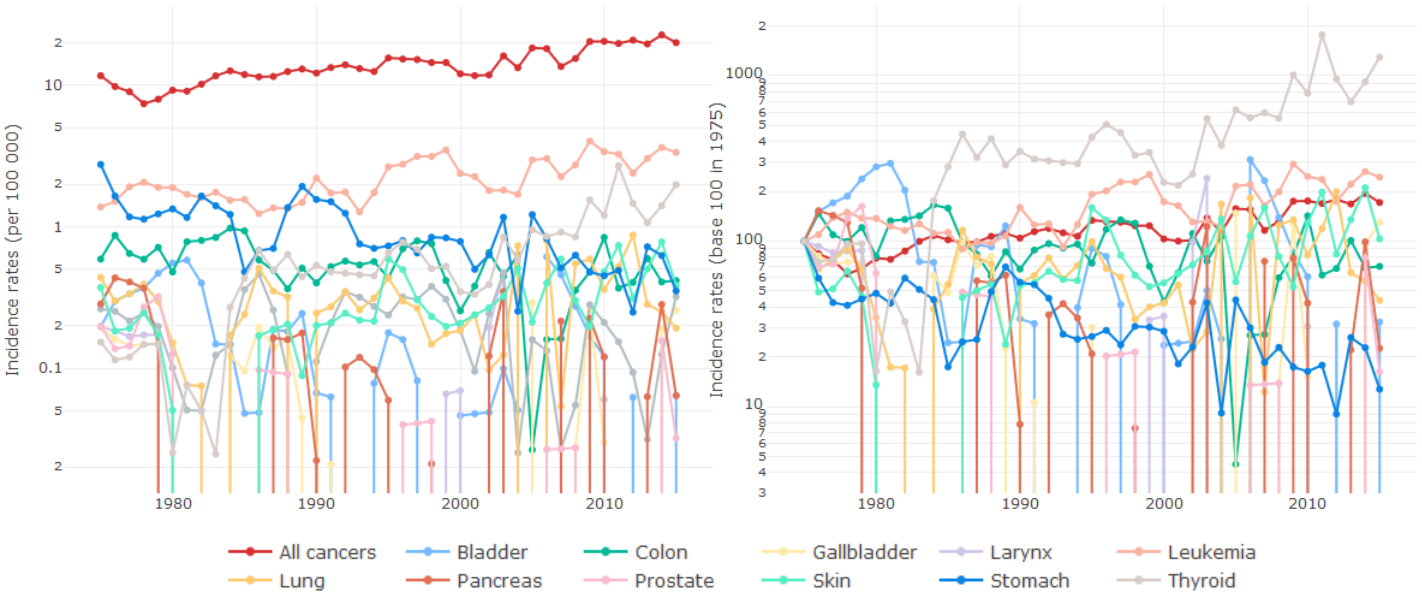
For male children aged 0 to 4, cancer all causes annual incidence rate has been around 18 for 100,000. As in Korea, the most common cancer is leukemia and the second one thyroid cancer. Other cancer sites incidence is very small and erratic. For this age class and other young ages, the sum of cancer incidence per site does not match the cancer incidence all causes in this database.



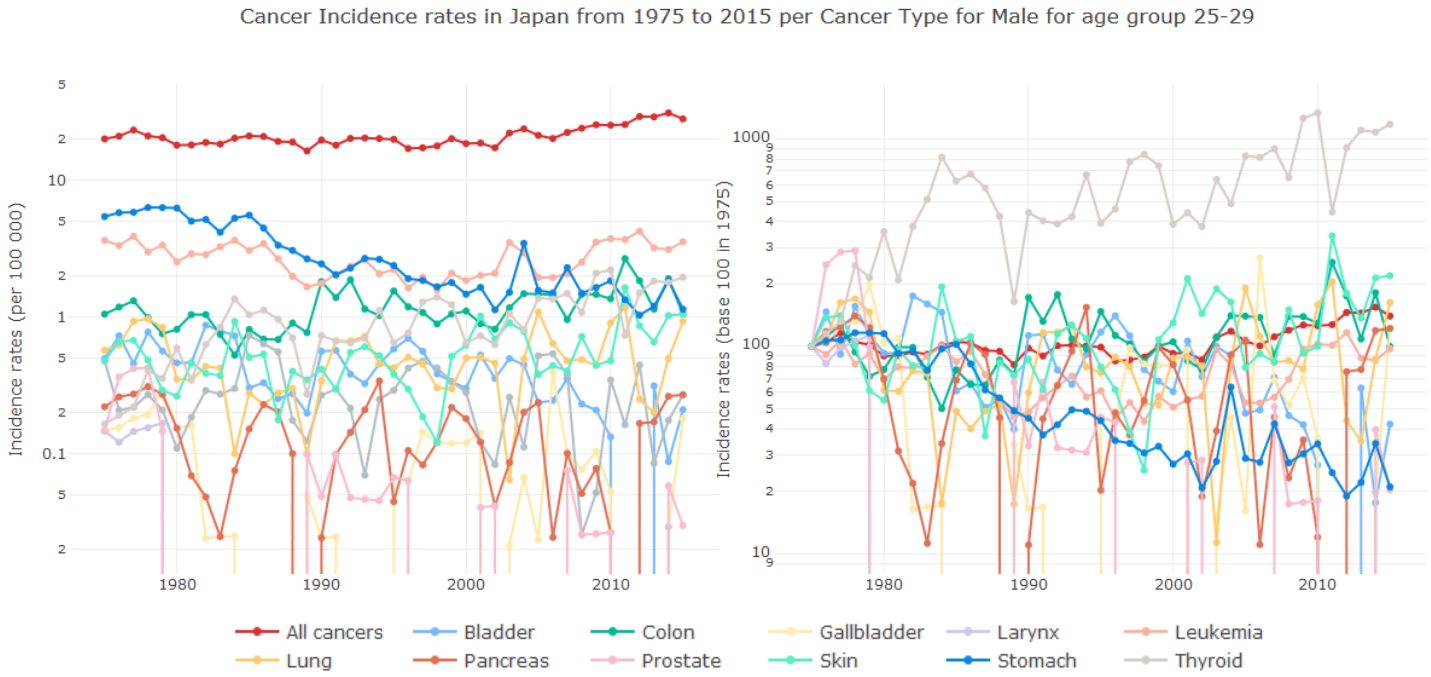


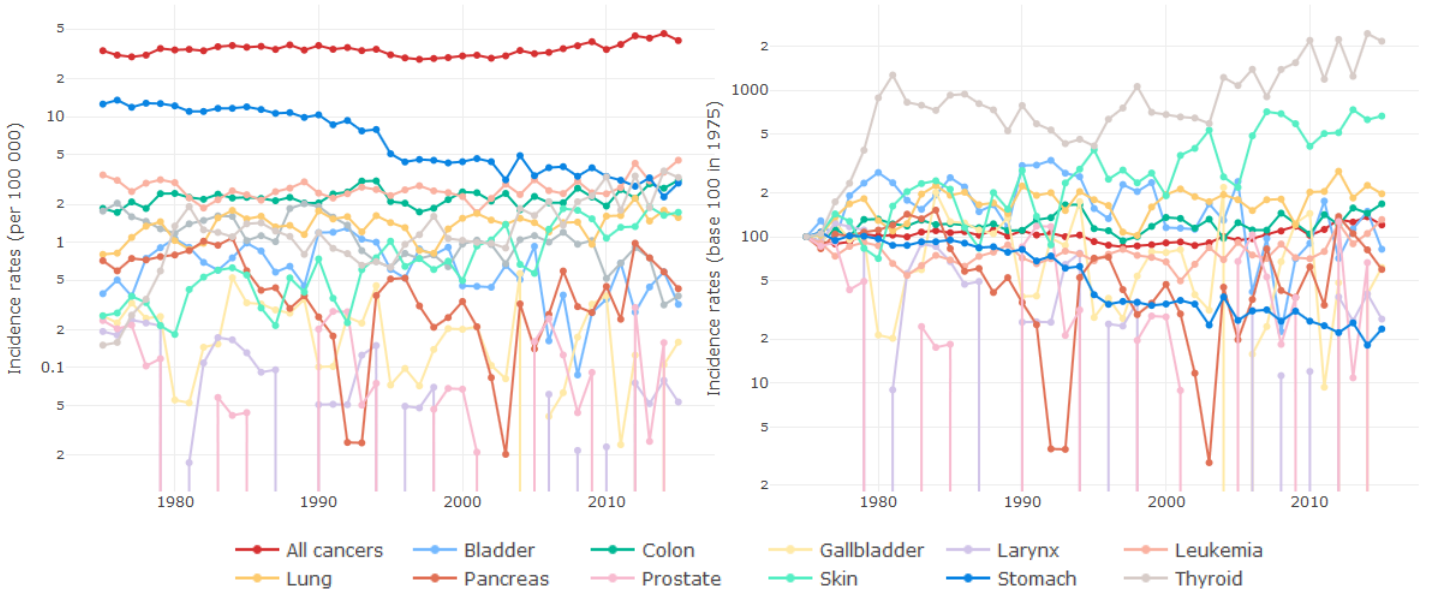
Cancer all causes annual incidence rate reaches 10 for 100,000 for both male children aged 5 to 9 and 10 to 14. Again, the most common cancer is leukemia.



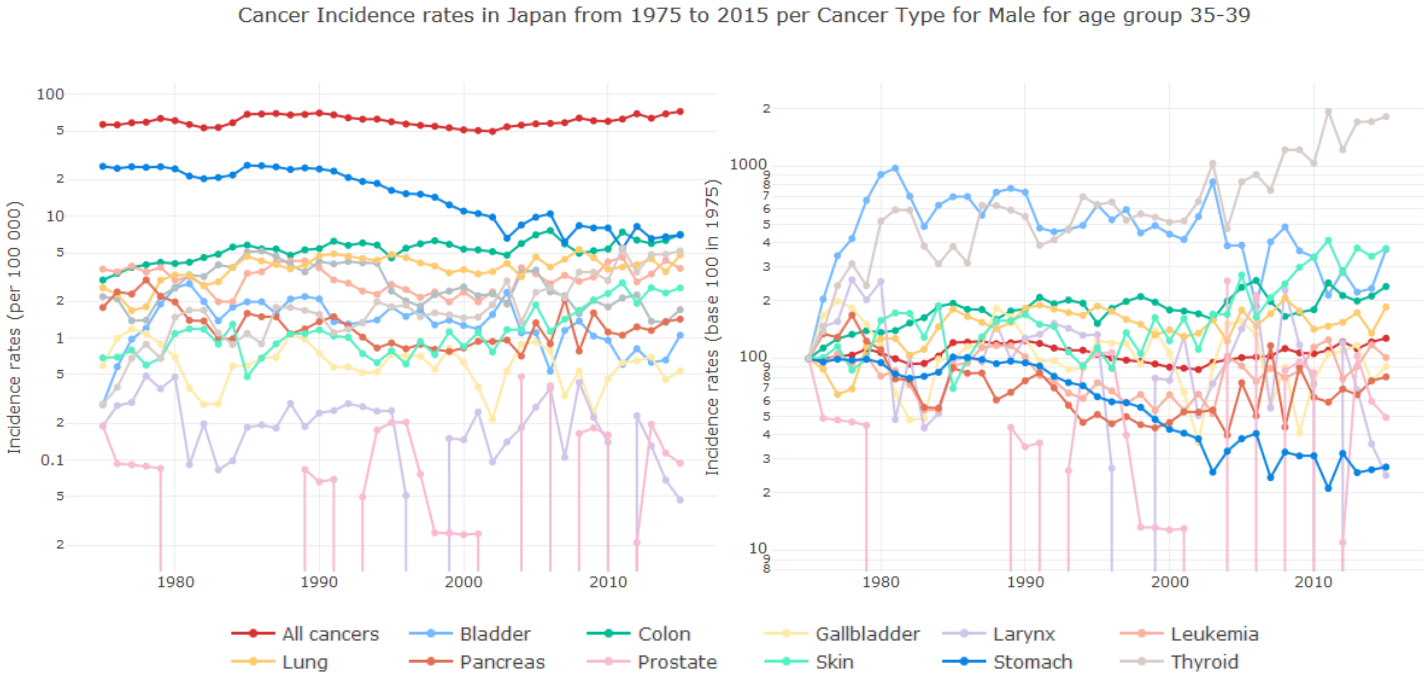


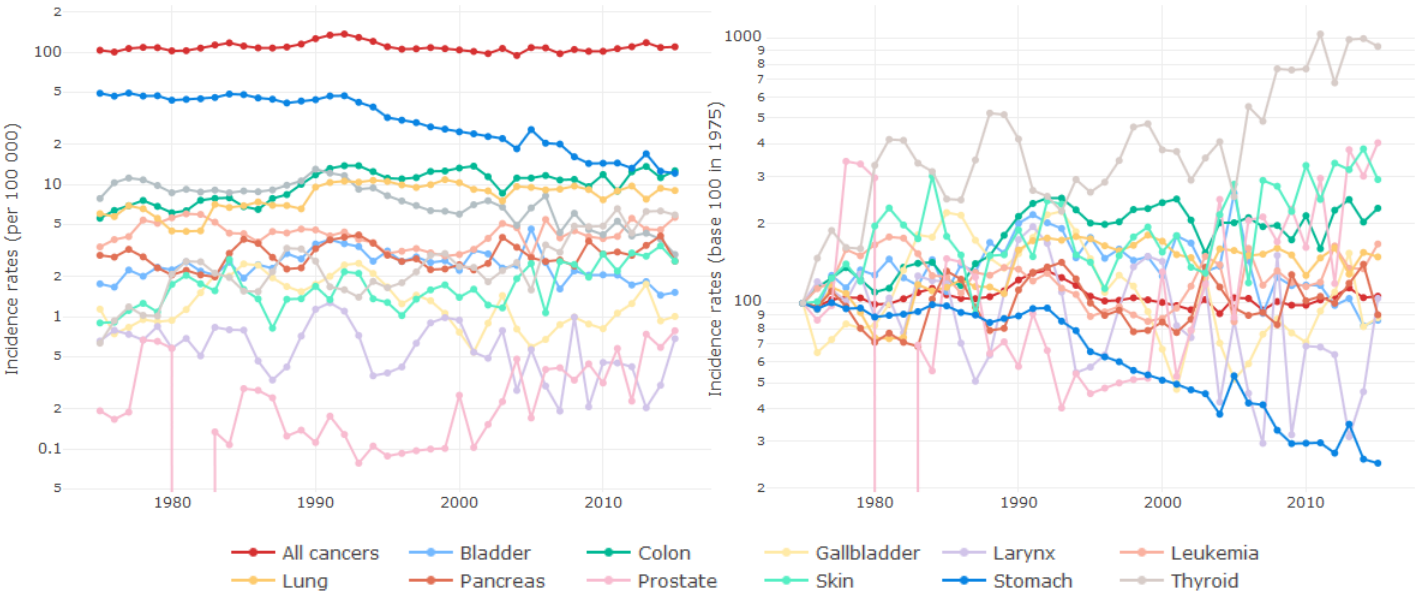
Cancer all causes annual incidence rate reaches respectively 14 and 20 per 100,000 for male children aged 15 to 19 and 20 to 24. Both show an upward trend. The most common cancer is leukemia and then thyroid cancer.



