

**Mémoire présenté le :**  
**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA**  
**et l'admission à l'Institut des Actuaires**

Par : Louis-Marie GOULUT

Titre : Tarification de garanties Maladies graves et redoutées

Confidentialité :  NON  (Durée :  1 an  2 ans)

*Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus*

*Membres présents du jury de Signature  
l'Institut des Actuaires*

.....

.....

.....

*Membres présents du jury de  
l'ISFA*

.....

.....

.....

*Entreprise*

*Nom : Covéa Cooperations*

*Signature :*

*Directeur de mémoire en  
entreprise :*

*Nom : Medhi HIMEUR*

*Signature :*



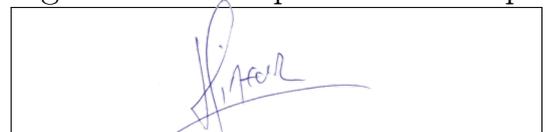
*Invité :*

*Nom :*

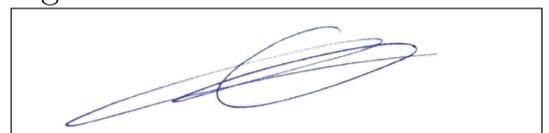
*Signature :*

***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 du candidat





# Table des matières

<b>Table des matières</b>	<b>2</b>
<b>Résumé</b>	<b>4</b>
<b>Note de Synthèse</b>	<b>7</b>
<b>Remerciements</b>	<b>19</b>
<b>Introduction</b>	<b>21</b>
<b>1 Contexte et présentation du produit</b>	<b>23</b>
1.1 Contexte actuel des principales maladies garanties . . . . .	23
1.2 Présentation du produit . . . . .	27
<b>2 Présentation des données</b>	<b>39</b>
2.1 Données en incidence . . . . .	39
2.2 Données de mortalité . . . . .	48
<b>3 Tarification d'une garantie Maladies graves et redoutées</b>	<b>51</b>
3.1 Modèle markovien à plusieurs états . . . . .	51
3.2 Modèle bayésien avec lois bêta binomiale . . . . .	81
<b>Conclusion</b>	<b>95</b>
<b>Bibliographie</b>	<b>97</b>

## Résumé

Avec le vieillissement de la population mondiale et le perfectionnement des technologies médicales, de plus en plus de maladies graves et redoutées sont diagnostiquées chaque jour. Les maladies graves et redoutées peuvent être définies telles que des maladies qui diminuent fondamentalement l'espérance de vie de l'individu et qui entraînent des soins médicaux conséquents ainsi qu'une diminution des ressources physiques de la personne atteinte.

En ce sens, un tel diagnostic va entraîner un coût important que ce soit pour les soins, la perte de revenus et autres encore. En France la perte de revenus ainsi que les frais médicaux sont bien couverts, mais ce n'est pas le cas partout dans le monde. Il existe des pathologies graves dont la gravité et le coût peut se situer entre la couverture pour les frais de soins et pour l'invalidité, il apparaît donc nécessaire de s'en protéger si les protections à notre disposition ne sont pas suffisantes. Pour cela, une garantie dite *Maladie grave et redoutée* ou *Affection redoutée* voit le jour pour garantir un capital en cas de diagnostic d'une telle maladie.

Certaines difficultés apparaissent aussi bien sur le plan technique que sur le plan commercial dans la conception d'un tel produit de prévoyance :

- la tarification nécessite des données et des statistiques qui ne sont pas toujours disponibles ou accessibles,
- les phénomènes d'antisélection que peuvent générer un tel produit,
- la définition précise du risque garanti.

Contrairement aux produits décès en prévoyance, où le décès ne nécessite pas de définition précise, les *Maladies redoutées* doivent être énumérées et définies précisément dans le contrat pour qu'il n'y ait pas de litiges.

Dans la première partie de cette étude, nous nous proposons de tarifier différentes garanties *Maladie graves et redoutées*, dans un premier temps, nous avons d'abord considéré la garantie cancer puis nous avons élargie cette étude à l'AVC et l'infarctus du myocarde afin d'obtenir un tarif pour les trois affections les plus courantes dans le cadre de contrats *Maladies graves et redoutées*. La tarification est basée sur la théorie des processus de Markov continus à plusieurs états.

Dans un second temps, nous nous sommes intéressés à comment intégrer les résultats que l'on obtiendrait après la commercialisation d'un tel produit. Nous avons utilisé la théorie d'inférence statistique bayésienne pour modéliser la sinistralité future.

*Mots-clés : Maladies graves et redoutées, Tarification, Prévoyance, Modèle markovien à plusieurs états, Weibull, Cancer, Accident vasculaire cérébral, Infarctus du myocarde, Inférence bayésienne, Crédibilité.*

### Abstract

As the world's population ages and medical technologies improve, more and more dread diseases are being diagnosed every day. These diseases are also called *Critical Illnesses* can be defined as diseases that fundamentally decrease the life expectancy of the insured and which lead to substantial medical care as well as a reduction of the physical resources of the insured.

In this sense, such a diagnosis will lead to a significant cost in terms of care, loss of income and other costs.

In France, loss of income and medical expenses are well covered, but this is not the case worldwide. There are serious pathologies whose severity and cost can range between medical expenses and disability coverage, it's essential to protect yourself if the protection available isn't sufficient.

In this order, a provident insurance product emerges. This guarantees a capital sum in the event of the diagnosis of such an illness, this guarantee is known as *Serious and Dreaded Illness* or *Dreaded Affection*. Some difficulties appear on both the technical and commercial levels in the conception of such a product:

- pricing requires data and statistics that are not always available or accessible,
- the anti-selection phenomena that such a product can generate,
- the precise definition of the guaranteed risk.

Unlike death products in provident insurance, where death does not require a precise definition, *Critical Illness* definition, the dread disease must be listed and precisely defined in the contract so that there is no issues.

In the first part of this study, we propose to price different *Critical Illness and Dread Disease* benefits, initially considering only cancer and then extending the study to stroke and heart attack in order to obtain a price for the three most common conditions guaranteed by most of *Critical Illness and Dread Disease* contracts. The pricing is based on the theory of multi-state Markov processes.

In a second time, we have looked into how to integrate the results that would be obtained after the commercialization of such a product. We used Bayesian statistical inference theory to model future loss experience.

*Keywords* : *Critical Illness, Pricing, Provident insurance, Markovian multi state model, Weibull, Cancer, Stroke, Heart attack, Bayesian Inference, Credibility.*



# Note de Synthèse

## Problématique

Les produits d'assurance couvrant les maladies graves et redoutées ou Critical Illness sont des produits garantissant à l'assuré un capital forfaitaire en cas de diagnostic d'une des des maladies figurant dans le contrat d'assurance selon les définitions statuées par celui-ci.

Ce produit diffère de l'assurance frais de santé (mutuelle) ayant pour objet de prendre en charge tout ou partie des dépenses de santé concernant la maladie, l'accident et la maternité, non couvertes par l'assurance maladie obligatoire. En effet, les produits Critical Illness ont pour vocation d'aider financièrement l'assuré et sa famille les dépenses que vont occasionner un tel diagnostic comme par exemple la perte de revenus, la nécessité d'obtenir une aide extérieure, etc. C'est pour ces raisons que ces produits sont des produits de prévoyance.

L'objectif de ce mémoire est de proposer une méthodologie de tarification pour ce type de produits à partir de données publiques.

## Contexte

Par simplification, notre étude se concentrera sur les cancers, l'AVC et l'infarctus du myocarde, en effet, ces ALD représentent plus de 85% des cas.

Les cancers sont à l'origine de dix millions de décès en 2020 dans le monde, et représentent près d'un décès sur six, ils se placent en première cause de la mortalité mondiale. Les cancers, ou tumeurs malignes, les plus fréquents dans le monde sont le cancer du sein, le cancer du poumon, le cancer de la prostate, le cancer de la peau et le cancer colorectal.