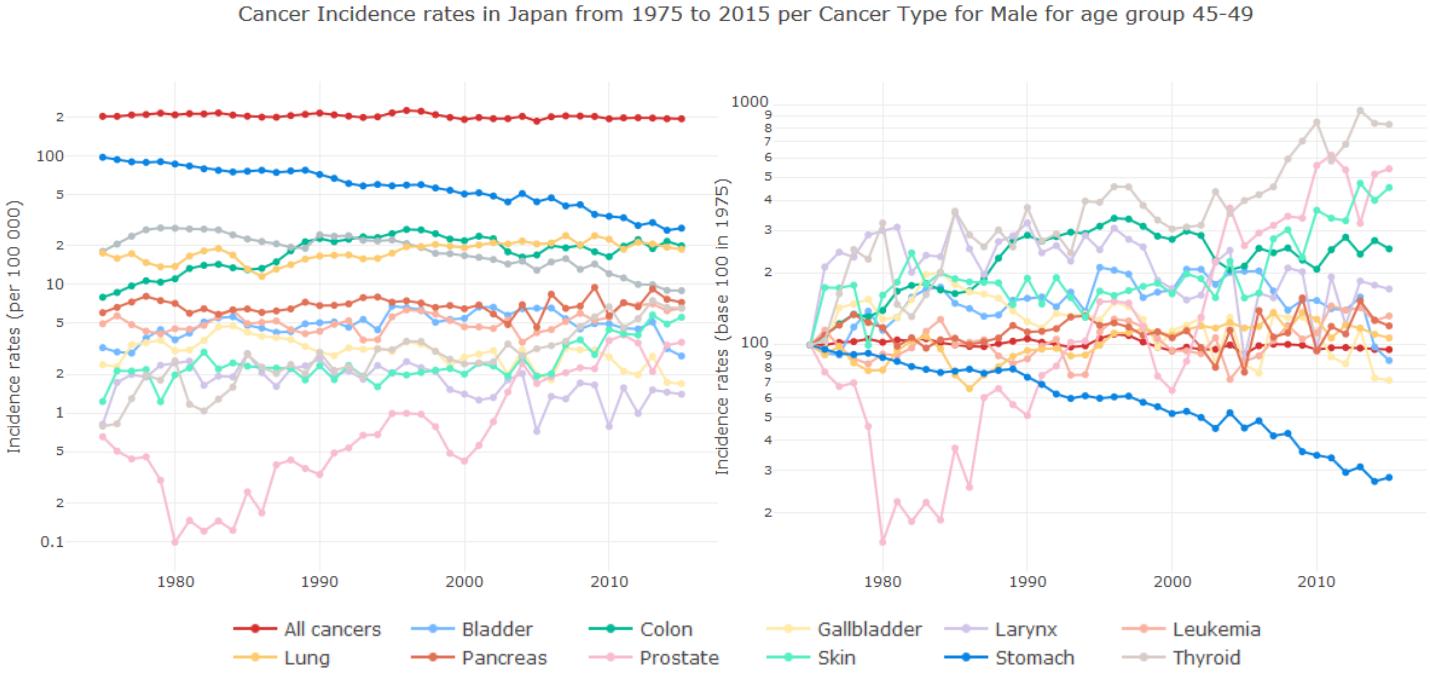


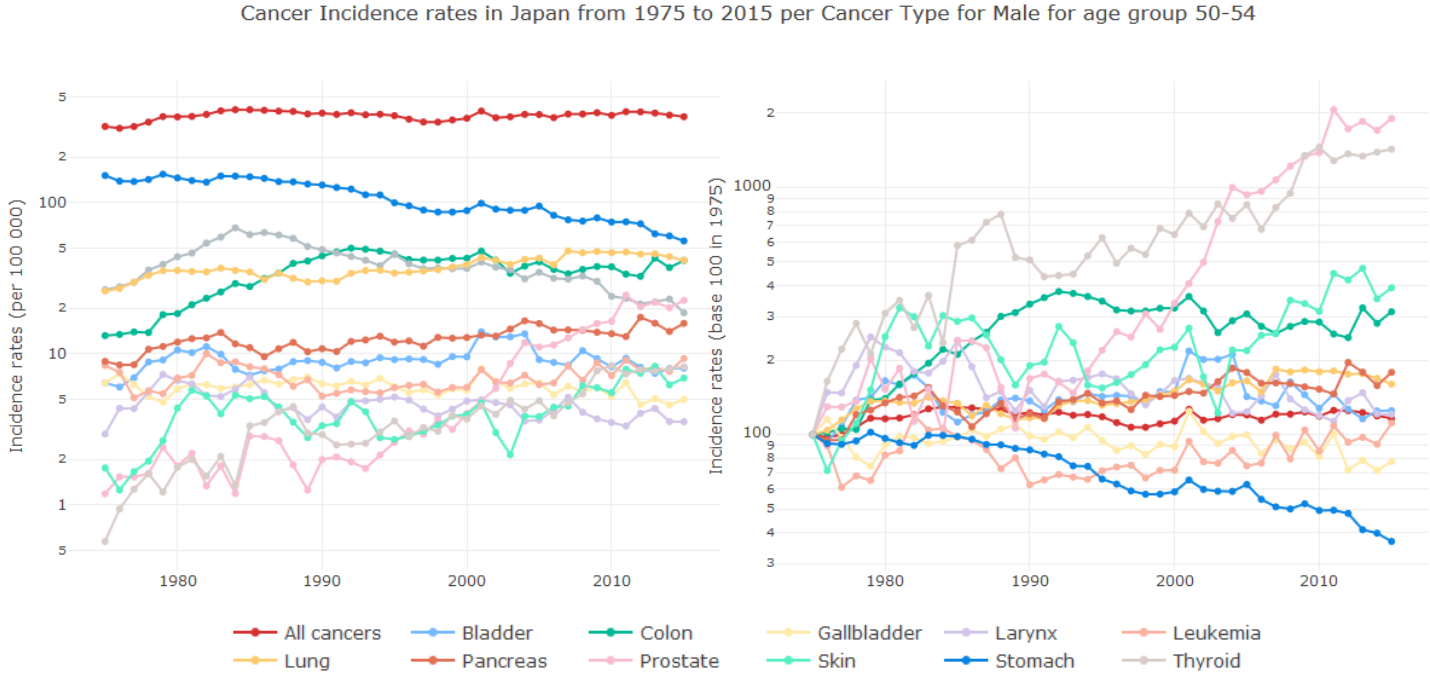
Cancer all causes annual incidence rate reaches respectively 28 and 40 per 100,000 for male aged 25 to 29 and 30 to 34. Both show an upward trend. The most common cancer has historically been stomach cancer, for which incidence has strongly decreased over time. Leukemia is the most frequent cancer in 2015.



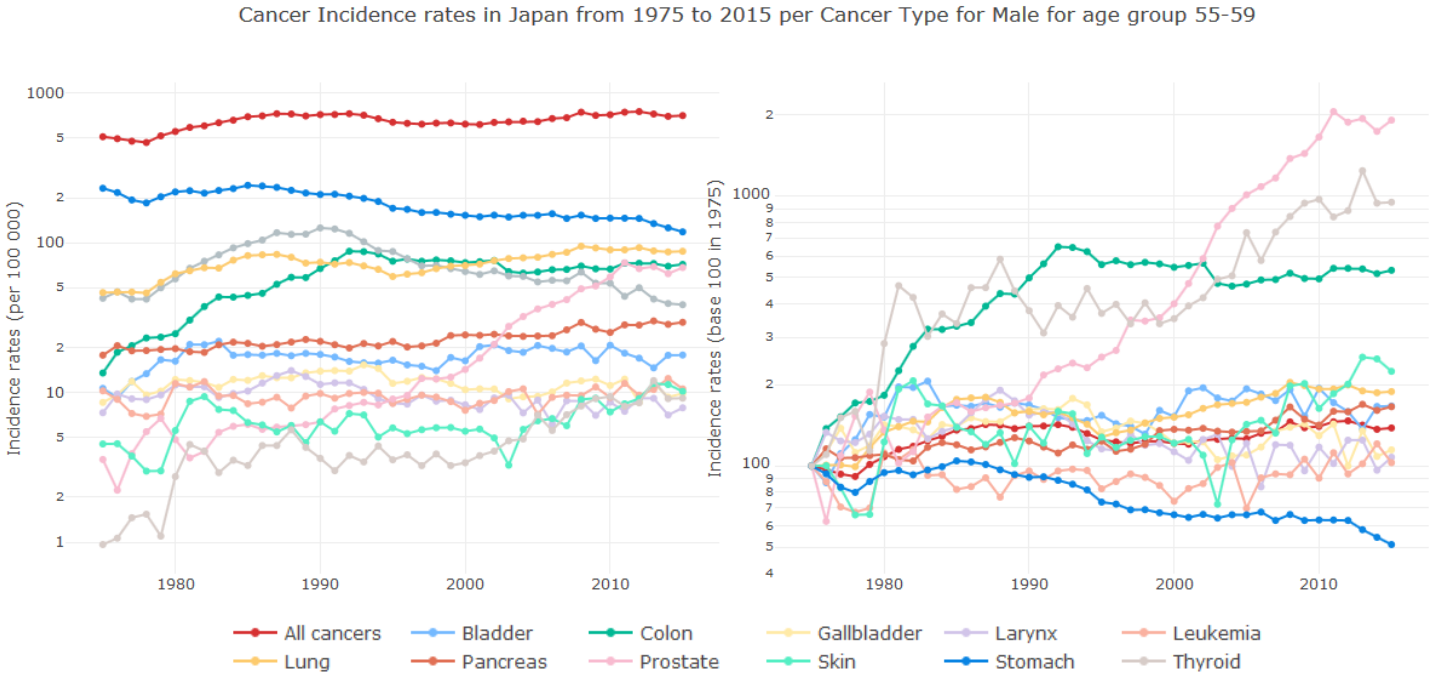


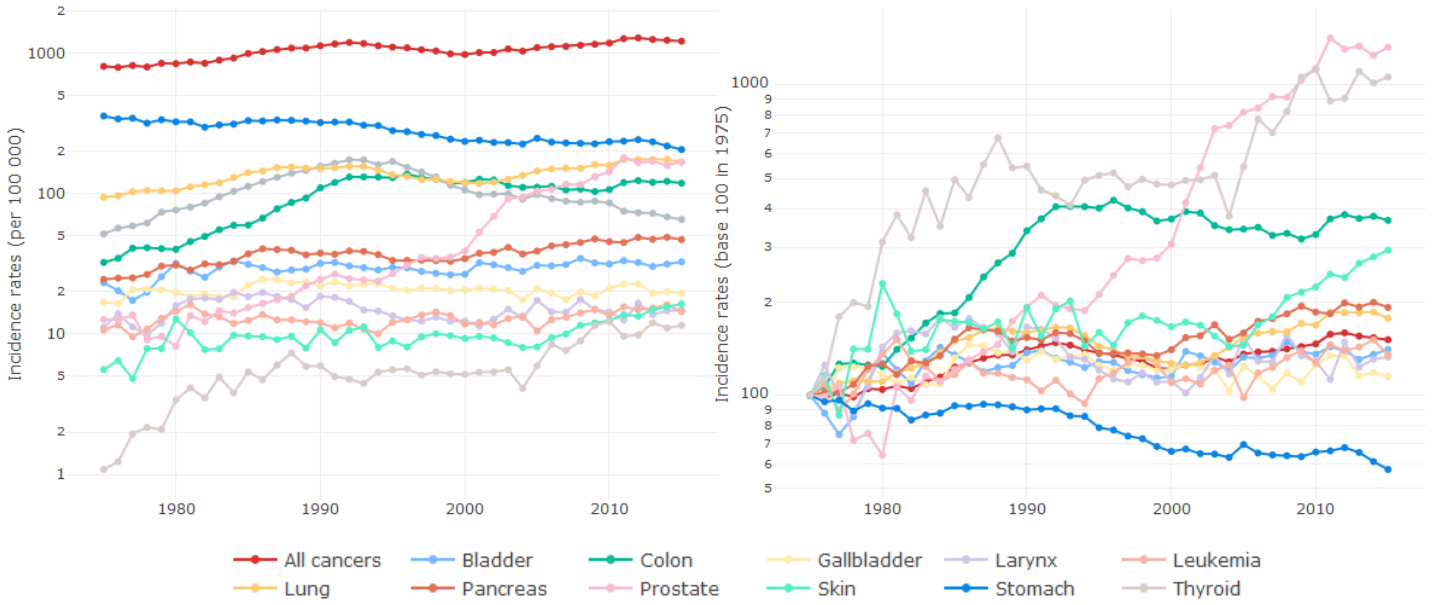
Cancer all causes annual incidence rate reaches respectively 72 and 110 per 100,000 for male aged 35 to 39 and 40 to 44. Both show some variations during the study period. Stomach cancer has strongly decrease, while colon cancer has became the most common cancer.



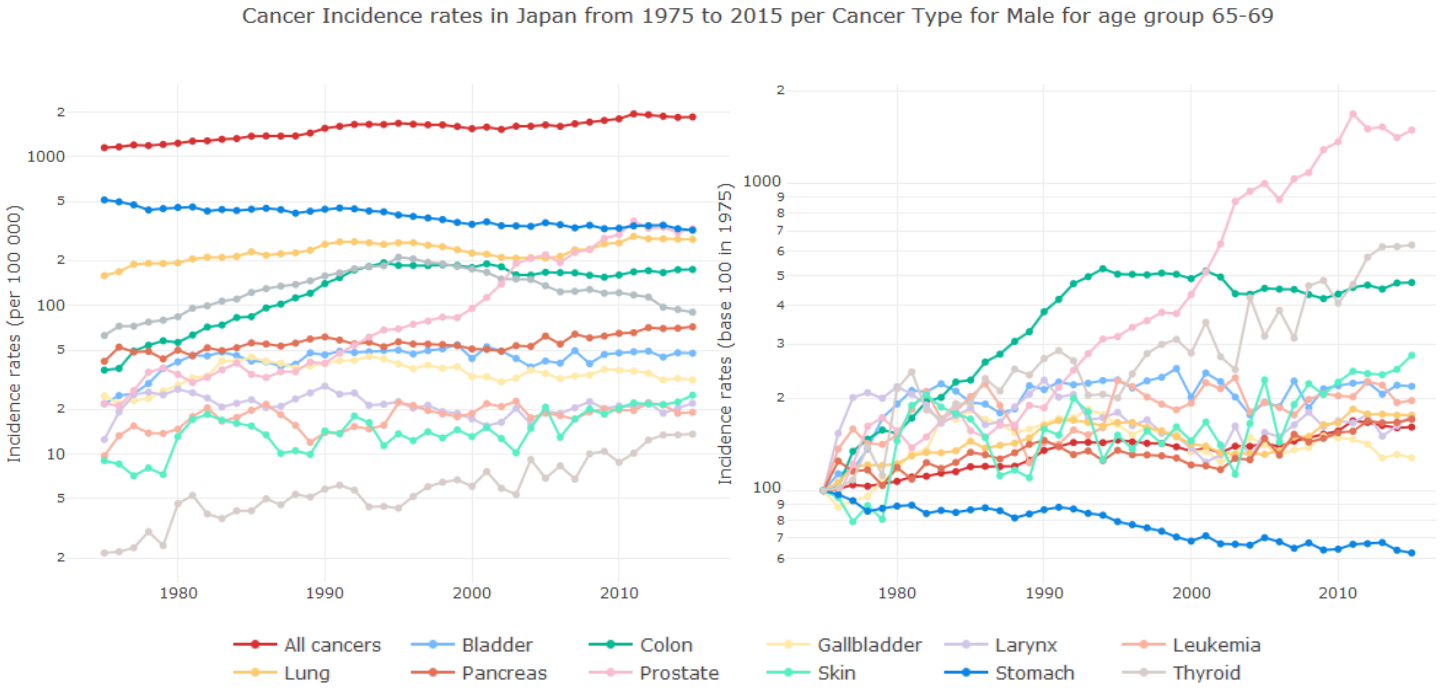


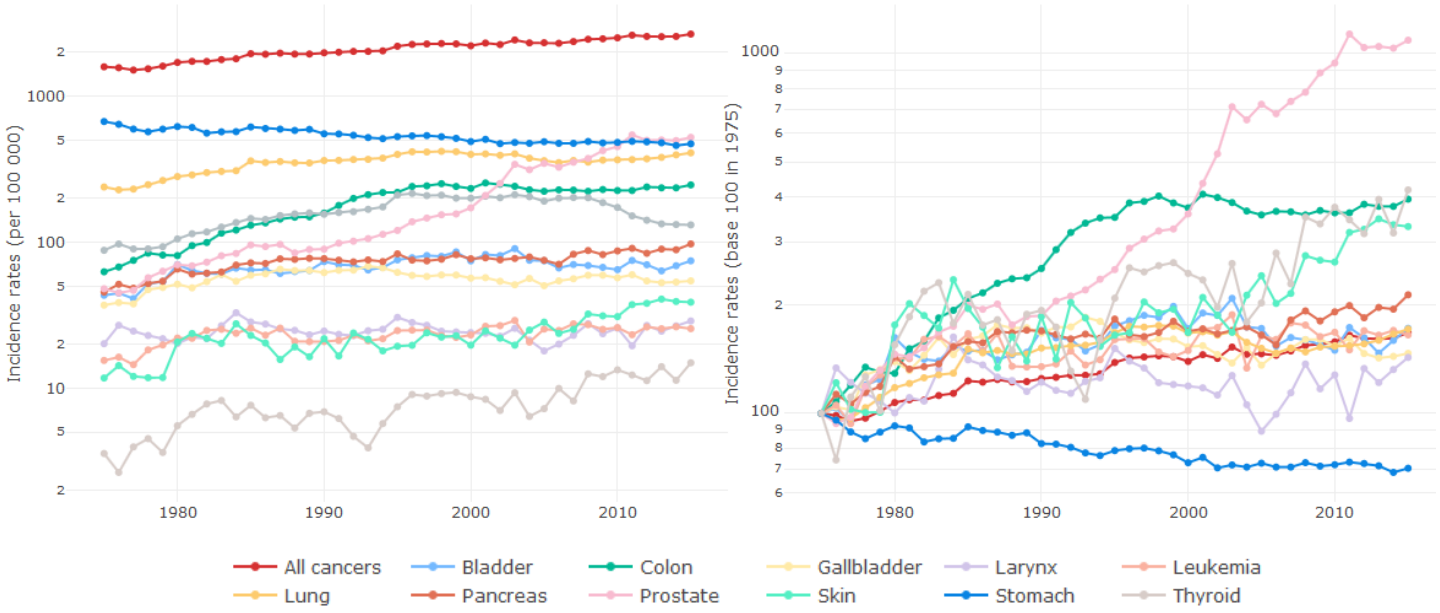
Cancer all causes annual incidence rate reaches respectively 2 and 3.7 per 1,000 for male aged 45 to 49 and 50 to 54. As the previous age classes, stomach cancer has decreased, while colon cancer has strongly increased. Colon and lung cancer are the most frequent cancer sites.