L'incidence du cancer augmente à mesure que la population vieillit, en effet le vieillissement s'accompagne d'une perte d'efficacité de nos mécanismes de régénération et d'une intensification des facteurs de risques de cancers comme le tabagisme, la consommation d'alcool, l'obésité résultante d'une mauvaise alimentation ou d'un manque d'activité physique.

Selon l'OMS, dans le monde en 2019, il a été dénombré 12,22 millions d'AVC dont 6,55 millions de décès ( $\simeq 53\%$ ) confirmant l'AVC comme la deuxième cause de la mortalité mondiale. En France, on dénombre près de 80 000 infarctus du myocarde par an. D'après l'Institut national de Santé et de Recherche Médicale (INSERM), 10% des personnes ayant un tel infarctus décèdent dans l'heure, ce chiffre est en baisse grâce aux progrès de la médecine et à la plus grande disponibilité des équipes d'interventions. Les facteurs prédisposants sont : le tabagisme, le diabète, l'obésité, l'hypertension et le stress.

## Garanties

La garantie maladies graves et redoutées est un produit de prévoyance, pouvant être commercialisé en contrats individuels ou en contrat collectif. Il permet le déblocage d'une somme forfaitaire en cas de diagnostic pour subvenir aux besoins (absence de travail, perte de revenus, etc...) de l'assuré. Un accompagnement et des services complètent souvent cette garantie : la garde d'enfants, l'achat de matériel médicalisé, un réaménagement du domicile et une aide au retour au travail. Cette assurance a été introduite en 1983 en Afrique du Sud. Elle proposait alors seulement la couverture de quatre pathologies : le cancer, l'accident vasculaire cérébral (AVC), l'infarctus du myocarde et le pontage coronarien.

Actuellement, le volume de primes annuelles d'assurance Maladies Redoutées dans le monde est estimé à plus de 20 milliards d'euros ce qui représente 1% des primes d'assurance vie.

Les ALD entraînent souvent une hospitalisation et un suivi très long, engendrant des dépenses en santé conséquentes : pour un assuré aux Etats-Unis, un diagnostic de cancer engendre des frais de reste à charge moyen de l'ordre de 4.000\$ par patient. En France, le cancer et les différentes pathologies garanties sont dans la liste des 30 Affections de longue durée (ALD) dites "exonérantes". La Sécurité Sociale prend donc en charge l'intégralité du ticket modérateur des actes et médicaments afférents à la maladie. Cependant, le forfait journalier, la chambre particulière et le dépassement d'honoraire ne sont pas pris en charge par la Sécurité Sociale et restent à la charge de l'assuré en fonction de sa complémentaire santé.

Concernant le maintien de salaire, celui-ci est pris en charge à 50% par la Sécurité Sociale pendant 1095 jours (3 ans en incapacité de travail) et après une pension si passage en invalidité. En 2020, la loi de financement de la sécurité sociale crée un parcours de soin post traitement cancer. Ce parcours a pour vocation d'accompagner le patient sur trois différents postes en fonction de ses besoins : un soutien psychologique, des conseils diététiques et un accompagnement

à la pratique d'une activité sportive. Cependant, ce parcours est très limité : seules certaines structures portent ce projet (essentiellement en Ile de France), l'indemnisation totale ne peut excéder 180€ par personne et il est non renouvelable.

On voit donc l'importance de garanties Critical Illness pour compléter ce dispositif. On peut distinguer trois types de garanties "Maladies Redoutées" ou "Critical Illness" :

- ***Accelerated CI*** :

Garantit le paiement anticipé d'une partie du montant assuré en cas du diagnostic d'une des maladies garantie par le contrat. Le reste du montant assuré est versé lors du décès.

- ***Additional CI*** :

Garantit le paiement d'une prestation en cas du diagnostic d'une des maladies garantie par le contrat indépendamment du capital décès garanti.

- ***Stand-alone CI*** :

Ce type de contrat ne garantit pas le décès, il couvre le paiement d'une prestation unique en cas du diagnostic d'une des maladies garantie par le contrat et prend fin à la survenance de la pathologie.

## Données

Pour tarifer nous allons utiliser les données en incidence des cancers, de l'AVC et de l'infarctus du myocarde. Ainsi que la survie à horizon 1 an, 3 ans et 5 ans en cas de diagnostic du cancer, et enfin la mortalité due à ces ALD et toutes causes de la population française. Les statistiques d'incidence et de mortalité des cancers que nous avons utilisées sont extraites d'une étude conduite par Santé Publique France, le Registres des cancers Francim, les Hopitaux Civils de Lyon, l'Institut National de la Santé et de la Recherche Médicale (INSERM) et l'Institut National du Cancer : *Données globales d'épidémiologie des cancers (2020)*.

Les statistiques de survie à horizon 1 an, 3 ans et 5 ans en cas de diagnostic du cancer sont issues d'une autre étude conduite par l'Institut de la Veille Sanitaire (InVS), le Registres des cancers Francim, les Hopitaux Civils de Lyon, et l'Institut National du Cancer, *Survie des personnes atteintes de cancer en France métropolitaine (1989-2018) (2018)*. Les statistiques d'incidence des ALD que nous avons utilisées sont extraites du site de l'Assurance Maladie : *Nombre, incidence et caractéristiques des nouvelles exonérations pour ALD en 2019 (2020)*. Les statistiques de mortalité toute cause ainsi que l'effectif français par âge sont disponibles sur le site de l'INSEE.

## Tarification

Nous allons proposer une approche de tarification pour les produits *Stand Alone CI* et *Accelerated CI*, basée sur un modèle markovien à plusieurs états. La garantie *Stand Alone CI* indemnise un capital en cas de diagnostic d'une des maladies garanties par le contrat alors que la garantie *Accelerated CI* indemnise une avance du capital garanti en cas de diagnostic d'une des maladies garanties par le contrat et le restant du capital en cas de décès quelle qu'en soit la cause. Une police *Accelerated CI* peut donc être vue comme une garantie *Stand Alone CI* à laquelle on adjoint une garantie décès. [Gogola & Kopecká \(2017\)](#).

Nous pouvons donc définir les différents états que nous allons représenter sur la Figure 3.1 de la même manière que [Dash & Grimshaw \(1993\)](#), mais en utilisant un état différent par pathologie :

- $S$  : individu sain
- $C$  : individu diagnostiqué cancéreux
- $D_C$  : décès dû au cancer
- $A$  : individu victime d'un AVC
- $D_A$  : décès dû à un AVC
- $I$  : individu victime d'un infarctus du myocarde
- $D_I$  : décès dû à un infarctus du myocarde
- $D$  : décès dû aux autres causes

Nous pouvons donc exprimer la matrice des probabilités de transition pour un assuré d'un âge  $x$  à un âge  $x + t$ , en considérant les états  $\mathcal{E} := \{S, C, A, I, D_C, D_A, D_I, D\}$ , comme :

$${}_tP_x = \begin{pmatrix} {}_tP_x^{SS} & {}_tP_x^{SC} & {}_tP_x^{SA} & {}_tP_x^{SI} & 0 & 0 & 0 & {}_tP_x^{SD} \\ 0 & {}_tP_x^{CC} & 0 & 0 & {}_tP_x^{CD_C} & 0 & 0 & {}_tP_x^{CD} \\ 0 & 0 & {}_tP_x^{AA} & 0 & 0 & {}_tP_x^{AD_A} & 0 & {}_tP_x^{AD} \\ 0 & 0 & 0 & {}_tP_x^{II} & 0 & 0 & {}_tP_x^{ID_I} & {}_tP_x^{ID} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Nous cherchons maintenant à obtenir l'expression des différents coefficients de cette matrice.