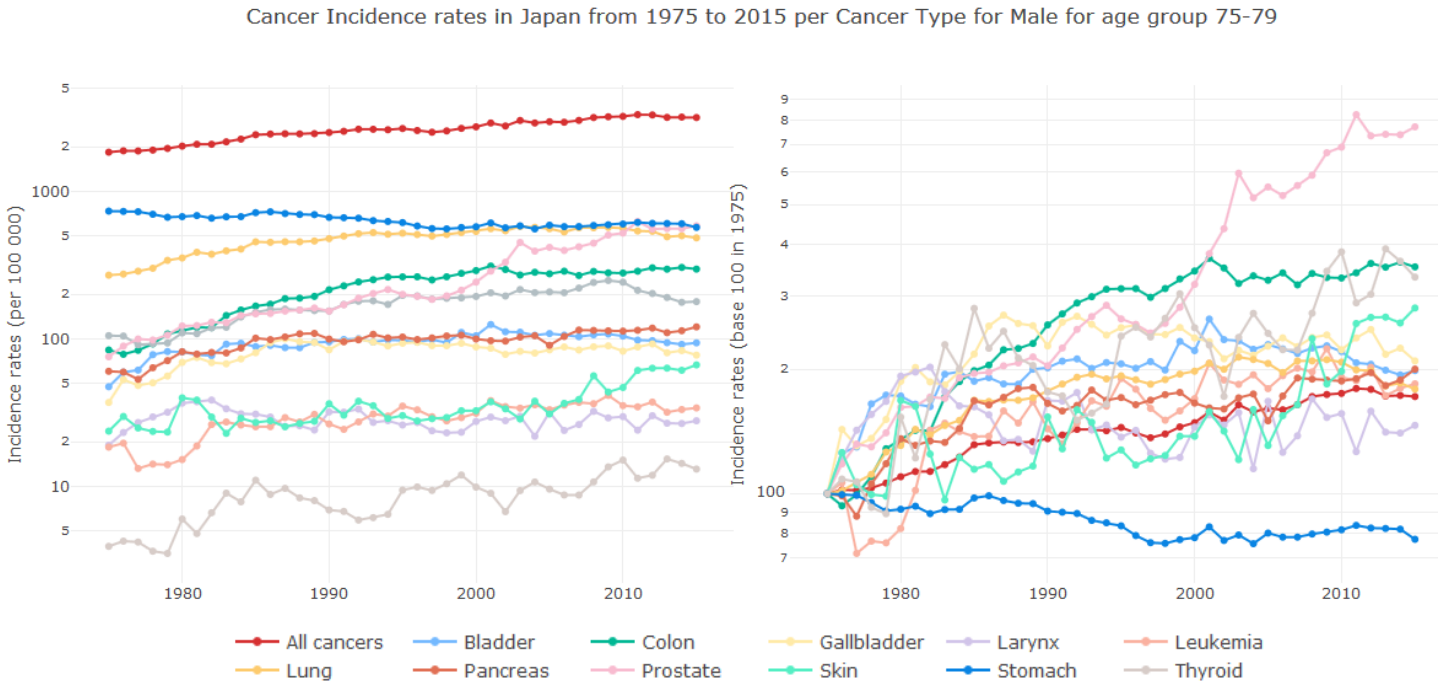


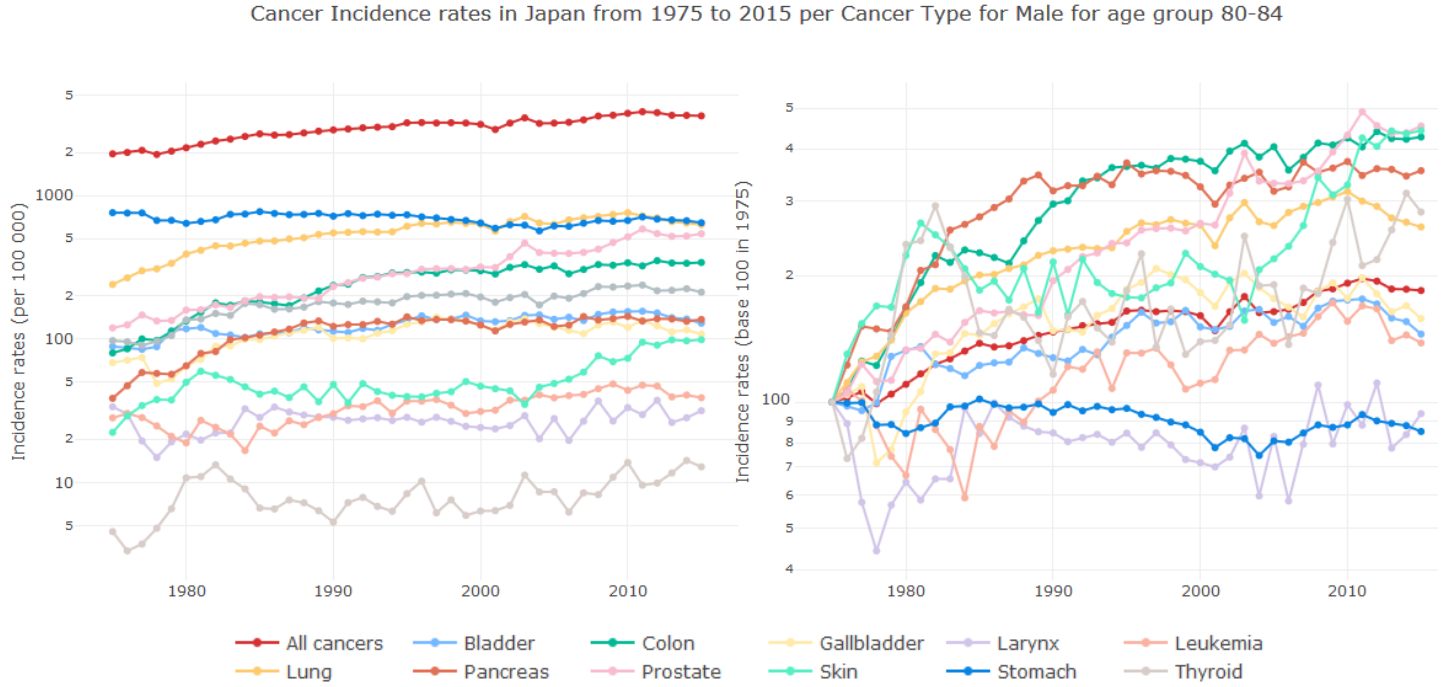
Cancer all causes annual incidence rate reaches respectively 7 per 1,000 and 1.2 per 100. Changes are similar with the previous age classes in the dynamics of stomach, colon and lung cancer. Prostate cancer has radically increased from 2000.



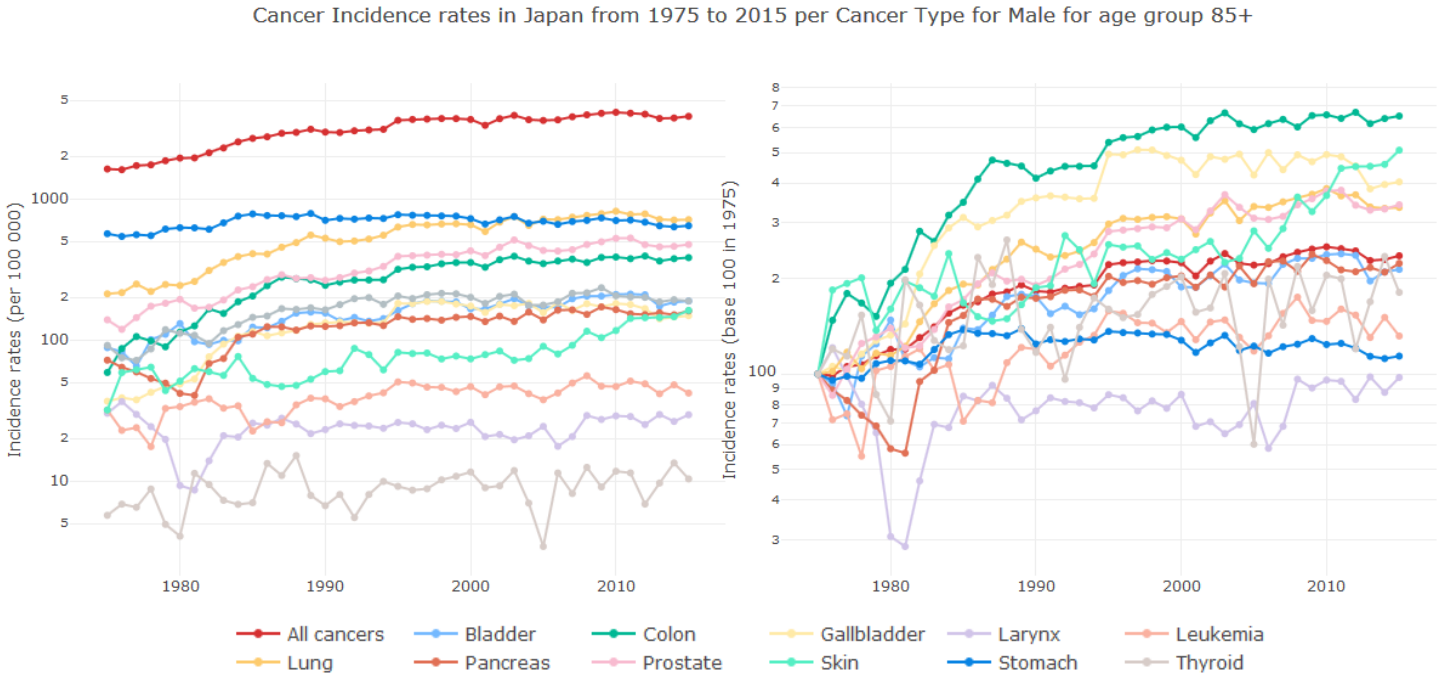


Cancer all causes annual incidence rate reaches respectively 1.8 and 2.7 per 100 for male aged 65 to 69 and 70 to 74. Prostate and lung cancers have become the most frequent cancer while stomach cancer has decreased. Colon cancer have strongly increased until 1995 while prostate cancer strongly increased from 2000.



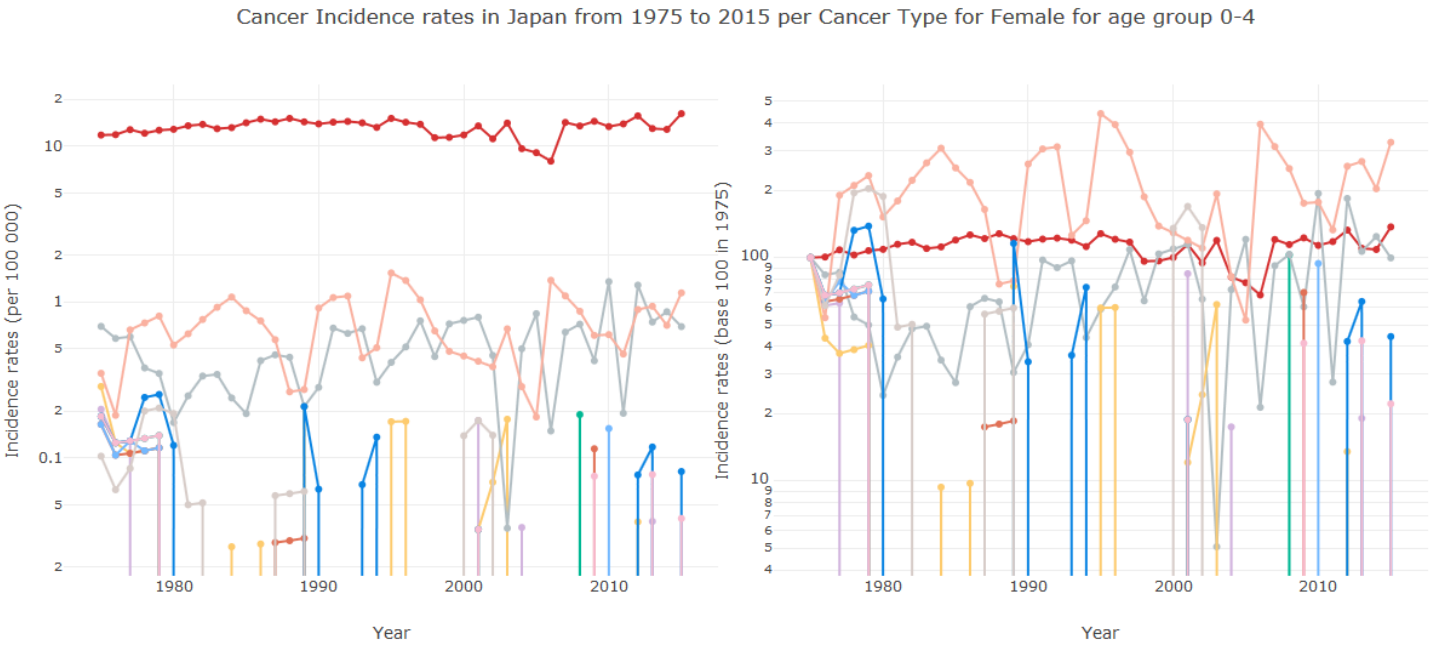


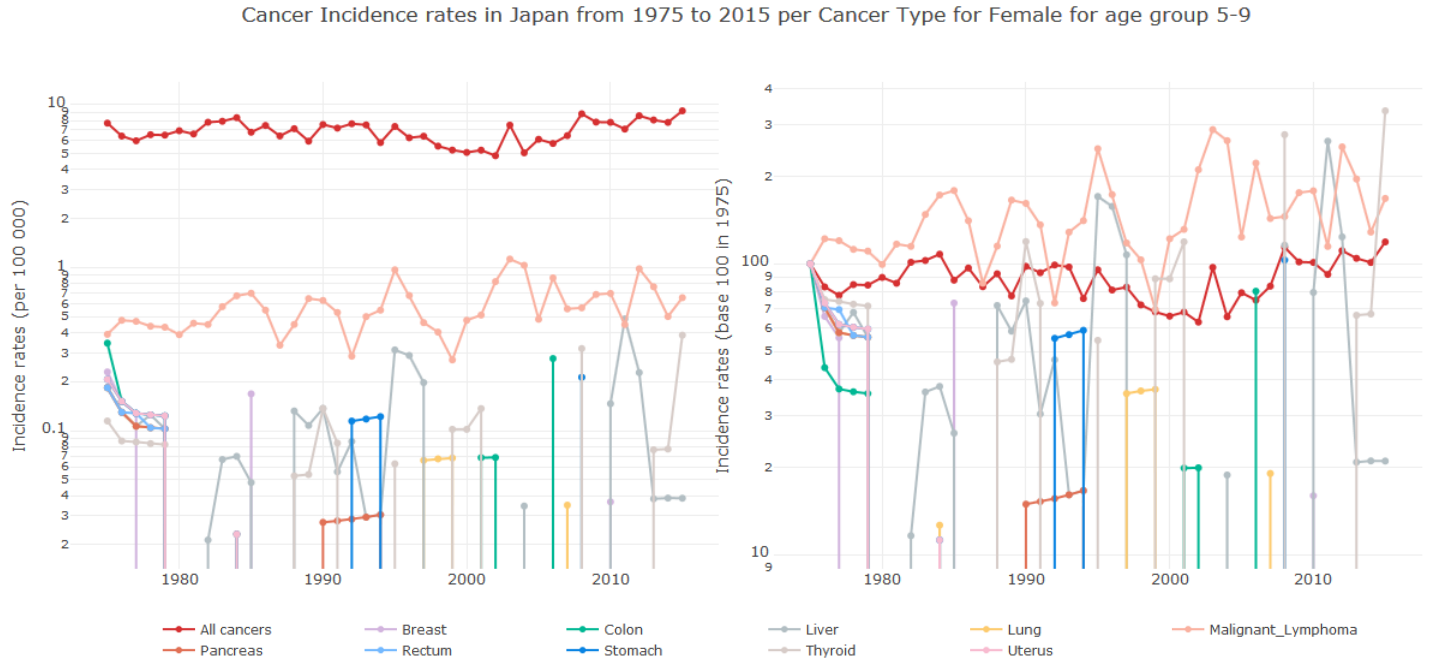
Cancer all causes annual incidence rate reaches respectively 3.2 and 3.6 per 100 for male aged 75 to 79 and 80 to 84. Prostate (especially for age 75-79) and colon have strongly increased. Pancreas cancer increased sharply for age 80 to 84 until 1990.



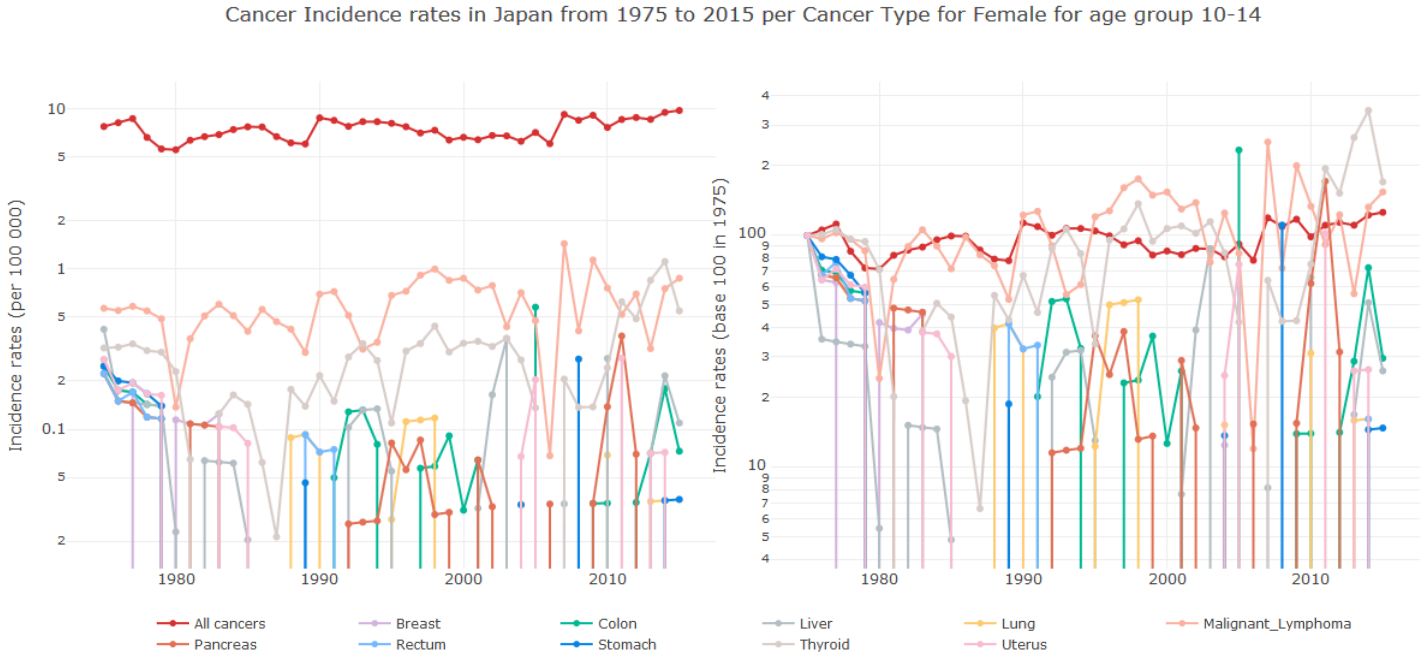
Cancer all causes annual incidence rate reaches 3.9 per 100 for male aged 85 or more and has more than doubled during the study period. Most frequent cancer in 1975, stomach cancer incidence is relatively constant. Lung cancer has been the most frequent cancer from 2000.

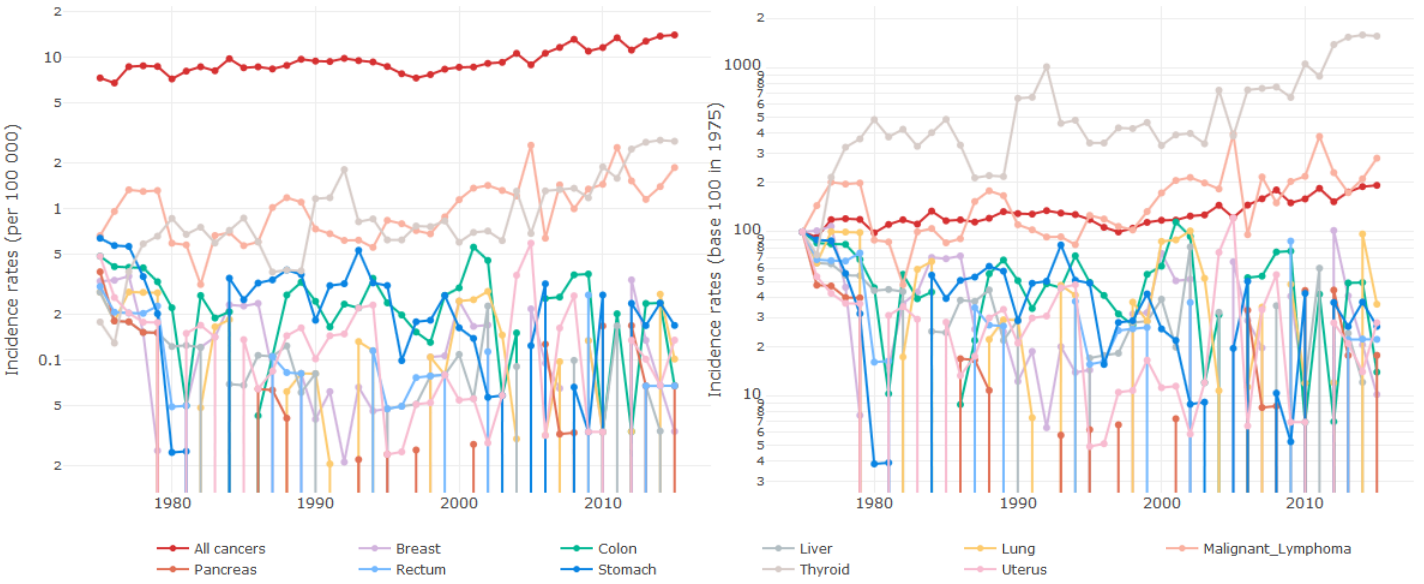
D.2 Japanese Data - data per age group for female



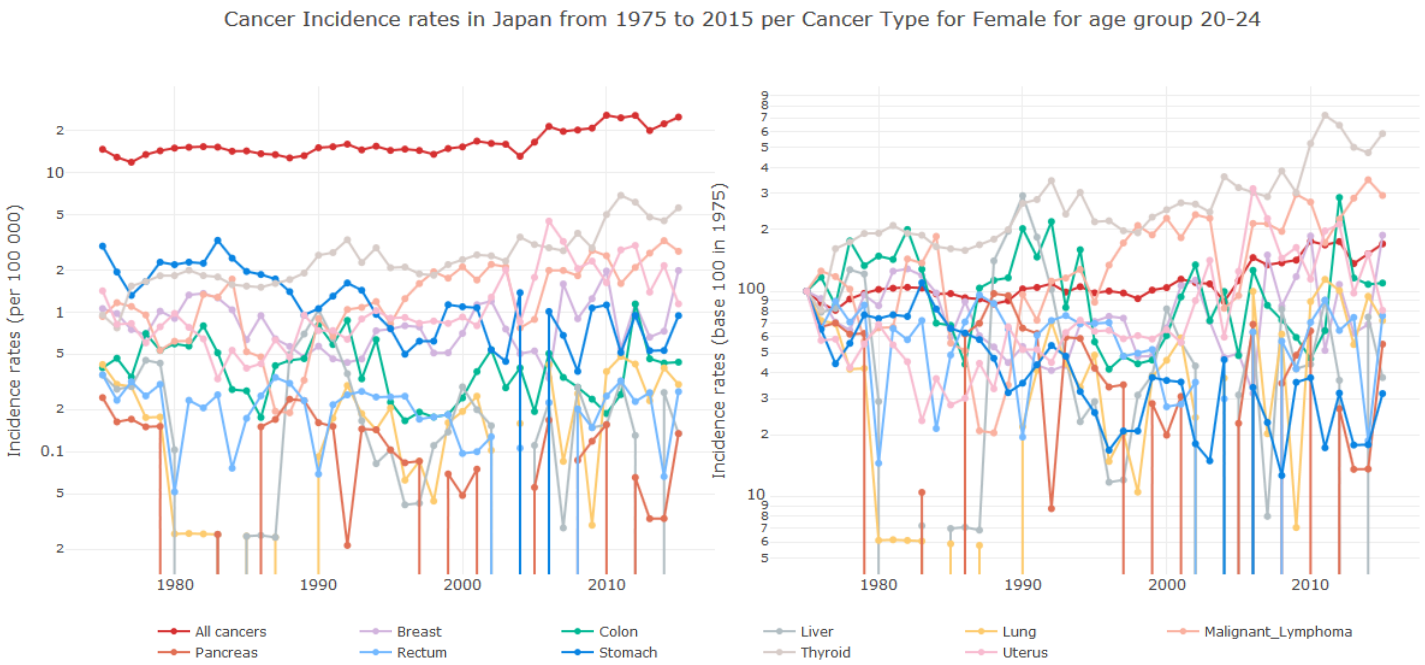


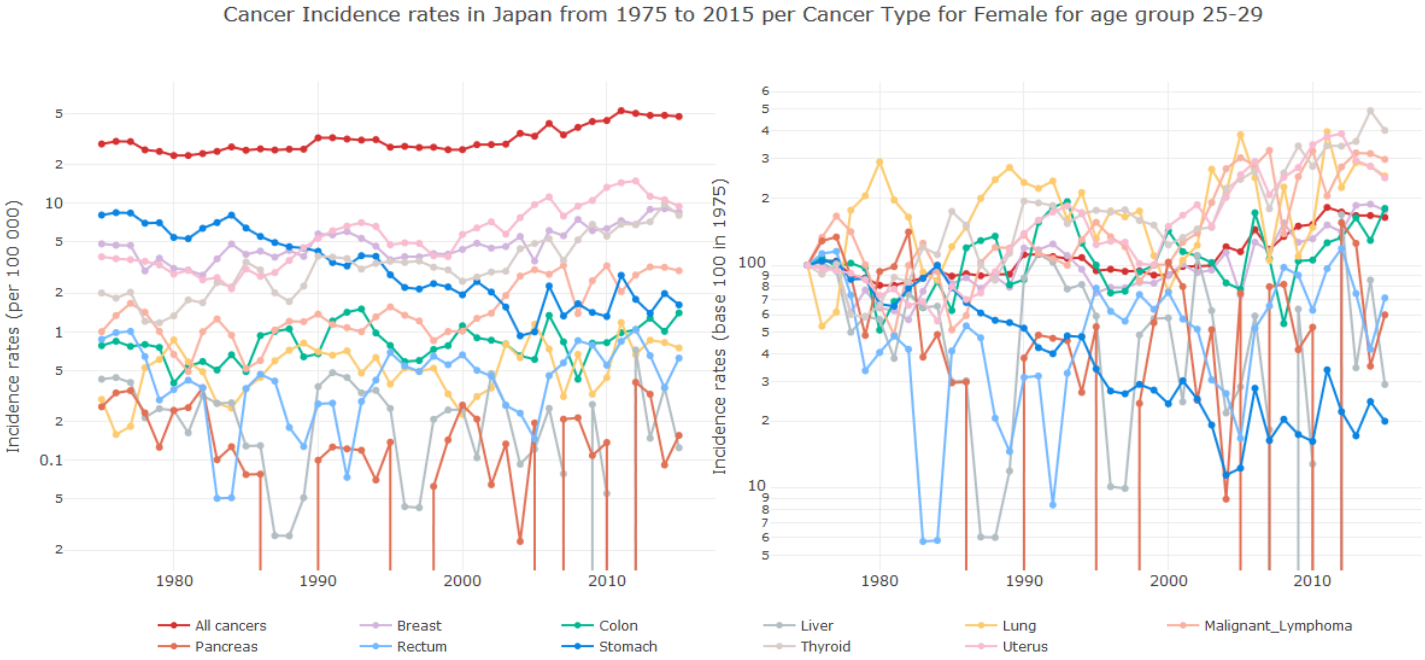
Cancer incidence all causes is relatively constant over the time period, with a slight increase of the past recent years. Leukemia (not represented in the figure) is the most common cancer while other sites are rare.



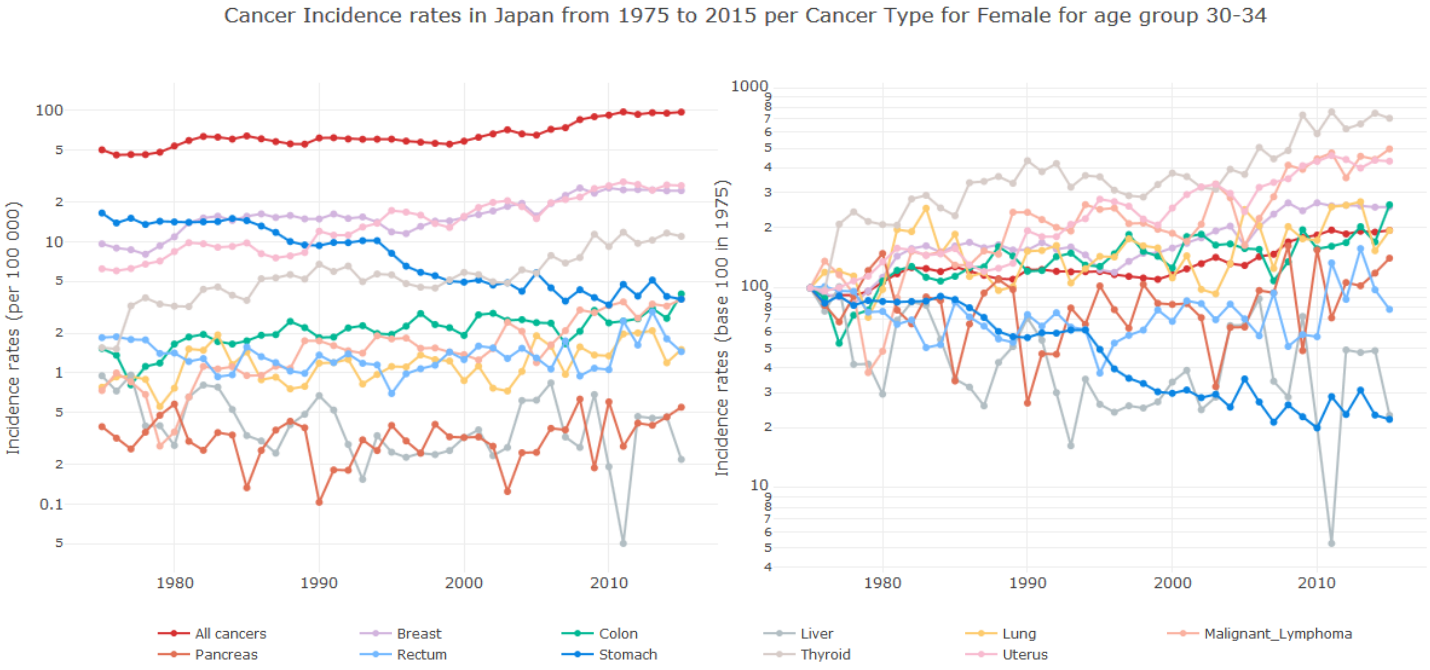


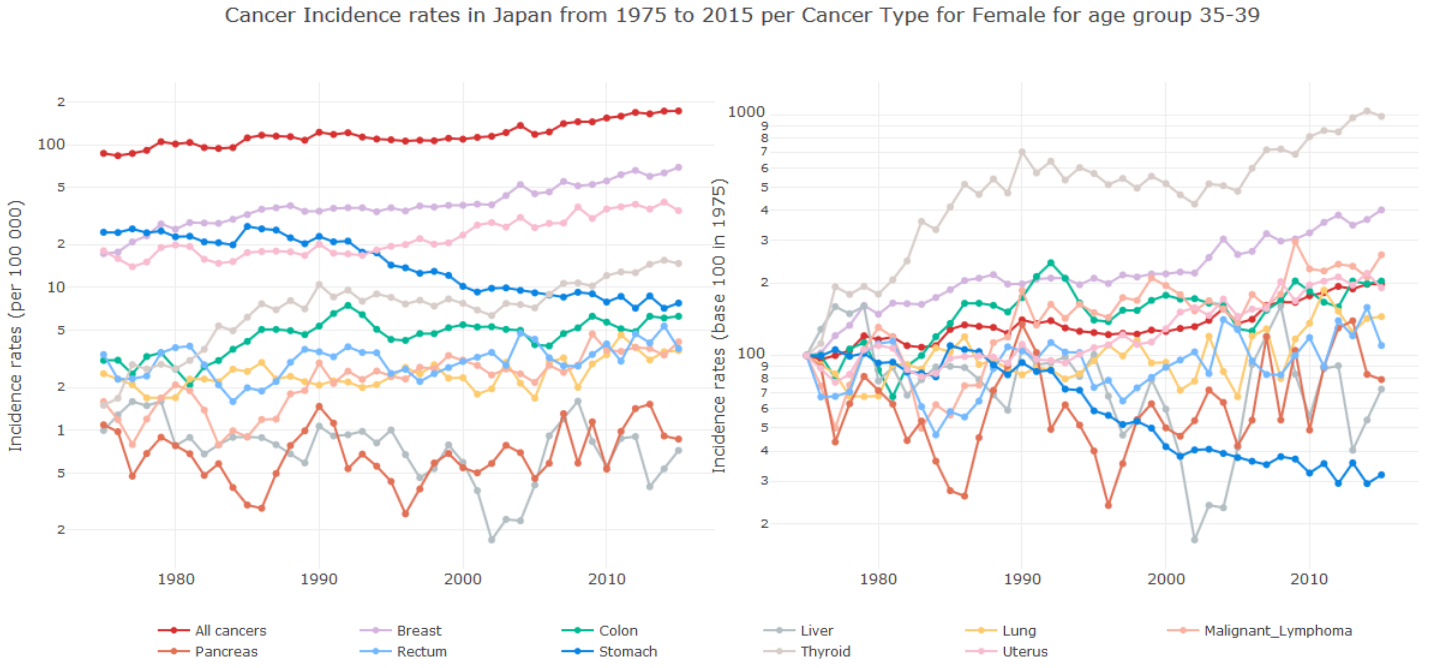
Cancer all causes annual incidence rate reaches respectively 10 and 14 per 100,000 for female children aged 10 to 14 and 15 to 19. Both do not show clear trend. The changes of incidences rates per site are quite erratic, except thyroid cancer for age group 15-19 which show a large increase through time.