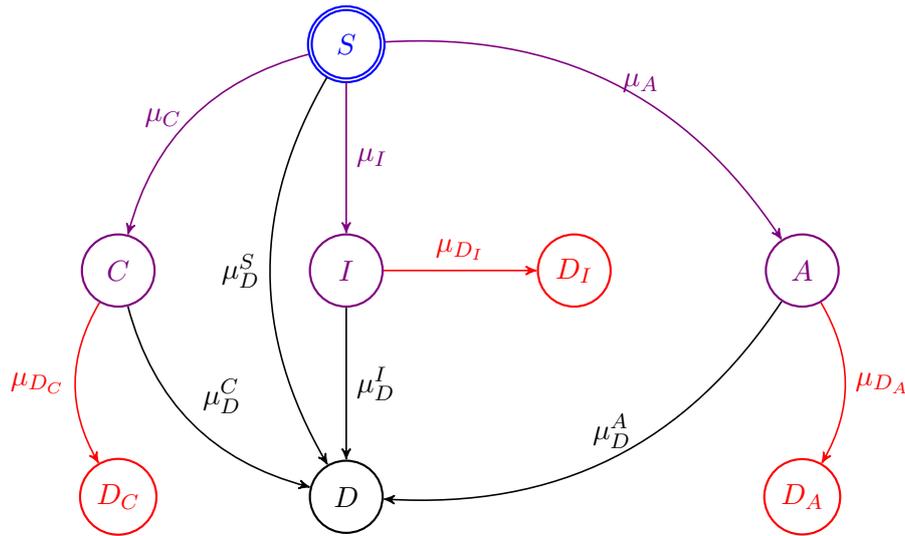


FIGURE 1 – Processus de Markov à plusieurs états, en fonction de l’âge

Nous utilisons les équations de Kolmogorov *forward* et *backward* dans le cadre des processus de Markov qui permettent de relier les probabilités de transitions entre elles. En effet cette matrice de transition possède les propriétés nécessaires des processus de Markov pour les utiliser.

Ce qui nous permet ensuite de déterminer la prime en fonction de la méthode d’estimation des probabilité de transition utilisée :

TABLE 1 – Primes nettes uniques en fonction de l’âge de souscription pour une garantie *Stand Alone*

age	méthode 1		méthode 2		méthode 3		méthode 4	
	H	F	H	F	H	F	H	F
35	198,51	449,81	202,21	448,05	295,35	392,94	289,94	394,49
40	345,38	632,05	356,20	630,81	392,57	426,59	380,64	427,44
45	691,74	859,52	725,79	861,46	575,01	486,94	548,03	485,85
50	1198,24	1119,74	1289,71	1132,07	1993,19	1385,64	1569,81	1260,41
55	1759,77	1386,36	1966,57	1424,21	2190,26	1528,72	1959,93	1488,09
60	2230,92	1618,84	2633,84	1708,01	2295,46	1677,49	2234,09	1706,00
65	2471,01	1766,10	3158,60	1942,16	3845,93	2211,68	3008,72	2011,18

Ces primes sont réalisées à l’aide de données publiques sur l’ensemble de la population française et non sur une population assurable. La différence de niveau de vie entraîne une réduction du risque et donc de la prime à appliquer sur un portefeuille collectif ou sélectionné.

Enfin, dans la dernière partie de ce mémoire, nous allons expliquer les méthodes d’inférence statistique classique et bayésienne pour estimer la probabilité des diagnostics des différentes maladies redoutées pour les hommes et les femmes et pour différentes classes d’âge.

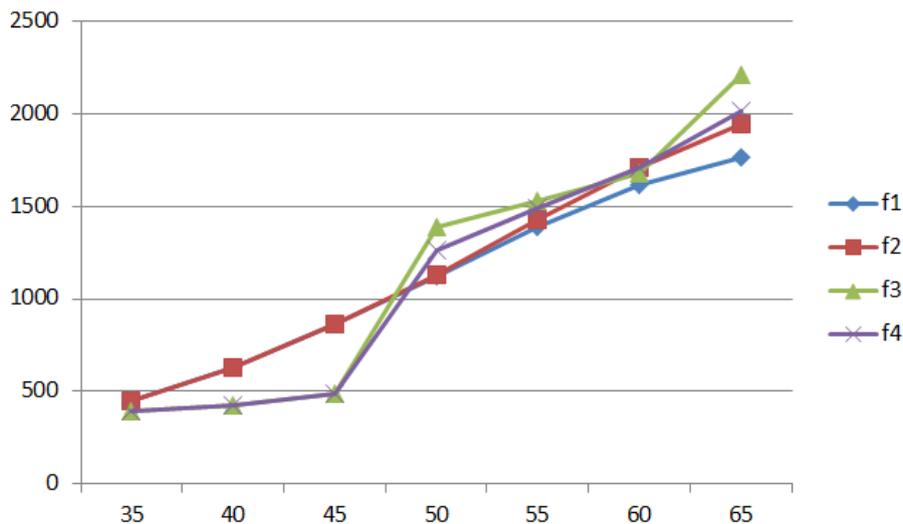


FIGURE 2 – Tracés des différentes Primes Unique Nettes en fonction de l'âge de souscription pour le sexe féminin

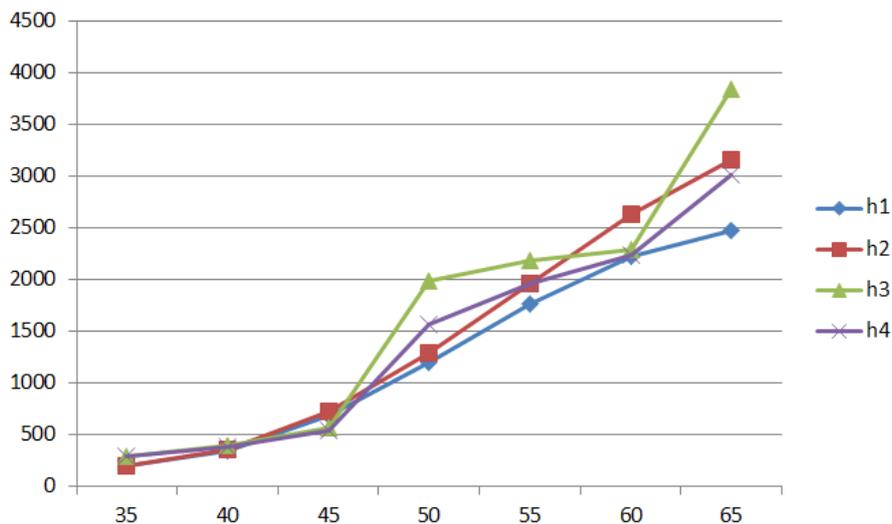


FIGURE 3 – Tracés des différentes Primes Unique Nettes en fonction de l'âge de souscription pour le sexe masculin

Les probabilités d'événements estimées sont ensuite utilisées pour fixer les primes de risque dans des groupes homogènes par sexe et par âge. Et ainsi ajuster le tarif déterminé dans la première partie à l'aide des résultats obtenus par l'assureur.

# Executive Summary

## Objective

Critical Illness insurance products guarantee the insured a lump-sum payment in the event of the diagnosis of one of the illnesses listed in the insurance contract, according to the definitions written in the contract.

This product differs from health insurance (Medical Expenses), whose purpose is to cover all or part of health expenses relating to illness, accident and maternity, not covered by compulsory health insurance. Actually, Critical Illness products are designed to provide financial assistance to policyholders and their families to cover the expenses incurred by such a diagnosis, such as loss of income, the need for outside help, etc. It is for these reasons that Critical Illness products are known as protection products.

The aim of this thesis is to propose a pricing methodology for this type of product based on public data.