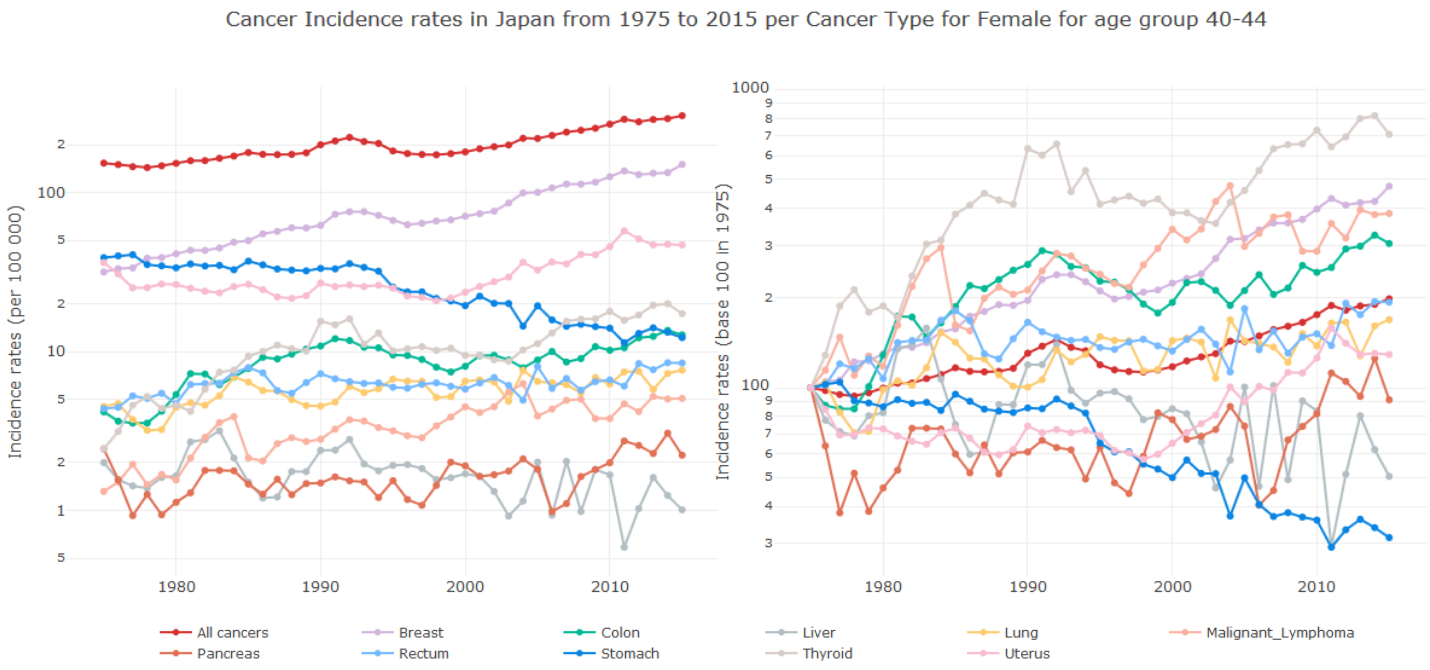


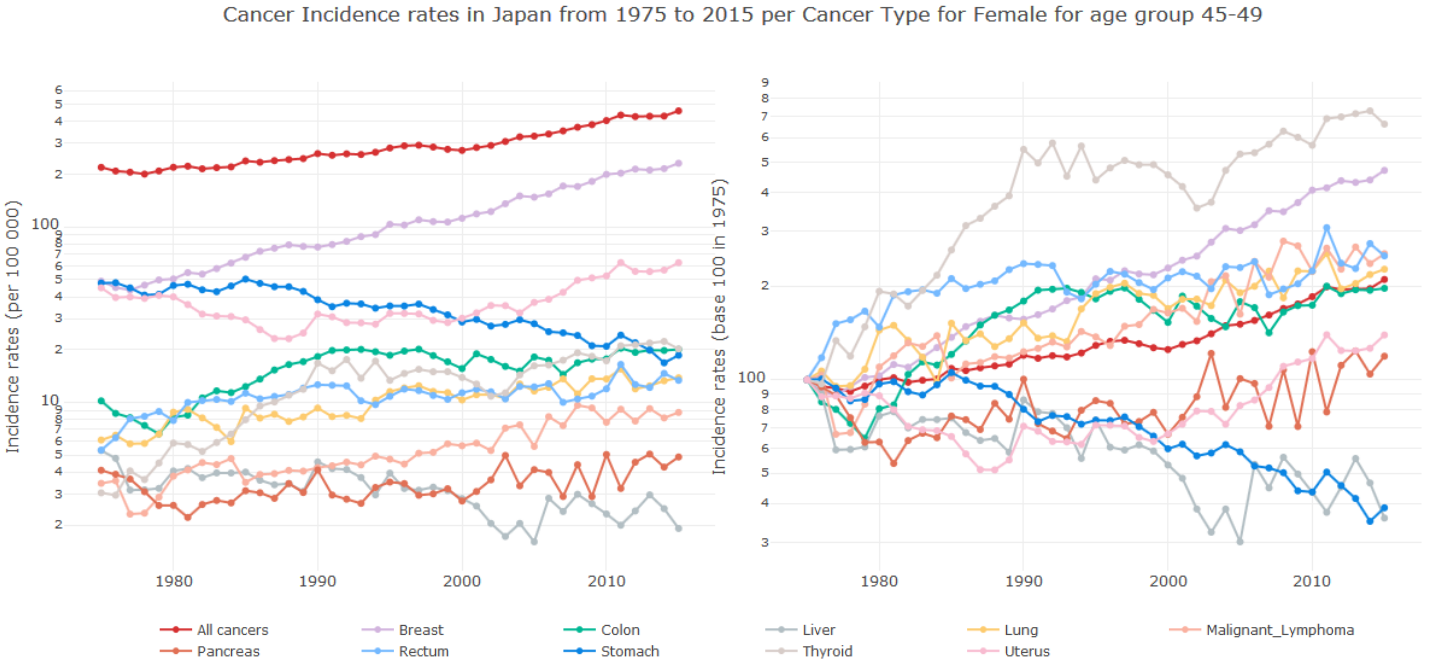
Cancer all causes annual incidence rate reaches respectively 25 and 48 per 100,000 for female aged 20 to 24 and 25 to 29. Both show an upward trend from year 2000. Stomach cancer incidence has decreased while thyroid, malignant lymphoma, uterus and breast are the most common cancer sites.



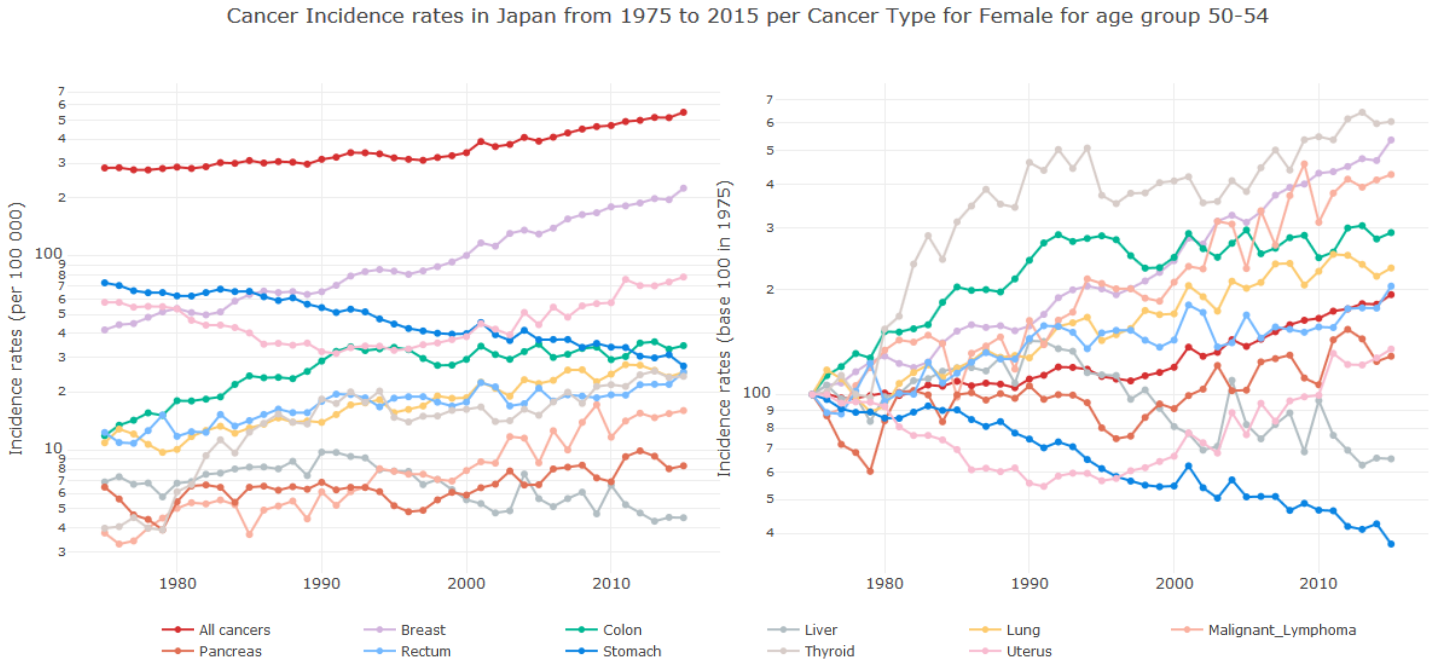


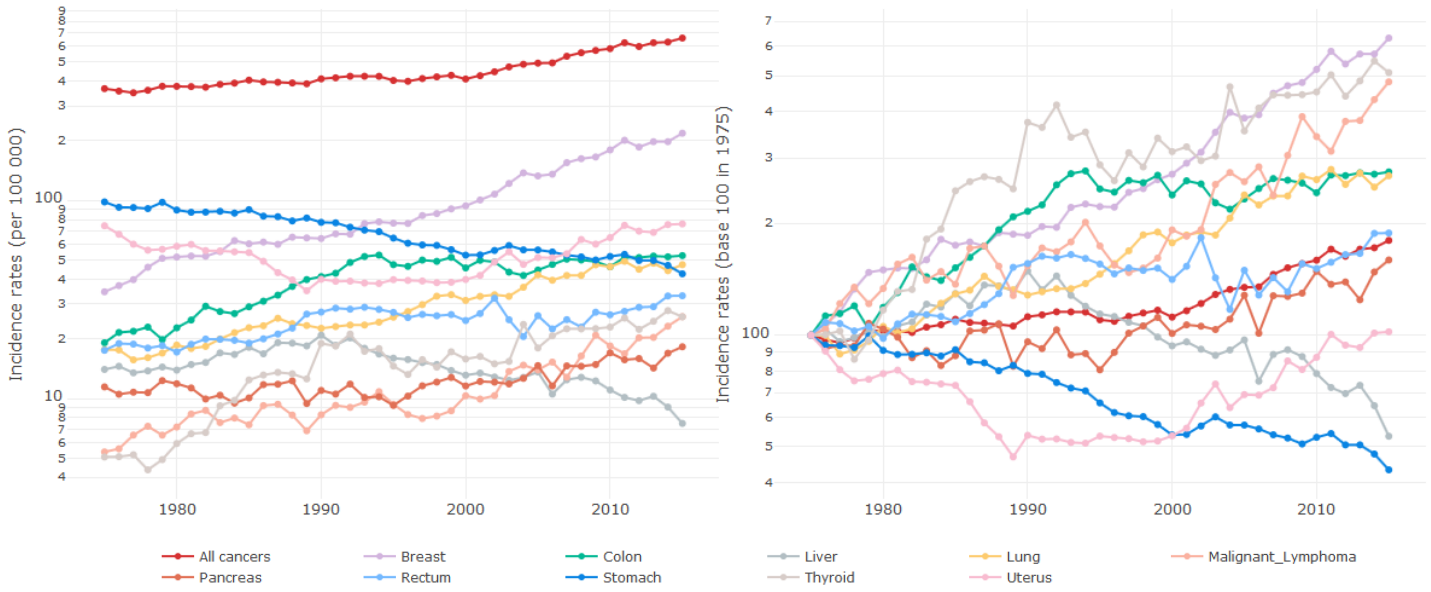
Cancer all causes annual incidence rate reaches respectively 1 and 1.7 per 1,000 for female aged 30 to 34 and 35 to 39. Both show an upward trend in the recent years and he most common cancer sites are uterus and breast cancer, while a large increase over the time period. Thyroid cancer incidence has been multiplied by 10 from 1975 to 2015.



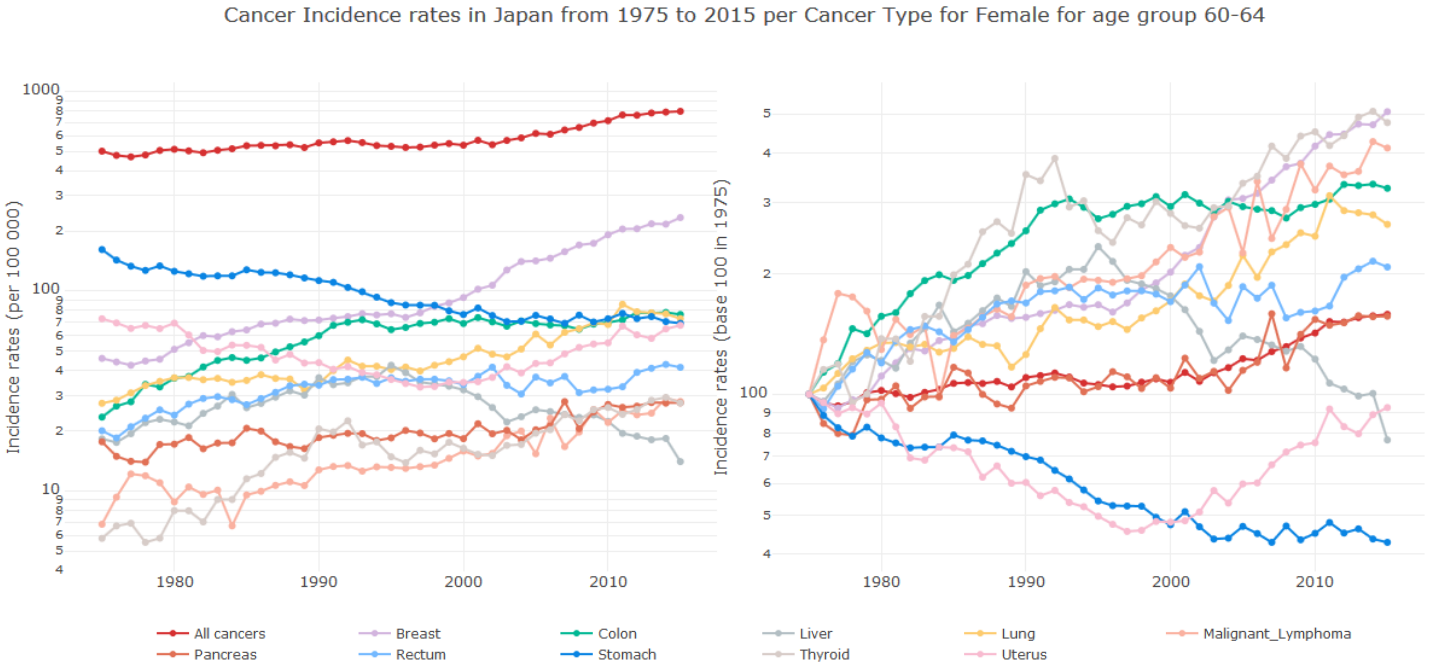


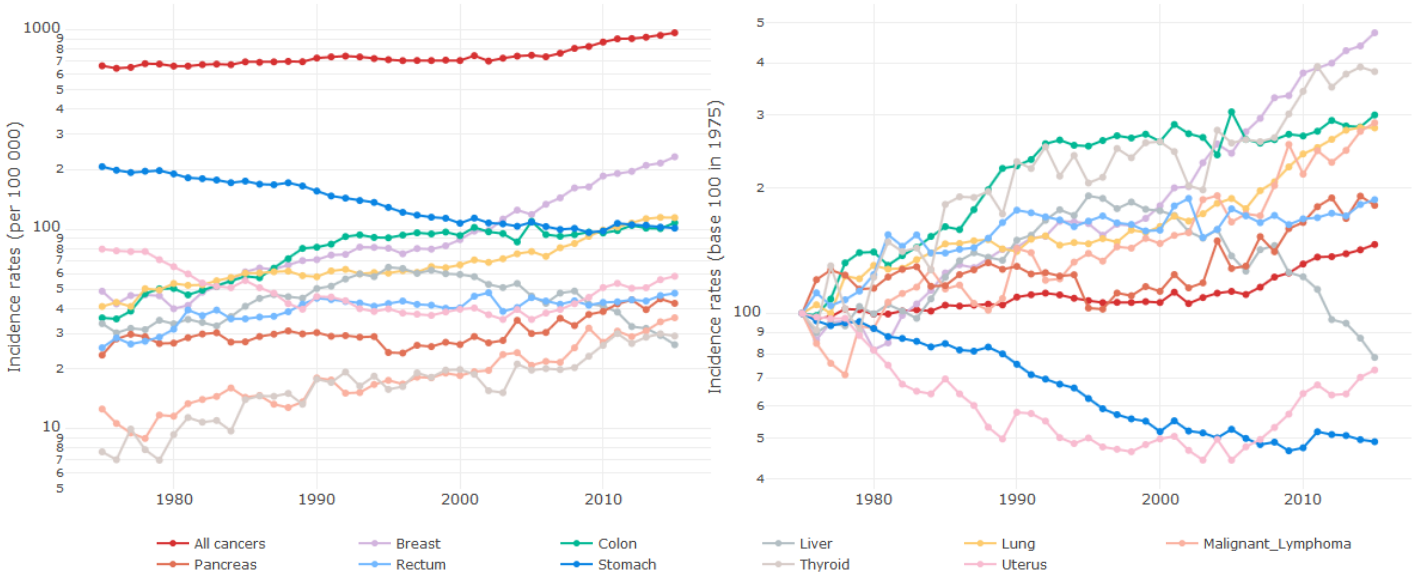
Cancer all causes annual incidence rate reaches respectively 3 and 4.6 per 100 for female aged 40 to 44 and 45 to 49. Originally stomach, uterus and breast cancer were the most prevalent cancer. While stomach cancer has decreased and uterus cancer is constant (with first a decrease, then an increase to return to the original level), breast cancer incidence has been multiplied by 5 over the study period.