## Context

For simplicity's sake, our study will focus on cancer, stroke and myocardial infarction, as these ALDs account for over 85% of cases.

Cancers were responsible for ten million deaths worldwide by 2020, accounting for almost one in every six deaths, making them the world's leading cause of mortality. The most common cancers, or malignant tumors, worldwide are breast cancer, lung cancer, prostate cancer, skin cancer and colorectal cancer.

The incidence of cancer rises as the population ages, since aging is accompanied by a loss of efficiency in our regenerative mechanisms and an intensification of cancer risk factors such as smoking, alcohol consumption, obesity resulting from poor diet or lack of physical activity.

According to the WHO, worldwide in 2019, there were 12.22 million strokes, including 6.55

million deaths ( $\simeq 53\%$ ), confirming stroke as the second leading cause of death worldwide. In France, there are almost 80,000 myocardial infarctions every year. According to the French National Institute for Health and Medical Research (INSERM), 10% of people who have such an infarction die within the hour, a figure that is falling thanks to advances in medicine and the greater availability of intervention teams. Predisposing factors include: smoking, diabetes, obesity, hypertension and stress.

## Guarantees

Critical illness coverage is a protection product that can be sold as individual or group contracts. In the event of diagnosis, a lump sum is released to cover the insured's needs (absence from work, loss of income, etc.). Support and services are often added to this guarantee, such as childcare, the purchase of medical equipment, home reorganization and assistance in returning to work. This type of insurance was introduced in South Africa in 1983. At this time, it covered only four illnesses: cancer, stroke, myocardial infarction and coronary bypass surgery.

At present, the volume of annual Dread Disease insurance premiums worldwide is estimated at over 20 billion euros, representing 1% of life insurance premiums.

Long-term illnesses often require lengthy hospitalization and follow-up, resulting in substantial healthcare costs: for an insured person in the United States, a cancer diagnosis generates average out-of-pocket expenses of around \$4,000 per patient. In France, cancer and the various pathologies it covers are included in the list of 30 "exonerating" long-term illnesses. The French Social Security system therefore covers the full co-payment for procedures and drugs related to the disease. However, the fixed daily rate, private room and excess fees are not covered by Social Security, and remain the responsibility of the policyholder, depending on his or her supplementary health insurance.

Regarding salary continuance, Social Security pays 50% of the salary for 1,095 days, and after a pension if you become disabled. In 2020, a French Social Security Law created a post-cancer treatment care pathway. The aim of this pathway is to support patients in three different areas according to their needs: psychological support, dietary advice and support in practicing a sporting activity. However, this program is very limited: only certain structures are involved (mainly in the Ile-de-France region), total compensation cannot exceed €180 per person, and it is non-renewable.

This highlights the importance of Critical Illness cover to complete the protection. There are three types of Critical Illness cover:

- ***Accelerated CI*** :

Guarantees early payment of a portion of the sum insured in the event of diagnosis of one of the illnesses covered by the policy. The remainder of the sum insured is paid upon death.

- ***Additional CI*** :

Guarantees the payment of a benefit in the event of the diagnosis of one of the illnesses covered by the policy, independently of the guaranteed death benefit.

- ***Stand-alone CI*** :

This type of contract does not guarantee death, but covers the payment of a single benefit in the event of the diagnosis of one of the illnesses covered by the contract, and ends when the illness occurs.

## Datas

For pricing purposes, we will use incidence data for cancer, stroke and myocardial infarction. We will also use 1-year, 3-year and 5-year survival rates for cancer diagnoses, and all-cause mortality for these long-term illnesses in the French population. The cancer incidence and mortality statistics we have used are taken from a study conducted by Santé Publique France, the Francim cancer registry, the Lyon Civil Hospitals, the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Institut National du Cancer. : [Données globales d'épidémiologie des cancers \(2020\)](#).

Survival statistics for 1, 3 and 5 years after cancer diagnosis are taken from another study conducted by the Institut de la Veille Sanitaire (InVS), the Francim cancer registry, the Hopitaux Civils de Lyon, and the Institut National du Cancer. [Survie des personnes atteintes de cancer en France métropolitaine \(1989-2018\) \(2018\)](#). The pathologies' incidence statistics we have used are taken from the Assurance Maladie website: [Nombre, incidence et caractéristiques des nouvelles exonérations pour ALD en 2019 \(2020\)](#). All-cause mortality statistics and the French population by age are available on the INSEE website.

## Pricing

We will propose a pricing methodology for the following products *Stand Alone CI* and *Accelerated CI*, based on a multi-state Markov model. The guarantee *Stand Alone CI* pays a

lump sum in the event of the diagnosis of one of the illnesses covered by the policy whereas the guarantee *Accelerated CI* indemnifies an advance of the capital sum guaranteed in the event of diagnosis of one of the illnesses covered by the policy, and the remainder of the capital sum in the event of death from any cause. An *Accelerated CI* policy can be seen as a *Stand Alone CI* to which a death benefit is added. [Gogola & Kopecká \(2017\)](#).

We can now define the different states that we will show on Figure 3.1 as is done in [Dash & Grimshaw \(1993\)](#), but with a different state by pathology :

- $H$ : healthy
- $C$ : cancer diagnosis
- $D_C$  : death due to cancer
- $S$  : victim of a stroke
- $D_S$  : death due to stroke
- $I$  : victim of a myocardial infarction
- $D_I$  : death due to a myocardial infarction
- $D$  : other death cause

We can therefore express the transition probability matrix for an insured person of age  $x$  to an age  $x + t$ , considering the states  $\mathcal{E} := \{H, C, S, I, D_C, D_S, D_I, D\}$ , like :

$${}_tP_x = \begin{pmatrix} {}_tP_x^{HH} & {}_tP_x^{HC} & {}_tP_x^{HS} & {}_tP_x^{HI} & 0 & 0 & 0 & {}_tP_x^{HD} \\ 0 & {}_tP_x^{CC} & 0 & 0 & {}_tP_x^{CD_C} & 0 & 0 & {}_tP_x^{CD} \\ 0 & 0 & {}_tP_x^{SS} & 0 & 0 & {}_tP_x^{SD_S} & 0 & {}_tP_x^{SD} \\ 0 & 0 & 0 & {}_tP_x^{II} & 0 & 0 & {}_tP_x^{ID_I} & {}_tP_x^{ID} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We now seek to obtain the expression of the different coefficients of this matrix. We use the Kolmogorov equations *forward* and *backward* in the context of Markov processes, which enable transition probabilities to be linked together. In fact, this transition matrix has all the properties needed for Markov processes.

This then allows us to determine the premium according to the transition probability estimation method used:

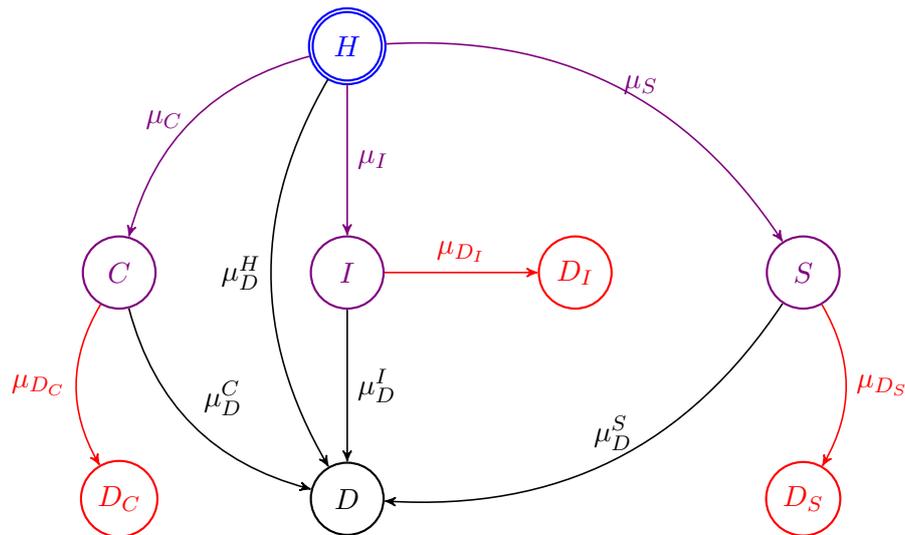


Figure 4 – Age-dependent multi-state Markov processes

Table 2 – Net single premiums as a function of age of purchase for a *Stand Alone* benefit

age	method 1		method 2		method 3		method 4	
	H	F	H	F	H	F	H	F
35	198,51	449,81	202,21	448,05	295,35	392,94	289,94	394,49
40	345,38	632,05	356,20	630,81	392,57	426,59	380,64	427,44
45	691,74	859,52	725,79	861,46	575,01	486,94	548,03	485,85
50	1198,24	1119,74	1289,71	1132,07	1993,19	1385,64	1569,81	1260,41
55	1759,77	1386,36	1966,57	1424,21	2190,26	1528,72	1959,93	1488,09
60	2230,92	1618,84	2633,84	1708,01	2295,46	1677,49	2234,09	1706,00
65	2471,01	1766,10	3158,60	1942,16	3845,93	2211,68	3008,72	2011,18

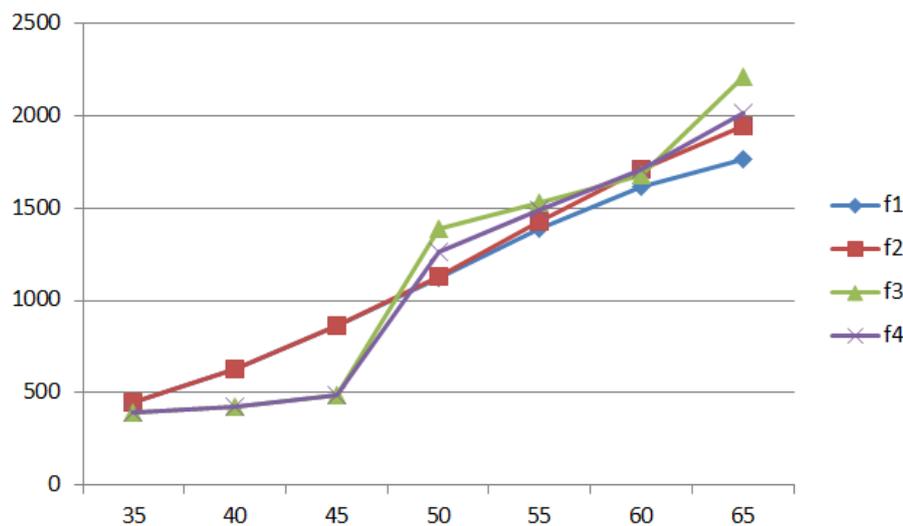


Figure 5 – Plots of the various Net Single Premiums as a function of underwriting age for the female gender

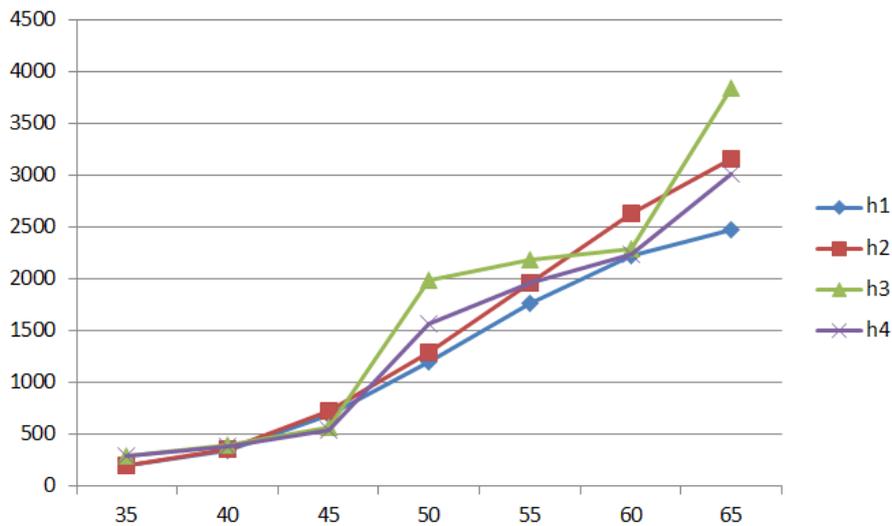


Figure 6 – Plots of the various Net Single Premiums as a function of underwriting age for the male gender

These premiums are calculated using public data on the entire French population, not on an insurable population. The difference in standard of living leads to a reduction in risk, and therefore in the premium to be applied to a collective or selected portfolio.

Finally, in the last part of this thesis, we will explain classical and Bayesian statistical inference methods for estimating the probability of diagnoses of different dread diseases for men and women and for different age classes.

The estimated event probabilities are then used to set risk premiums for homogeneous groups by gender and age. The results obtained by the insurer are then used to adjust the rate determined in the first part.

# Remerciements

En premier lieu, je tiens à remercier mon tuteur d'entreprise Medhi Himeur pour son accompagnement et ses précieux conseils tout au long de mon alternance qui a permis la rédaction de ce mémoire.

Je souhaite aussi remercier l'équipe de souscription de Covéa Coopération qui m'a accompagné dans ma démarche sur les plans moral et intellectuel.

Je tiens également à remercier Thomas Tabernero, Thomas Meerman, Kristi Hoti, Vincent Habib, Jurnerio Syziu et Gilles Chabanol, Responsable du service *Actuariat Recherche et Développement* pour leur aide, leur relecture et leurs précieux conseils durant la conduite de ce projet.

J'adresse également mes remerciements à l'ensemble de l'équipe pédagogique de l'ISFA pour l'encadrement au cours de mon cursus en actuariat, et plus particulièrement à Pierre-Olivier Goffard mon tuteur pédagogique pour son soutien et sa réactivité face à mes questionnements.



# Introduction

D'après l'Institut national du cancer (INCa), pour la France métropolitaine, le nombre de nouveaux cas de cancer s'élève à près de 382 000 en 2018, 204 600 chez l'homme et 177 400 chez la femme. Ce nombre est en constante augmentation depuis 30 ans. En effet, la population est vieillissante : la proportion de personnes âgées de plus de 60 ans est passée de 19% à 27% en 30 ans. Ce vieillissement est responsable de l'aggravation du nombre de cancers et plus particulièrement du sein ou de la prostate, qui sont respectivement les cancers les plus fréquents pour les femmes et pour les hommes.

Un autre facteur de cette augmentation est le perfectionnement des diagnostics par les professionnels de santé grâce aux avancées technologiques. Mais, simultanément, ce facteur est aussi responsable de la baisse de la létalité des cancers. Par ailleurs, certains comportements et pratiques à risque favorisent cette hausse du nombre de cas : par exemple, pour le cancer du poumon, les chiffres concernant les femmes augmentent en lien avec la consommation de tabac qui s'est installée dans leurs habitudes. Le cancer est donc devenue la première cause de la mortalité en France.

Bien que les frais de soins courants soient bien couverts par la Sécurité Sociale, en cas de survenance d'un cancer ou d'une maladie grave, les dépenses liées à une longue hospitalisation et des chirurgies complexes peuvent entraîner un reste à charge important pour l'assuré. Et en l'absence de complémentaire arrêt de travail ou chômage, une perte de revenus importante peut être subie pour le patient. Par conséquent, les produits de prévoyance "maladies graves et redoutées" sont de plus en plus demandés. En effet, ces produits aident à faire face à la maladie en garantissant à l'assuré un capital en cas de diagnostic de l'une des maladies graves prévues par le contrat souscrit.