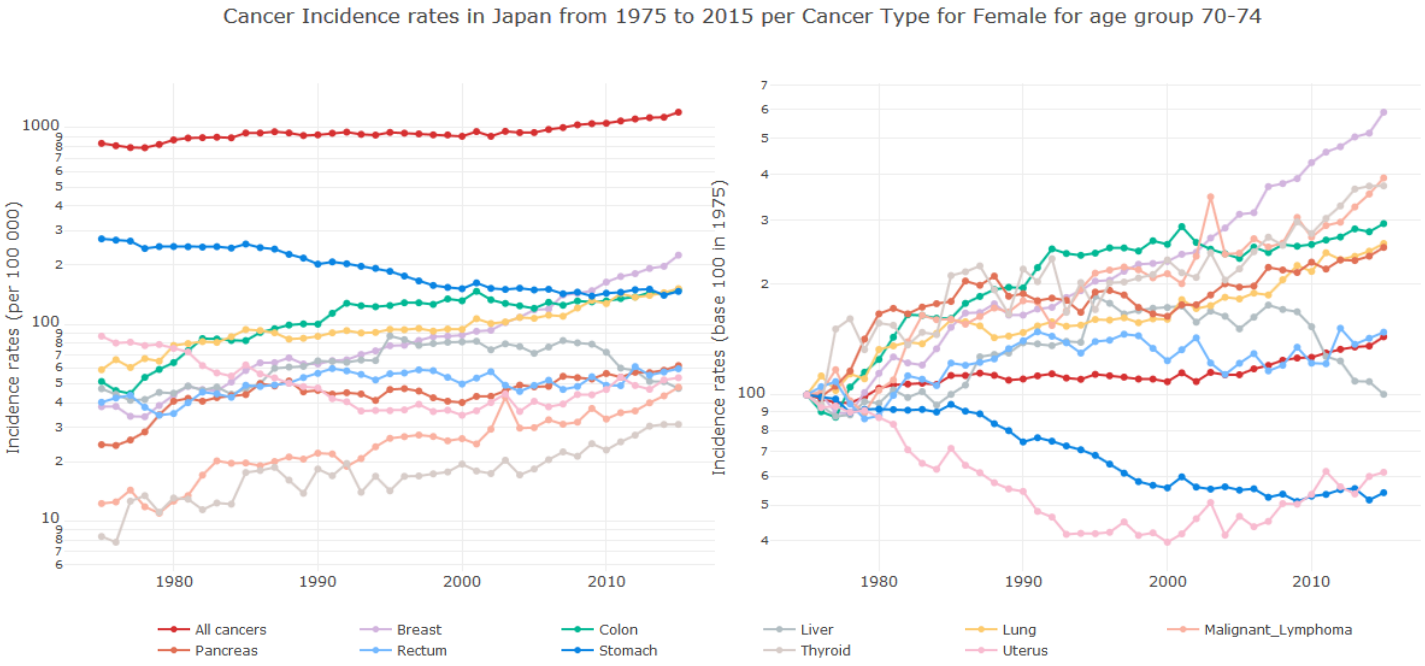


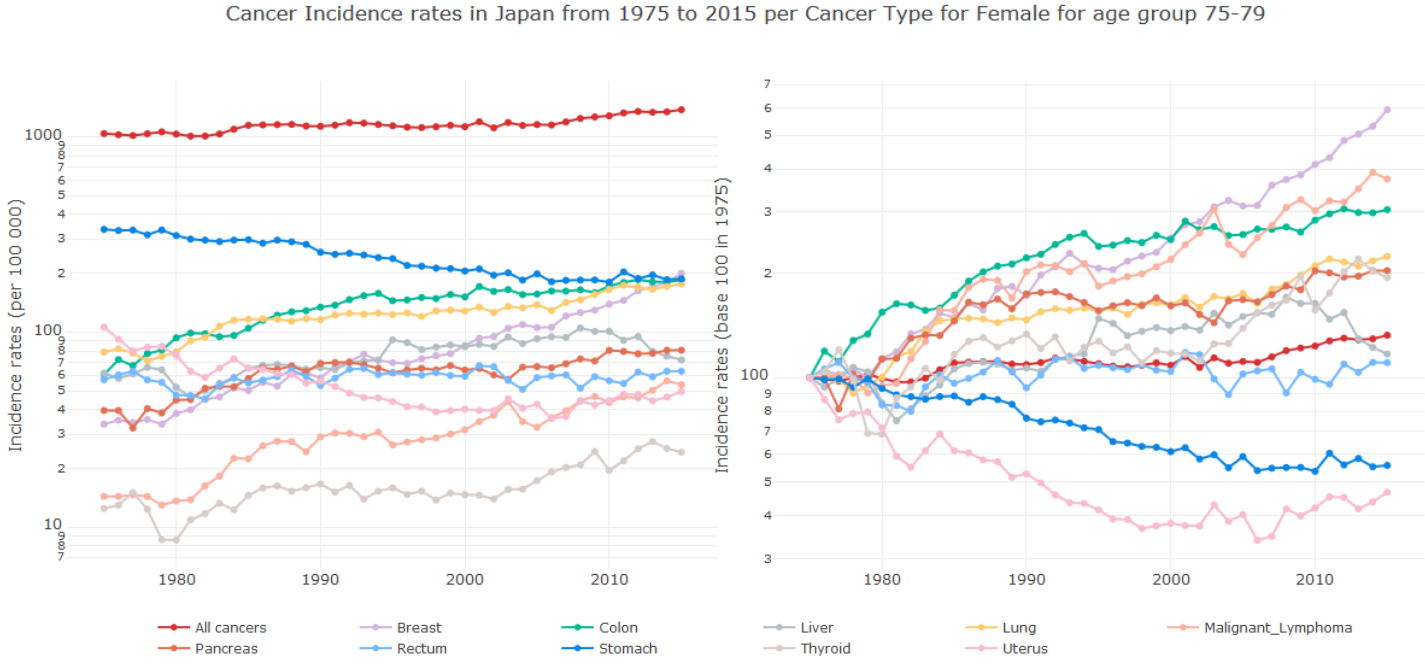
Cancer all causes annual incidence rate reaches respectively 5.5 and 6.6 per 1,000 for female aged 50 to 54 and 55 to 59. The evolutions are similar to the previous age classes. One cancer out of 3 is breast cancer. Breast cancer incidence has been multiplied by 5 in 40 years.



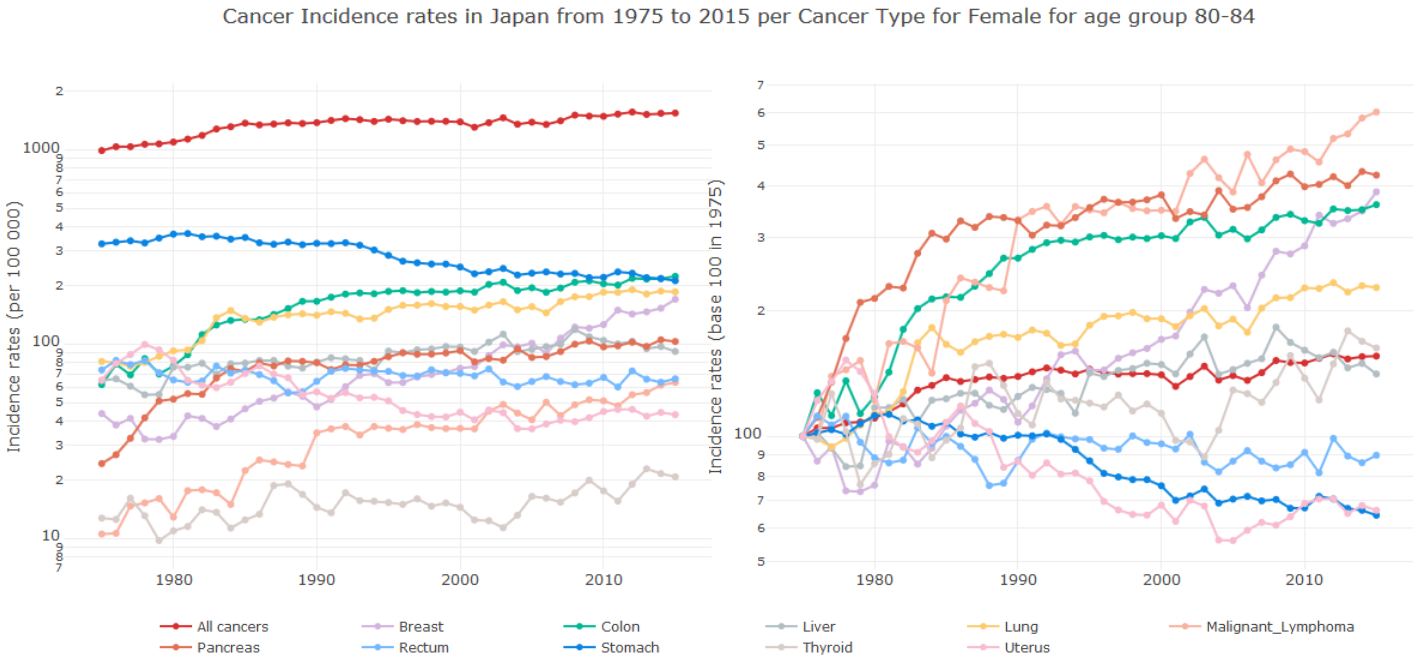


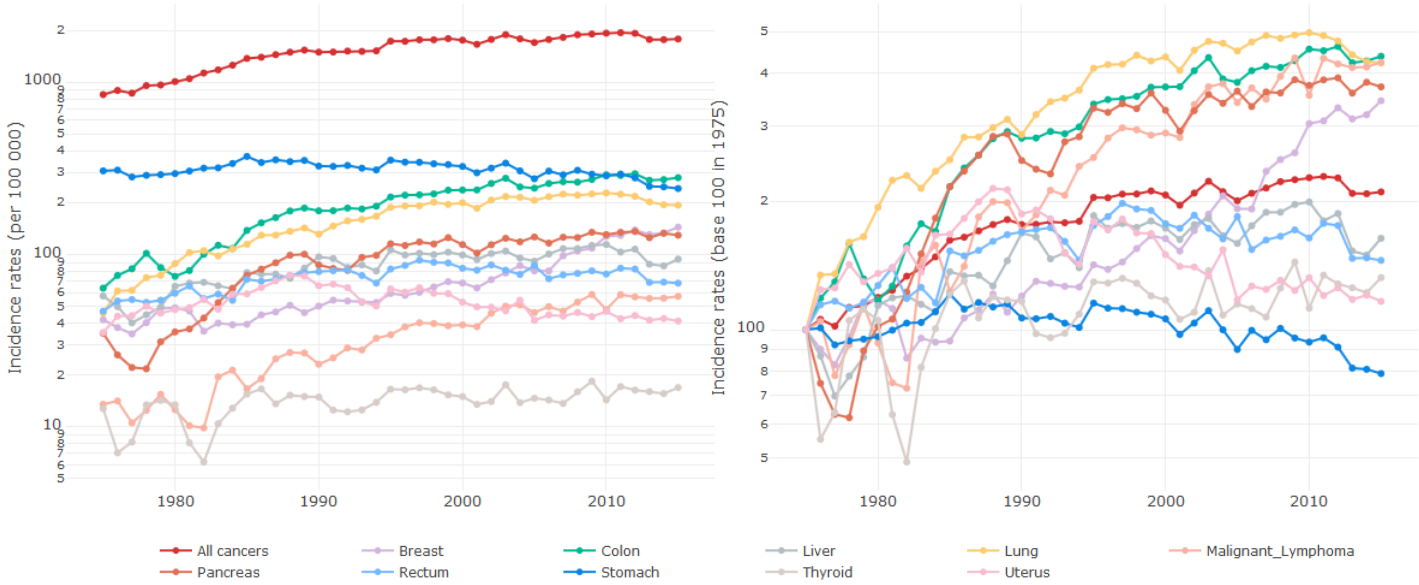
Cancer all causes annual incidence rate reaches respectively 8 and 9.7 per 1,000 for female aged 60 to 64 and 65 to 69. The cancer all causes incidence has increased from 2000. All cancer sites, except stomach and uterus, has increased over the study period.





Cancer all causes annual incidence rate reaches respectively 1.2 and 1.4% for female aged 70 to 74 and 75 to 79. The cancer all causes incidence has increased from 2005. The most frequent cancer sites are breast, stomach, colon and lung, with similar incidences in 2015. Again, stomach cancer incidence has strongly decreases while for most cancer sites incidence has increased over the study period.





Cancer all causes annual incidence rate reaches respectively 1.8% for female aged 80 to 84 and more than 85. The cancer all causes incidence has increased from 1975 to 1995 and is then relatively constant. The most frequent cancer sites are stomach and colon, and then lung, breast and pancreas.

E Appendix E: Statistical Tests & Additional Estimation Results

This section details statistical tests used to assess the validity of model estimation assumptions. Additional estimations results are also presented.

E.1 Statistical Tests

E.1.1 Jarque-Bera Test

The Jarque-Bera test tests whether a distribution has properties of a normal distribution, that is if the empirical skewness S and kurtosis K match a normal distribution. The null hypothesis H_0 is that the distribution match the normal distribution, whether it does not match in the alternative hypothesis H_1. The test statistic JB is as follows:

$$JB = \frac{n}{6} \cdot (S^2 + \frac{1}{4}(K-3)^2) \tag{20}$$

E.1.2 Box-Pierce Test

The Box-Pierce test tests whether a distribution is auto-correlated, that is there is a dependence between value of the distribution. The null hypothesis H_0 is that the distribution does not exhibit auto-correlation up to lag m, whether it does in the alternative hypothesis H_1. The test statistic Q is as follows:

$$Q = n \cdot \sum_{k=1}^{m} \hat{r_k^2} \tag{21}$$

with $\hat{r_k^2}$ the estimated auto-correlation of the distribution at lag k

E.1.3 Ljung-Box Test

The Ljung-Box test is a similar test to Box-Pierce test, testing whether a distribution is auto-correlated, that is there is a dependence between value of the distribution. The null hypothesis H_0 is that the distribution does not exhibit auto-correlation up to lag h, whether it does in the alternative hypothesis H_1. The test statistic Q is as follows:

$$Q = n(n+2) \cdot \sum_{k=1}^{h} \frac{\hat{r_k^2}}{n-k} \tag{22}$$

with $\hat{r_k^2}$ the estimated autocorrelation of the distribution at lag k

E.2 Estimations Results

This section adds some details of the section model estimations. The first sub-section details additional results related to the data of Japan, and the second of Korea.

E.2.1 Additional Results - Japan

Age Group	P-Value	Age Group	P-Value	Age Group	P-Value
Age 20-24	0.07	Age 45-49	0.88	Age 70-74	0.42
Age 25-29	0.24	Age 50-54	0.65	Age 75-79	0.65
Age 30-34	0.02	Age 55-29	0.09	Age 80-84	0.87
Age 35-39	0.37	Age 60-64	0.23	Age 85	0.83
Age 40-44	0.97	Age 65-69	0.53		

Table 24: P-Value of Ljung-Box tests for cancer all sites in Japan in Sparse VAR estimation for Male

The hypothesis H_0 is accepted at the significance level of 5% for the majority of age groups. It is rejected only for the age group 30-34. This procedure does univariate tests on residuals for each age group, and therefore does not perform tests at multivariate level. Despite the uncorrelation at multivariate level cannot be accepted with such univariate tests, it gives some possibilities of such results.

Age Group	P-Value	Age Group	P-Value	Age Group	P-Value
Age 20-24	0.29	Age 45-49	0.27	Age 70-74	0.47
Age 25-29	0.04	Age 50-54	0.56	Age 75-79	0.51
Age 30-34	0.22	Age 55-29	0.68	Age 80-84	0.85
Age 35-39	0.32	Age 60-64	0.78	Age 85	0.77
Age 40-44	0.08	Age 65-69	0.67		

Table 25: P-Value of Ljung-Box tests for cancer all sites in Japan in Sparse VAR estimation for Female

The hypothesis H_0 is accepted at the significance level of 5% for the majority of age groups. It is rejected only for the age group 25-29.

Lee-Carter

The following figures represent the values of residuals depending on age, calendar year and birth date for cancer all causes estimation:

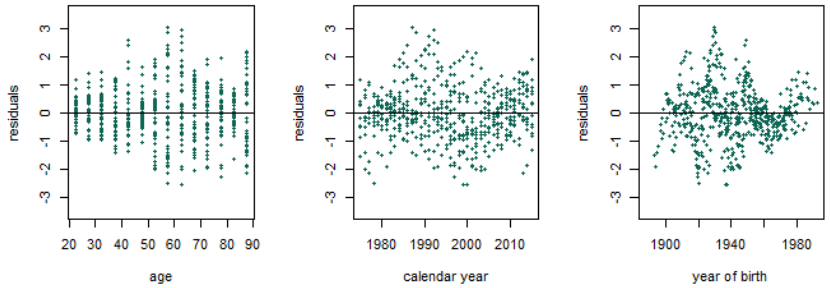


Figure 59: Lee-Carter estimation residuals for Cancer all sites for Male

The pattern of residuals following insured age shows some variance differences between young ages (with small variance) and middle & old ages (with greater variance). The pattern following calendar year reveals some irregularities, with some high residuals around 1990. The pattern following the year of birth presents an irregular form, with higher residuals for 1920 to 1950 years of births. This would indicate that a cohort effect could be tested in further studies.

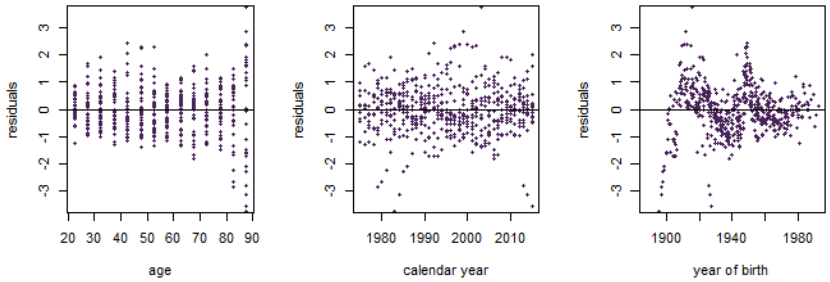


Figure 60: Lee-Carter estimation residuals for Cancer all sites for Female

For female, the variance on residuals at age 85+ is very high. The pattern of residuals depending calendar year is quite regular, whereas the pattern depending year of birth shows some important regularities.

E.2.2 Estimation Results - Japan

This section details the results of the nested cross-validation procedure per site and gender with the data from Japan. The model with the minimal MSE is highlighted in green.

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Cancer All causes	Naïve	6,760	5.5%	Naïve	4,523	8.5%
C00-C96	VAR	8,930	7.6%	VAR	2,598	6.4%
	Sparse VAR	3,985	5.0%	Sparse VAR	2,730	6.1%
	NN I	6,239	5.7%	NN I	5,279	8.8%
	NN I∗	10,017	6.2%	NN I∗	4,751	8.7%
	NN II	9,282	5.6%	NN II	4,658	8.0%
	Lee-Carter	4,331	5.5%	Lee-Carter	4,290	7.8%
Stomach	Naïve	177	13.5%	Naïve	25	10.2%
C16	VAR	297	12.6%	VAR	118	14.6%
	Sparse VAR	240	11.1%	Sparse VAR	107	11.7%
	NN I	195	15.9%	NN I	118	15.3%
	NN I∗	202	15.4%	NN I∗	54	11.0%
	NN II	185	15.0%	NN II	81	13.1%
	Lee-Carter	279	12.9%	Lee-Carter	74	12.1%
Colon	Naïve	54	9.0%	Naïve	43	9.7%
C18	VAR	203	12.7%	VAR	36	10.2%
	Sparse VAR	82	9.5%	Sparse VAR	19	8.0%
	NN I	207	12.1%	NN I	82	10.9%
	NN I∗	87	11.6%	NN I∗	66	12.4%
	NN II	75	10.2%	NN II	38	10.0%
	Lee-Carter	471	15.9%	Lee-Carter	41	8.7%
Rectum	Naïve	42	9.9%	Naïve	15	14.8%
C19-C20	VAR	51	10.7%	VAR	19	16.2%
	Sparse VAR	25	8.5%	Sparse VAR	12	14.2%
	NN I	47	10.9%	NN I	21	17.0%
	NN I∗	82	12.3%	NN I∗	22	17.0%
	NN II	55	9.9%	NN II	16	14.1%
	Lee-Carter	45	9.4%	Lee-Carter	8	11.9%
Pancreas	Naïve	24	16.4%	Naïve	13	18.7%
C25	VAR	31	20.5%	VAR	11	18.4%
	Sparse VAR	17	17.0%	Sparse VAR	8	18.5%
	NN I	40	18.2%	NN I	32	22.1%
	NN I∗	61	20.8%	NN I∗	24	21.9%
	NN II	40	15.5%	NN II	19	17.5%
	Lee-Carter	19	15.0%	Lee-Carter	25	20.0%
Liver	Naïve	314	28.1%	Naïve	62	39.9%
C22	VAR	784	36.2%	VAR	112	48.4%
	Sparse VAR	529	27.1%	Sparse VAR	97	42.3%
	NN I	784	39.5%	NN I	76	46.3%
	NN I∗	572	30.6%	NN I∗	107	50.9%
	NN II	478	27.1%	NN II	92	41.0%
	Lee-Carter	453	58.1%	Lee-Carter	189	52.1%

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Esophagus	Naïve	29	14.9%	Naïve	1.9	17.4%
C15	VAR	35	18.1%	VAR	4.3	24.3%
	Sparse VAR	26	16.3%	Sparse VAR	2.9	18.6%
	NN I	54	16.3%	NN I	2.5	21.1%
	NN I∗	57	20.1%	NN I∗	2.9	23.7%
	NN II	75	16.4%	NN II	5.7	27.9%
	Lee-Carter	38	17.8%	Lee-Carter	3.7	22.4%
Lung	Naïve	371	9.5%	Naïve	118	13.3%
C33-C34	VAR	1,098	15.3%	VAR	244	20.4%
	Sparse VAR	640	10.5%	Sparse VAR	62	12.4%
	NN I	438	9.4%	NN I	131	14.8%
	NN I∗	501	10.5%	NN I∗	102	16.1%
	NN II	466	9.7%	NN II	95	14.0%
	Lee-Carter	1,571	11.6%	Lee-Carter	89	12.7%
Prostate	Naïve	2,254	21.1%	Naïve	-	-
C61	VAR	2,630	22.3%	VAR	-	-
	Sparse VAR	1,166	18.1%	Sparse VAR	-	-
	NN I	8,392	26.0%	NN I	-	-
	NN I∗	2,195	21.3%	NN I∗	-	-
	NN II	4,378	23.4%	NN II	-	-
	Lee-Carter	1,121	18.2%	Lee-Carter	-	-
Breast	Naïve	-	-	Naïve	1,011	13.8%
C50	VAR	-	-	VAR	635	12.3%
	Sparse VAR	-	-	Sparse VAR	430	9.0%
	NN I	-	-	NN I	586	11.2%
	NN I∗	-	-	NN I∗	342	9.1%
	NN II	-	-	NN II	319	8.5%
	Lee-Carter	-	-	Lee-Carter	208	6.5%
Uterus	Naïve	-	-	Naïve	82	12.8%
C53-C55	VAR	-	-	VAR	125	16.2%
	Sparse VAR	-	-	Sparse VAR	114	14.9%
	NN I	-	-	NN I	119	15.7%
	NN I∗	-	-	NN I∗	116	15.7%
	NN II	-	-	NN II	88	13.6%
	Lee-Carter	-	-	Lee-Carter	273	24.2%
Ovary	Naïve	-	-	Naïve	6.5	11.2%
C56	VAR	-	-	VAR	7.1	11.5%
	Sparse VAR	-	-	Sparse VAR	3.6	9.2%
	NN I	-	-	NN I	7.5	13.6%
	NN I∗	-	-	NN I∗	10.1	14.3%
	NN II	-	-	NN II	14.7	13.6%
	Lee-Carter	-	-	Lee-Carter	3.8	8.5%

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Skin	Naïve	31	20.2%	Naïve	12	23.3%
C43-C44	VAR	33	22.8%	VAR	10	23.5%
	Sparse VAR	27	18.5%	Sparse VAR	10	21.3%
	NN I	37	26.1%	NN I	26	31.2%
	NN I∗	45	27.5%	NN I∗	14	26.6%
	NN II	43	25.0%	NN II	15	24.3%
	Lee-Carter	19	16.8%	Lee-Carter	6	18.6%
Thyroïd	Naïve	5	21.5%	Naïve	15	13.8%
C73	VAR	14	30.2%	VAR	16	13.8%
	Sparse VAR	4	19.1%	Sparse VAR	8	10.4%
	NN I	7	25.4%	NN I	24	18.1%
	NN I∗	7	24.1%	NN I∗	21	18.1%
	NN II	6	21.7%	NN II	18	15.1%
	Lee-Carter	3	16.2%	Lee-Carter	13	12.9%
Brain	Naïve	1.4	17.8%	Naïve	1.1	21.4%
C70-C72	VAR	4.2	26.3%	VAR	7.2	37.6%
	Sparse VAR	2.0	19.6%	Sparse VAR	1.6	24.8%
	NN I	1.2	18.1%	NN I	1.8	25.7%
	NN I∗	1.7	22.7%	NN I∗	1.7	22.4%
	NN II	1.6	19.3%	NN II	0.9	17.9%
	Lee-Carter	4.0	20.4%	Lee-Carter	19.5	35.4%
Bladder	Naïve	13	23.1%	Naïve	1.1	31.9%
C67	VAR	43	38.6%	VAR	2.4	42.9%
	Sparse VAR	28	26.7%	Sparse VAR	1.5	34.5%
	NN I	39	30.9%	NN I	1.8	34.8%
	NN I∗	21	37.8%	NN I∗	3.1	53.8%
	NN II	15	28.9%	NN II	1.2	30.9%
	Lee-Carter	65	46.3%	Lee-Carter	0.9	24.1%
Gallbladder	Naïve	8	17.8%	Naïve	5	22.4%
C23-C24	VAR	15	24.0%	VAR	59	34.9%
	Sparse VAR	14	19.0%	Sparse VAR	6	22.9%
	NN I	14	24.1%	NN I	7	27.4%
	NN I∗	18	22.0%	NN I∗	10	29.7%
	NN II	26	21.5%	NN II	20	49.3%
	Lee-Carter	20	23.6%	Lee-Carter	4	17.6%
Kidney	Naïve	60	17.1%	Naïve	8	20.4%
C64-C66 C68	VAR	39	16.3%	VAR	6	19.9%
	Sparse VAR	31	14.1%	Sparse VAR	4	18.0%
	NN I	36	18.2%	NN I	6	20.8%
	NN I∗	70	18.9%	NN I∗	7	21.9%
	NN II	48	16.8%	NN II	9	19.7%
	Lee-Carter	29	12.9%	Lee-Carter	3	14.8%

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Leukemia	Naïve	3.2	13.4%	Naïve	1.4	14.9%
C91-C95	VAR	6.3	16.7%	VAR	2.6	18.8%
	Sparse VAR	4.4	14.0%	Sparse VAR	1.5	15.3%
	NN I	4.4	15.2%	NN I	2.0	17.1%
	NN I∗	10.3	18.8%	NN I∗	2.8	20.7%
	NN II	11.4	15.3%	NN II	5.7	20.3%
	Lee-Carter	5.6	16.2%	Lee-Carter	1.3	16.0%
Malignant Lymphoma	Naïve	35	14.7%	Naïve	24	16.2%
C81-C85 C96	VAR	43	20.4%	VAR	19	16.7%
	Sparse VAR	17	13.4%	Sparse VAR	13	13.5%
	NN I	65	22.5%	NN I	18	16.2%
	NN I∗	53	23.1%	NN I∗	24	19.6%
	NN II	41	19.8%	NN II	17	15.0%
	Lee-Carter	22	15.1%	Lee-Carter	11	13.0%

E.2.3 Estimation Results - Korea

This section details the results of the nested cross-validation procedure per site and gender with the data from Korea. The model with the minimal MSE is highlighted in green.

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Cancer All causes	Naïve	2,942	5.7%	Naïve	1,936	6.1%
C00-C96	Sparse VAR	4,452	6.3%	Sparse VAR	5,418	10.3%
(direct)	NN I	4,101	7.7%	NN I	6,322	12.4%
	Lee-Carter	11,554	12.3%	Lee-Carter	7,786	11.9%
Cancer All causes	Naïve	2,942	5.7%	Naïve	1,936	6.1%
(indirect)	Sparse VAR	3,291	12.6%	Sparse VAR	11,831	22.9%
C00-C96	NN I	4,040	7.7%	NN I	6,325	12.3%
	Lee-Carter	3,362	12.9%	Lee-Carter	11,617	22.8%
Stomach	Naïve	53	7.3%	Naïve	461	10.0%
C16	Sparse VAR	35	11.1%	Sparse VAR	282	6.4%
	NN I	31	15.9%	NN I	476	7.7%
	Lee-Carter	34	8.0%	Lee-Carter	274	7.6%
Colonrectum	Naïve	295	9.0%	Naïve	60	8.9%
C18-C20	Sparse VAR	399	11.2%	Sparse VAR	124	10.6%
	NN I	578	12.7%	NN I	60	10.0%
	Lee-Carter	16	14.1%	Lee-Carter	216	14.7%
Pancreas	Naïve	15	10.9%	Naïve	10	14.3%
C25	Sparse VAR	11	11.8%	Sparse VAR	6	13.8%
	NN I	9	10.3%	NN I	11	16.8%
	Lee-Carter	11	10.7%	Lee-Carter	7	13.6%
Liver	Naïve	99	12.0%	Naïve	17	12.9%
C22	Sparse VAR	33	6.9%	Sparse VAR	9	11.5%
	NN I	44	9.6%	NN I	24	14.2%
	Lee-Carter	42	6.5%	Lee-Carter	12	11.6%
Esophagus	Naïve	15	22.0%	Naïve	1.9	27.3%
C15	Sparse VAR	10	23.6%	Sparse VAR	2.9	29.8%
	NN I	12	20.6%	NN I	2.5	29.9%
	Lee-Carter	8	14.8%	Lee-Carter	NA	35.8%
Lung	Naïve	105	7.0%	Naïve	17	12.2%
C33-C34	Sparse VAR	105	7.5%	Sparse VAR	18	13.2%
	NN I	1398	8.6%	NN I	53	12.0%
	Lee-Carter	100	5.7%	Lee-Carter	19	11.2%
Prostate	Naïve	262	9.3%	Naïve	-	-
C61	Sparse VAR	178	10.0%	Sparse VAR	-	-
	NN I	674	11.7%	NN I	-	-
	Lee-Carter	624	14.4%	Lee-Carter	-	-
Testis	Naïve	0.1	44.4%	Naïve	-	-
C62	Sparse VAR	0.1	48.0%	Sparse VAR	-	-
	NN I	0.2	53.4%	NN I	-	-
	Lee-Carter	0.1	52.0%	Lee-Carter	-	-
Breast	Naïve	-	-	Naïve	197	9.3%
C50	Sparse VAR	-	-	Sparse VAR	45	5.0%
	NN I	-	-	NN I	51	5.0%
	Lee-Carter	-	-	Lee-Carter	32	4.2%

Cancer Site	Male	MSE	MAPE	Female	MSE	MAPE
Cervix	Naïve	-	-	Naïve	4.8	9.1%
C53	Sparse VAR	-	-	Sparse VAR	4.3	8.2%
	NN I	-	-	NN I	6.4	9.9%
	Lee-Carter	-	-	Lee-Carter	5.9	10.4%
Corpus	Naïve	-	-	Naïve	6.4	13.0%
C54	Sparse VAR	-	-	Sparse VAR	3.2	9.5%
	NN I	-	-	NN I	5.1	10.9%
	Lee-Carter	-	-	Lee-Carter	2.4	8.8%
Ovary	Naïve	-	-	Naïve	3.2	10.7%
C56	Sparse VAR	-	-	Sparse VAR	2.2	9.3%
	NN I	-	-	NN I	5.3	14.4%
	Lee-Carter	-	-	Lee-Carter	1.9	8.4%
Thyroid	Naïve	87	21.9%	Naïve	1,900	26.1%
C73	Sparse VAR	203	33.4%	Sparse VAR	8,901	58.8%
	NN I	191	32.8%	NN I	24,995	45.9%
	Lee-Carter	765	63.5%	Lee-Carter	11,878	66.0%
Brain	Naïve	1.5	14.6%	Naïve	1.3	19.8%
C70-C72	Sparse VAR	1.8	15.8%	Sparse VAR	1.3	20.2%
	NN I	1.8	14.5%	NN I	1.1	18.0%
	Lee-Carter	2.4	13.8%	Lee-Carter	0.8	16.0%
Bladder	Naïve	5.5	13.1%	Naïve	1.1	20.2%
C67	Sparse VAR	5.7	12.9%	Sparse VAR	1.1	23.0%
	NN I	5.4	11.5%	NN I	1.5	24.2%
	Lee-Carter	3.8	8.9%	Lee-Carter	1.4	18.6%
Kidney	Naïve	5.5	9.6%	Naïve	2.6	12.3%
C64	Sparse VAR	8.8	9.1%	Sparse VAR	2.4	12.0%
	NN I	11.5	11.2%	NN I	3.0	13.7%
	Lee-Carter	12.2	9.8%	Lee-Carter	2.3	10.7%
Leukemia	Naïve	2.8	13.2%	Naïve	1.7	11.8%
C91-C95	Sparse VAR	3.7	15.1%	Sparse VAR	1.9	12.9%
	NN I	2.7	15.3%	NN I	2.7	11.9%
	Lee-Carter	1.8	11.9%	Lee-Carter	1.0	8.2%
Malignant Lymphoma	Naïve	35	14.7%	Naïve	24	16.2%
C81-C85 C96	Sparse VAR	17	13.4%	Sparse VAR	13	13.5%
	NN I	65	22.5%	NN I	18	16.2%
	Lee-Carter	22	15.1%	Lee-Carter	11	13.0%
Gallbladder	Naïve	8	17.8%	Naïve	5	22.4%
C23-C24	Sparse VAR	14	19.0%	Sparse VAR	6	22.9%
	NN I	14	24.1%	NN I	7	27.4%
	Lee-Carter	20	23.6%	Lee-Carter	4	17.6%

E.3 Predictions Results

The following table collects the evolution of the different cancer site incidence rates to give more details about the results.