La plupart des garanties « maladies graves et redoutées » assurent : le cancer, l'Accident

Vasculaire Cérébral (AVC) avec séquelles neurologiques, l'infarctus du myocarde, le pontage coronarien, la greffe d'un organe vital, la sclérose en plaques, l'insuffisance rénale nécessitant un traitement par dialyse, la maladie de Parkinson et la maladie d'Alzheimer. Les maladies couvertes diffèrent d'un contrat à l'autre. Ce capital, choisi par l'assuré en fonction du contrat, permet par exemple l'achat de matériel médical, la rémunération d'une assistance, le remboursement de prêts, etc...

Ces produits sont aussi populaires et, selon The Actuary, sont en développement en Asie, notamment en Corée du sud, Taïwan, Singapour, Hong Kong, etc... On constate par exemple une évolution de la somme assurée en Chine de \$3,000 à \$18,000 en 15 ans.

Le marché mondial des produits *Critical Illness* devrait constater une hausse de 16,2% dans les prochaines années : en 2021, le chiffre d'affaire était de 185,53 milliards de dollars, en 2022 il devrait atteindre les 215,49 milliards et poursuivre jusqu'à 330 milliards en 2026 selon Globe Newswire.

Dans cette étude, nous allons utiliser un modèle markovien à plusieurs états pour aboutir à une tarification de garanties "*Critical Illness*" ou "*CI*" standards. Les différents états introduits pour un individu sont : Sain, Diagnostiqué CI (un état pour chaque maladie garantie par le contrat) et Décédé. Les données que nous allons utiliser sont les taux d'incidence et de prévalence des cancers et des différentes affections de longues durées (ALD) garanties. Ces données nous permettront d'estimer les probabilités de transition entre ces différents états et donc de tarifier les différents contrats.

Dans un second temps, nous nous sommes demandé par quel moyen incorporer des données au fur et à mesure que l'historique de données s'enrichit, ce que nous n'avons pas encore pour l'instant, c'est pourquoi cette partie reste théorique. En effet Covéa Coopération ne possède pas de données pour de tels contrats, mais cela est à anticiper pour la suite. Une solution envisagée est alors l'utilisation de la théorie de la crédibilité et plus particulièrement la théorie Bayésienne.

# Chapitre 1

## Contexte et présentation du produit

Dans ce premier chapitre, nous allons tout d'abord présenter les différentes pathologies garanties par ces produits dans un contexte local et international. Par simplification, notre étude se concentrera sur les cancers, l'AVC et l'infarctus du myocarde, en effet, ces ALD représentent plus de 85% des contrats, selon GenRe : [Lim \(2022\)](#) et [Li \(2021\)](#) (ces études ont été réalisées en Asie mais cela nous paraît raisonnable concernant le périmètre français d'après la fiche produit de la garantie CARCEPT de Klésia [Fiche produit Maladies Redoutées-Carcept \(2022\)](#)). Ensuite nous définirons les différents types de produits "maladies redoutées" existants et nous détaillerons les notations actuarielles utiles à la suite de notre étude.

### 1.1 Contexte actuel des principales maladies garanties

Les Maladies Redoutées font partie de ce que la Sécurité sociale nomme « Affections Longue durée » (ALD). Ce sont des pathologies invalidantes qui nécessitent une prise en charge thérapeutique coûteuse et un accompagnement particulièrement important, bouleversant la vie de nombreuses personnes.

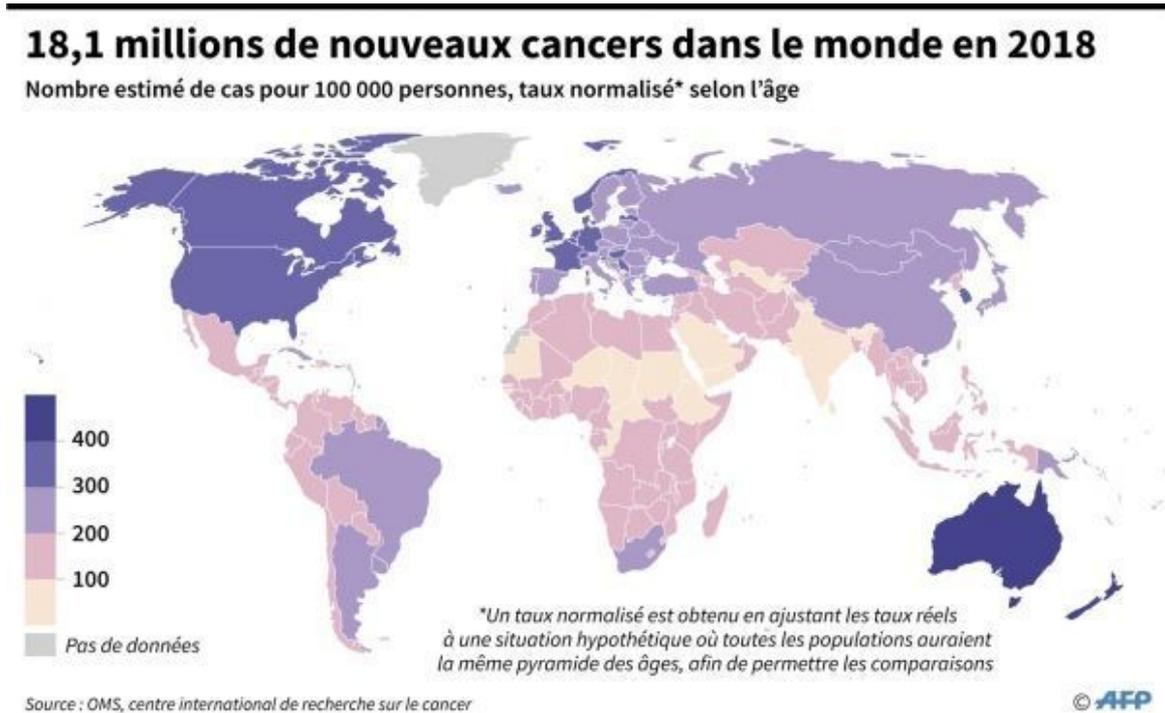
En France, on dénombre, en 2018, 11,4 millions de personnes prises en charge pour ALD et 157 400 décès selon l'Assurance Maladie.

#### 1.1.1 Le cancer

Les cancers sont à l'origine de dix millions de décès en 2020 dans le monde, et représentent près d'un décès sur six, ils se placent en première cause de la mortalité mondiale. Les cancers,

ou tumeurs malignes, les plus fréquents dans le monde sont le cancer du sein, le cancer du poumon, le cancer de la prostate, le cancer de la peau et le cancer colorectal.

L'incidence du cancer augmente à mesure que la population vieillit, en effet le vieillissement s'accompagne d'une perte d'efficacité de nos mécanismes de régénération et d'une intensification des facteurs de risques de cancers comme le tabagisme, la consommation d'alcool, l'obésité résultante d'une mauvaise alimentation ou d'un manque d'activité physique.



On voit sur cette carte que l'incidence des cancers est majoritaire dans les pays développés. Notamment en Asie où, d'après l'OMS, près de la moitié des nouveaux cas y seraient recensés. En effet, le développement économique de ces pays permet un meilleur recensement des cancers et le vieillissement de la population entraîne une incidence plus forte. De plus, le taux de nouveaux cancers chez la femme est en nette augmentation en Amérique du Nord, en Europe, en Chine et en Océanie. La croissance démographique et industrielle serait aussi un facteur favorisant ces pathologies.

L'Institut National du Cancer estime qu'entre 30 et 50 % des cancers pourraient être évités en réduisant ces différents facteurs de risque et en détectant plus rapidement ces tumeurs. En effet, pour un bon nombre de cancers, une détection précoce et une prise en charge adaptée conduisent à une probabilité de guérison importante.