The forecasts of esophagus, pancreas and skin cancer are discussed, while liver cancer, for which Naïve model is selected, is omitted. The combination of esophagus, pancreas and liver cancer, which terminates the policy in the numerical illustration, is also discussed.

E.3.1 Esophagus Cancer

Selected models are Sparse VAR for male and Naïve for female. Results are as follows:

Age	Male (Sparse VAR model)						Female (Naïve model)					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
30-34	-16.3%	-32.7%	-65.4%	-98%	-100%	-100%	0%	0%	0%	0%	0%	0%
35-39	-7.1%	-14.1%	-28.2%	-42.3%	-56.4%	-70.6%	0%	0%	0%	0%	0%	0%
40-44	7.5%	15.1%	30.2%	45.3%	60.4%	75.5%	0%	0%	0%	0%	0%	0%
45-49	2.6%	5.2%	10.3%	15.5%	20.6%	25.8%	0%	0%	0%	0%	0%	0%
50-54	4.6%	9.4%	18.9%	28.4%	38.0%	47.5%	0%	0%	0%	0%	0%	0%
55-59	5.4%	10.9%	21.8%	32.7%	43.6%	54.5%	0%	0%	0%	0%	0%	0%
60-64	6.8%	11.8%	21.8%	31.8%	41.8%	51.8%	0%	0%	0%	0%	0%	0%
65-69	4.8%	9.5%	18.9%	28.3%	37.7%	47.1%	0%	0%	0%	0%	0%	0%
70-74	6.9%	12.3%	23.1%	33.9%	44.7%	55.5%	0%	0%	0%	0%	0%	0%
75-79	3.6%	6.7%	12.9%	19.0%	25.1%	31.3%	0%	0%	0%	0%	0%	0%
80-84	0.9%	2.4%	5.4%	8.5%	11.5%	14.5%	0%	0%	0%	0%	0%	0%

Table 26: Evolution between last historical value & forecasts for Esophagus Cancer

Trends strongly differ depending male age, with an inflexion age at age 40, while no evolution is forecasted for female due to the Naïve model.

Ages 30 to 40 show a strong decrease. Forecasted rates are actually negative from time horizon 40 years and have been floored to 0. Because there is no constraints on negative forecasts in VAR model, it is possible to get negative incidence rates in the forecasts. Such evidences indicates model misspecification. However, it should be noted that incidence rates for esophagus cancer between ages 30 and 40 are very low and close to 0. Incidence rates in 2015 for age class 30-34 and 35-39 are respectively 0.2 and 0.5 per 100,000 (that is nationally between 6 and 21 cases annually). Then a forecasted null incidence rates for such ages remain a plausible forecast.

Middle ages show an increase of incidence rates around 50% at 50 years, except age 40-44 for which the increase reach 75.5% and 45-49 for which the increase is 25.8%. Increase is smaller at old ages, at the age esophagus cancer incidence is maximum.

E.3.2 Pancreas Cancer

Selected models are Sparse VAR for both male and female. Results are as follows:

Age	Female (Sparse VAR model)						Female (Sparse VAR model)					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
30-34	-4.9%	-9.9%	-19.8%	-29.7%	-39.6%	-49.5%	5.4%	10.8%	21.5%	32.3%	43%	53.8%
35-39	-8.7%	-17.3%	-34.6%	-51.9%	-69.3%	-86.6%	-1.6%	-3.3%	-6.6%	-9.8%	-13.1%	-16.4%
40-44	-0.9%	-1.9%	-3.7%	-5.6%	-7.5%	-9.4%	3.8%	7.7%	15.4%	23.1%	30.7%	38.4%
45-49	-2.6%	-2.5%	-2.4%	-2.2%	-2.1%	-1.9%	1.0%	3.8%	9.3%	14.9%	20.5%	26.1%
50-54	6.0%	12.0%	24.1%	36.1%	48.1%	60.1%	4.2%	8.4%	16.7%	25%	33.4%	41.7%
55-59	4.1%	8.0%	15.8%	23.6%	31.4%	39.3%	5.9%	11.3%	22.1%	32.8%	43.6%	54.3%
60-64	6.2%	12.2%	24.3%	36.5%	48.6%	60.7%	6.3%	12%	23.5%	35%	46.5%	58.0%
65-69	3.9%	7.7%	15.5%	23.3%	31.1%	38.8%	5.7%	10.2%	19.2%	28.2%	37.2%	46.2%
70-74	5.2%	11.4%	23.7%	36.1%	48.4%	60.7%	7.8%	15.6%	31.1%	46.7%	62.2%	77.8%
75-79	5.7%	12.2%	25.3%	38.4%	51.5%	64.5%	6.3%	12.9%	26.1%	39.3%	52.5%	65.7%
80-84	7.9%	16.4%	33.3%	50.2%	67.1%	84.0%	9.5%	19.1%	38.3%	57.5%	76.7%	95.9%

Table 27: Evolution between last historical value & forecasts for Pancreas Cancer

For male, incidence rates decrease for all ages below 49 and then increase. As esophagus cancer, pancreas cancer is a rare cancer at young ages. Incidence rates in 2015 for age class 30-34 and 35-39 are respectively 0.5 and 0.9 per 100,000 (that is nationally 20 and 36 cases annually). Forecasting such rare events would require specific modelling. For age class 35-39, the forecasted incidence reaches 0.2 per 100,000 at a 50 years horizon. An alternative is also to consider constant such rare events. Forecast is almost constant for age class 45-49, while forecasts for older ages show an incidence increase. The increase is maximal for age class 80-84.

For female, incidence rates increase for all age groups except age 35-39 for which there is a small decrease of -16.4% at 50 years. Increase is generally higher at old ages. While there are strong differences between the trends of male and female at young ages, the trends at middle and old ages are quite similar.

E.3.3 Special Cancer

This section summarizes the forecast of esophagus, pancreas and liver cancer (named special cancer), which is the three cancer sites used in the product design of the numerical illustration. Such cancer diagnosis allows a higher benefit level and terminate the policy. Recall that selected models for liver cancer are Naïve models for both male and female, omitted below.

Age	Male						Female					
Male	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
30-34	-3.9%	-7.8%	-15.5%	-23.3%	-28.3%	-33.1%	3.2%	6.3%	12.6%	19.0%	25.3%	31.6%
35-39	-4.3%	-8.5%	-17.1%	-25.6%	-34.2%	-42.7%	-0.7%	-1.4%	-2.7%	-4.1%	-5.4%	-6.8%
40-44	1.9%	3.7%	7.5%	11.2%	14.9%	18.7%	1.9%	3.8%	7.6%	11.5%	15.3%	19.1%
45-49	-0.2%	0.5%	1.8%	3.2%	4.6%	5.9%	0.6%	2.1%	5.3%	8.5%	11.7%	14.9%
50-54	3.3%	6.7%	13.5%	20.2%	27.0%	33.8%	2.1%	4.2%	8.3%	12.4%	16.6%	20.7%
55-59	2.9%	5.8%	11.6%	17.4%	23.2%	29.0%	3.3%	6.4%	12.4%	18.4%	24.5%	30.5%
60-64	4%	7.3%	14.0%	20.8%	27.5%	34.2%	3.4%	6.4%	12.6%	18.8%	24.9%	31.1%
65-69	2.8%	5.5%	11.0%	16.4%	21.9%	27.3%	3.0%	5.3%	10.1%	14.8%	19.5%	24.2%
70-74	3.8%	7.4%	14.5%	21.6%	28.8%	35.9%	3.9%	7.8%	15.5%	23.3%	31.1%	38.9%
75-79	2.7%	5.4%	10.9%	16.4%	21.9%	27.4%	3.0%	6.2%	12.5%	18.8%	25.2%	31.5%
80-84	2.6%	5.5%	11.3%	17.1%	22.9%	28.7%	4.7%	9.4%	18.9%	28.3%	37.8%	47.3%

Table 28: Evolution between last historical value & forecasts for special cancer

The incidences of the three cancer sites are quite similar. For male, at age 40-44 the incidence rates in 2015 are respectively 2.3, 2.9 and 2.6 per 100,000 for esophagus, liver and pancreas cancer, and reach 117.4, 131.7 and 97.4 at age 70-74. For female, esophagus cancer is less prevalent at old ages and at age 40-44 the incidence rates in 2015 are respectively 1.2, 1.0 and 2.2 per 100,000 for esophagus, liver and pancreas cancer, and reach 14.6, 47.2 and 61.7 at age 70-74.

Then, the general evolution reported in the table above is for male close to a simple average of the evolutions of the three cancer sites incidence rates. Some decrease is forecasted at young ages, whereas it is increase at middle and old ages. This increase is limited because no evolution is forecasted for liver cancer (both male and female) and esophagus cancer (for female only).

E.3.4 Skin Cancer

Selected models are Lee-Carter model for both male and female. Results are as follows:

Age	Male (Lee-Carter model)						Female (Lee-Carter model)					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
30-34	32.3%	65.2%	158%	302%	526%	877%	68.6%	115%	251%	472%	833%	1,421%
35-39	22.8%	49.7%	122%	230%	391%	629%	66.1%	104%	207%	362%	595%	946%
40-44	39.4%	61.6%	117%	192%	292%	427%	19.2%	42.3%	103%	189%	311%	486%
45-49	3.4%	21.8%	69.0%	134%	225%	351%	-8.7%	8.9%	55.3%	121%	215%	349%
50-54	37.0%	61.4%	124%	211%	333%	501%	9.9%	32.3%	91.7%	178%	303%	483%
55-59	20.3%	36.9%	77.4%	130%	198%	286%	8.3%	30.7%	90.1%	177%	302%	485%
60-64	1.3%	12.4%	38.5%	70.6%	110%	159%	9.0%	27%	72.4%	134%	218%	331%
65-69	12.1%	27.3%	64.1%	112%	173%	252%	20.7%	43.7%	104%	188%	308%	478%
70-74	22.1%	40.7%	87.0%	145%	223%	338%	23.2%	45.9%	105%	187%	303%	466%
75-79	15.5%	35.2%	85.2%	154%	248%	376%	16.1%	35.9%	86.1%	155%	249%	378%
80-84	21.5%	44.8%	105%	191%	313%	485%	18.8%	36.8%	81.4%	140%	219%	322%

Table 29: Evolution between last historical value & forecasts for skin cancer

The incidence rates in 2015 for male and female are respectively 2.6 and 3.1 per 100,000 at age 40-44, and 38.9 and 24.7 at age 70-74. Such incidence rates have strongly increased in the recent years. For example, the incidence rates in 2000 for male and female are respectively 1.4 and 1.3 per 100,000 at age 40-44, and 19.7 and 12.1 at age 70-74. For the age class discussed, the incidence rates have then doubled in 15 years. As a consequence, the forecasts of skin cancer diagnosis show an extreme increase in the future. the incidence rates in 2015 for male and female are respectively 14.2 and 18.1 per 100,000 at age 40-44, and 175.2 and 144.5 at age 70-74. Such increase may be thought as not plausible and therefore an expert judgment could especially object to such forecast results.

As assessed in the analysis of change, the impact of change from historical incidence tables to generation tables for skin cancer is relatively limited. Still, depending the weight of skin cancer cover in a product design, a review of the skin cancer modelling can be performed.

Skin cancer results stress the importance of reviewing results to prevent the impact of extreme forecasts. While Lee-Carter model have been selected throughout the cross-validation procedure, this model choice must be confirmed by plausible forecasts values. This is also the case for cancer all causes for female, discussed in the next section.

E.4 Additional Results

This section details the forecasts of VAR and Sparse VAR models for cancer all causes for female. Recall that in the cross-validation procedure, the MSE is minimal for VAR model (2,598 vs. 2,730 for Sparse VAR). However, the MSE is minimal for Sparse VAR (6.1% vs 6.4% for VAR). Sparse VAR and VAR models are then very close in terms of quality of test samples forecasts. The third best model, Lee-Carter, has a MSE of 4,290. Results of VAR and Sparse VAR forecasts are presented below:

Age	Male (VAR model)						Female (Sparse model)					
Horizon	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**	**5Y**	**10Y**	**20Y**	**30Y**	**40Y**	**50Y**
20-24	-9.8%	-6.5%	2.9%	12.6%	22.4%	32.1%	6.2%	12.4%	24.9%	37.3%	49.8%	62.2%
25-29	5.8%	8.3%	17.9%	28.1%	38.2%	48.4%	4.6%	9.3%	18.5%	27.8%	37.0%	46.3%
30-34	-2.8%	1.7%	13.5%	25.4%	37.4%	49.3%	6.8%	13.6%	27.2%	40.7%	54.3%	67.9%
35-39	-0.2%	5.3%	17.3%	29.3%	41.3%	53.3%	6.5%	13.0%	26.1%	39.2%	52.3%	65.4%
40-44	-6.4%	-2.2%	8.4%	19.0%	29.6%	40.2%	6.5%	13.0%	26.1%	39.1%	52.1%	65.1%
45-49	8.5%	15.8%	30.6%	45.4%	60.3%	75.1%	7.0%	14.0%	27.9%	41.9%	55.9%	69.9%
50-54	-4.9%	-0.9%	8.9%	18.8%	28.7%	38.6%	6.1%	12.2%	24.4%	36.7%	48.9%	61.2%
55-59	1.7%	6.8%	17.8%	28.9%	39.9%	50.9%	5.9%	11.8%	23.6%	35.4%	47.2%	59.0%
60-64	0.3%	4.5%	13.8%	23.1%	32.4%	41.6%	5.0%	10.1%	20.3%	30.4%	40.6%	50.8%
65-69	-2.5%	0.5%	7.4%	14.5%	21.5%	28.6%	4.1%	8.4%	16.9%	25.4%	34.0%	42.5%
70-74	0.1%	3.6%	11.3%	19.0%	26.7%	34.5%	3.8%	7.9%	16.1%	24.3%	32.5%	40.7%
75-79	-3.4%	-1.7%	2.9%	7.6%	12.3%	17.1%	2.4%	5.4%	11.6%	17.8%	24.0%	30.1%
80-84	4.7%	9.1%	18.0%	26.9%	35.7%	44.6%	4.5%	8.8%	17.4%	25.9%	34.5%	43.0%
85+	14.0%	21.7%	36.6%	51.5%	66.3%	81.2%	6.0%	12.3%	24.9%	37.5%	50.1%	62.7%

Table 30: Evolution between last historical value & forecasts for cancer all causes

VAR forecasts show a decrease of incidence rates for ages 20-24, 30-34, 35-39, 50-54, 65-69 and 75-79 at short term. While some forecasts evolution are negative at a 10-year horizon, all evolutions are positive from a 20-year time horizon.

F Appendix F: Numerical Illustrations with Japanese Population Profile

This section assess the profitability testing with a sales profile as the Japanese population profile for aged 30 to 70.

F.0.1 Data

The data used is the 2018 population estimates published by the Ministry of Internal Affairs and Communication[62]. This table collects estimated population per age and gender.

The Japanese population profile for aged 30 to 70 is then obtained by scaling the number of sales per age and gender to the population profile. Let l_x^g the population of age x and gender g from the table and s_x^g the number of sales for gender $g \in \{m; f\}$ and age $x \in [\![30; 70]\!]$. Then $s_x^g = 100,000 \times l_x^g / (\sum_{g \in m;f} \sum_{x \in [\![30;70]\!]} l_x^g)$ and the total number of sales is 100,000.

F.0.2 Numerical Results

Indicators for Scenario I: Constant Incidence Rates

Indicators	Value
Number of Claims	11,173 (9,744 General Cancer, 195 Skin Cancer, 1,233 Special Cancer)
Customer Value	78.2%
Pure Technical Loss Ratio	83.3%
EPV Technical Result	13.0 million
EPV Operational Result	5.1 million
EPV Profit (PVFP)	18.2 million

Indicators for Scenario II: Forecasted 2020 Incidence Rates

Indicators	Value
Number of Claims	12,805 (11,029 General Cancer, 363 Skin Cancer, 1,414 Special Cancer)
Customer Value	86.6%
Pure Technical Loss Ratio	95.1%
EPV Technical Result	5.7 million
EPV Operational Result	4.7 million
EPV Profit (PVFP)	10.4 million

[62] http://www.stat.go.jp/english/data/jinsui/index.htm