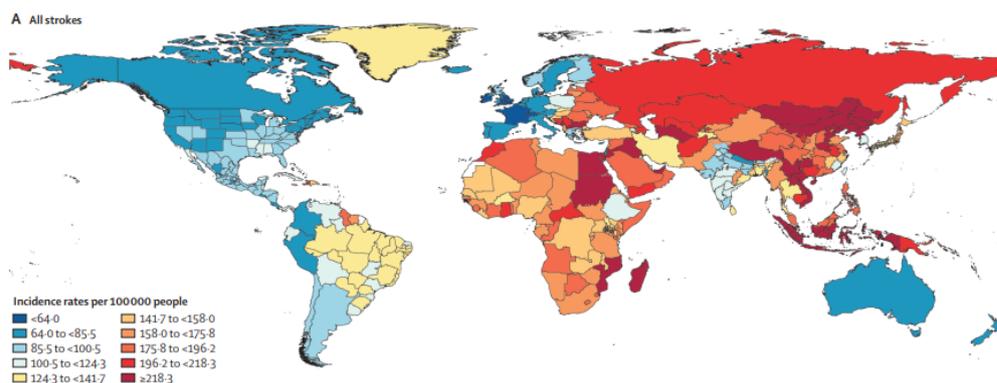


FIGURE 1.1 – Taux d'incidence d'AVC standardisés sur l'âge pour 100 000 personnes par pays, pour les deux sexes, 2019

### 1.1.2 L'AVC

Concernant l'Accident Vasculaire Cérébral (AVC), celui-ci survient lorsque la circulation sanguine est stoppée par un vaisseau sanguin bouché (AVC ischémique) ou lorsqu'un vaisseau sanguin est rompu (AVC hémorragique). L'AVC ischémique est le plus fréquent et le moins dangereux en termes de séquelles et de mortalité.

Il existe également l'AIT (Accident Ischémique Transitoire) ou mini AVC. Ses symptômes sont les mêmes que l'AVC, mais ils durent de quelques secondes à quelques minutes avant le retour à la normale. L'AIT peut donc passer inaperçu et être confondu avec un simple malaise. Il signale pourtant un risque important d'AVC plus grave et doit amener à consulter en urgence. Cependant l'AIT n'est souvent pas garanti par ces polices.

Selon l'OMS, dans le monde en 2019, il en a été compté 12,22 millions dont 6,55 millions de décès ( $\simeq 53\%$ ) confirmant l'AVC comme la deuxième cause de la mortalité mondiale.

Nous voyons sur cette carte que les taux d'incidence sont beaucoup plus importants en Afrique et en Asie. En France, on dénombre un AVC toutes les 4 minutes, selon la Fondation pour la recherche sur les AVC. Ils sont estimés aussi fréquents que l'infarctus du myocarde (ou crise cardiaque) et concernent essentiellement les plus de 65 ans. Les AVC mènent souvent à des séquelles neurologiques et psychologiques qui entraînent un suivi nécessaire du patient et potentiellement de son entourage. Son diagnostic entraîne un risque plus important (de l'ordre de +100%) de contracter des troubles cognitifs ou de la démence.

### 1.1.3 L'infarctus du myocarde

La troisième pathologie la plus souvent garantie par les contrats Critical Illness est l'infarctus du myocarde. Plus connue sous le nom de "crise cardiaque", elle correspond à la destruction partielle du muscle cardiaque en raison de l'obstruction d'une artère (artère coronaire), celle-ci ayant pour vocation d'alimenter le cœur en oxygène, l'absence d'oxygène (hypoxie) entraîne la destruction du muscle cardiaque.

En France, on dénombre près de 80 000 infarctus du myocarde par an. D'après l'Institut national de Santé et de Recherche Médicale (INSERM), 10% des personnes ayant un tel infarctus décèdent dans l'heure, ce chiffre est en baisse grâce aux progrès de la médecine et à la plus grande disponibilité des équipes d'interventions. Les facteurs prédisposants sont : le tabagisme, le diabète, l'obésité, l'hypertension et le stress.

Nous pouvons observer la différence d'incidence des infarctus du myocarde en fonction de la zone géographique sur ce graphique provenant d'une étude épidémiologique de cette pathologie en 2005, [Dujardin & Cambou \(2005\)](#).

On voit que ce type de garantie aurait le plus d'intérêt en Asie du Sud ainsi qu'en Amérique.

Nous avons donc vu que pour les trois pathologies précitées, les plus souvent garanties par les contrats de *Maladies Graves et Redoutées* : le cancer, l'AVC et l'infarctus du myocarde, le risque est grandissant et peut engendrer des difficultés économiques importantes pour l'assuré en fonction de sa couverture sociale. L'intérêt de ce type de produit est donc réel et en plein essor dans le monde.

### 1.1.4 La prévoyance en France

En France, le système de prévoyance est un ensemble de dispositifs sociaux qui permettent aux travailleurs et à leur famille de se prémunir contre les risques liés à la maladie, l'invalidité, le décès ou la dépendance. Il comprend notamment la Sécurité Sociale, qui offre une couverture médicale et des prestations sociales, ainsi que des régimes de retraite complémentaire obligatoires pour les salariés et des régimes de prévoyance complémentaire facultatifs pour les travailleurs non salariés et les salariés. Les employeurs ont également l'obligation de souscrire

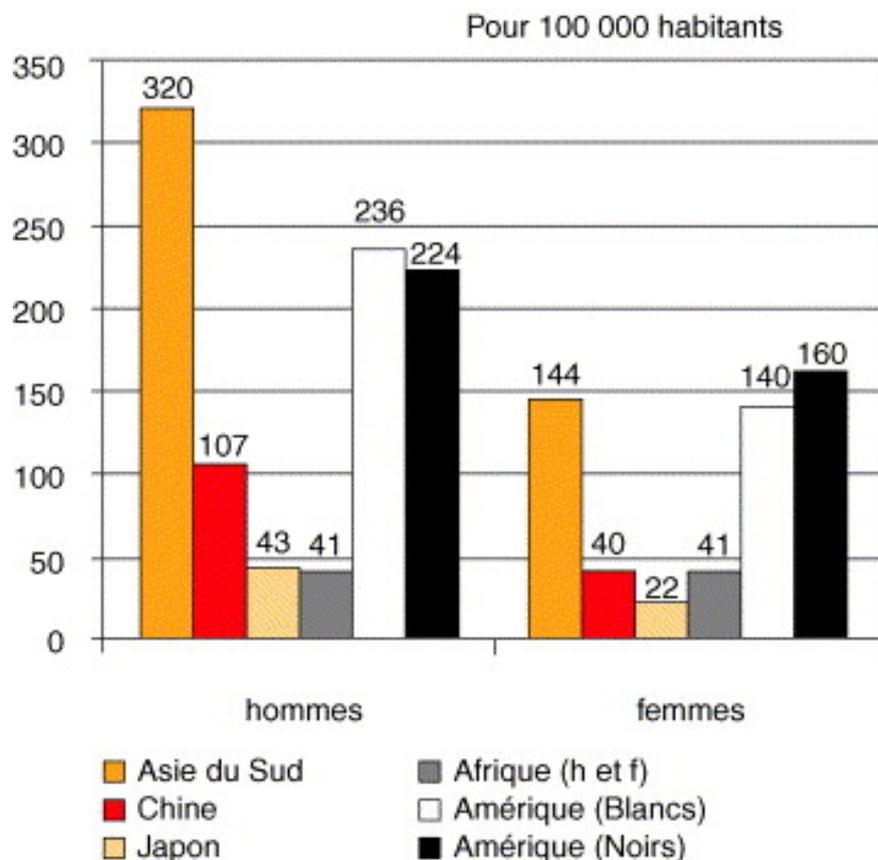


FIGURE 1.2 – Effectifs d'incidence d'infarctus du myocarde standardisés sur l'âge pour 100 000 personnes par zone géographique, pour les deux sexes, 2005 selon [Dujardin & Cambou \(2005\)](#)

une assurance en cas de décès ou d'invalidité pour leurs salariés. Le système de prévoyance en France repose sur le principe de solidarité et est financé par les cotisations des travailleurs, des employeurs et des contributions publiques.

Le rôle d'un contrat de prévoyance est de réduire au maximum les conséquences financières liées à ces événements et ainsi de protéger les familles face à une baisse de revenus. Contrairement à la complémentaire santé qui vient compléter les remboursements de la Sécurité Sociale, une assurance prévoyance permet de conserver son niveau de revenu en cas d'accident de la vie. La prévoyance couvre ainsi de nombreux domaines : prévoyance en cas d'accident, prévoyance famille, prévoyance santé, prévoyance dépendance, prévoyance décès, prévoyance retraite.

## 1.2 Présentation du produit

La garantie maladies graves et redoutées est un produit de prévoyance. Il permet le déblocage d'une somme forfaitaire en cas de diagnostic pour subvenir aux besoins (absence de travail, perte

## La prévoyance en France

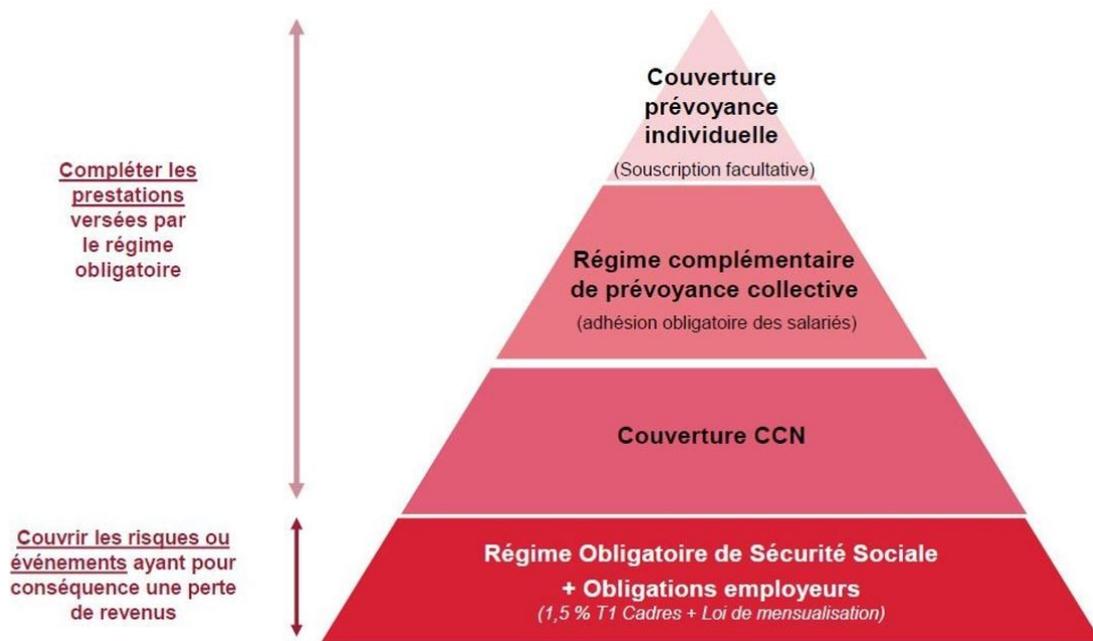


FIGURE 1.3 – Le système pyramidal de la prévoyance en France

de revenus, etc...) de l'assuré. Un accompagnement et des services complètent souvent cette garantie : la garde d'enfants, l'achat de matériel médicalisé, un réaménagement du domicile et une aide au retour au travail. Cette assurance a été introduite en 1983 en Afrique du Sud. Elle proposait alors seulement la couverture de quatre pathologies : le cancer, l'accident vasculaire cérébral (AVC), l'infarctus du myocarde et le pontage coronarien.

### 1.2.1 Historique de la garantie

Cette notion de couverture d'événement graves n'est pas nouvelle pour les assureurs, des tentatives de produits complémentaires aux garanties décès ont déjà été effectuées. En Australie, par exemple, avant la Seconde Guerre mondiale, un produit garantissant le paiement anticipé d'une partie du capital décès en cas d'intervention chirurgicale a été lancé. Aux États-Unis, en 1960, un produit nommé CANCER POLICY voit le jour : pour la première fois un capital est versé à l'assuré en cas de diagnostic d'une maladie. Ce produit connaît un grand succès : en 1986, l'American Family Life réalise un chiffre d'affaire de 287 millions de dollars pour un nombre d'assurés de l'ordre de deux millions de personnes. Il est ensuite introduit au Japon pour un succès similaire, en effet, le cancer est connu pour être la cause première de la mortalité au Japon.

En Afrique du Sud, après la première transplantation cardiaque réalisée, les maladies cardiovasculaires prennent toute leur importance et la couverture "cancer" se voit comme insuffisante. Et, en 1983, la première couverture "Dread Diseases" apparaît. Cette couverture est cependant, dans un premier temps, mal perçue par les réassureurs et les cédantes car les définitions des différentes maladies semblent imprécises au moment de la gestion des sinistres. De plus la tarification de tels produits se révèle complexe à apprécier.

Elle s'est depuis largement répandue dans le reste du monde et offre différentes prestations, avec une plus large couverture de maladies (ajout du VIH par exemple dans certains contrats), suivant les besoins des clients qui diffèrent selon le marché sur lequel elle est proposée. Les contrats précisent clairement les risques et maladies garanties et certaines exclusions ont été ajoutées comme notamment des formes de cancers peu sévères.

En France, parmi les premiers produits "*maladies redoutées*" on retrouve :

- CREAVIE en 1989 lancé par La Mondiale
- OMNIS en 1990 lancé par le groupe AZUR
- VIVANCE en 1989 lancé par EAGLE STAR VIE

Les deux premiers sont des produits temporaires décès avec anticipation d'une partie du capital en cas de diagnostic d'une des maladies couvertes. Le contrat VIVANCE, lui, garantit un capital en cas de décès ou en cas d'invalidité permanente et totale ou au diagnostic d'une des maladies couvertes. Nous voyons donc apparaître deux types de contrats intégrant la garantie décès.

### 1.2.2 Marché actuel

Actuellement, le volume de primes annuelles d'assurance Maladies Redoutées dans le monde est estimé à plus de 20 milliards d'euros ce qui représente 1% des primes d'assurance vie.

Plus de la moitié de ces primes proviennent d'Asie, en particulier de la Corée du sud, de la Chine, de la Malaisie et de Singapour, où le marché y est beaucoup plus développé : des produits couvrent la rechute ou simultanéité de plusieurs maladies pour un assuré. Il existe des options d'exonération de primes ou de rente mensuelle. Au Japon, des produits conjoints sur plusieurs risques comme la santé ou l'épargne existent, avec la possibilité de réduire ou d'étendre les couvertures.

Au Royaume Uni où ces produits ont connu un large succès, ces polices sont principalement associées à des polices d'assurance vie couvrant des prêts hypothécaires. En Australie, si ces produits sont aussi très populaires, ils sont moins liés aux prêts hypothécaires, [Brett & du Toit \(2007\)](#). Au Canada, le remboursement de la prime en cas de décès ou rachat si aucune indemnisation n'a été versée, a permis à cette assurance de se répandre largement car cela donne à l'assuré le sentiment d'être couvert dans toutes les situations. Des garanties juvéniles sont aussi fréquentes au Canada. En France, il existe quelques contrats "maladies redoutées" :

- La Caisse d'Epargne dans ses Garanties Famille Prévoyance propose une garantie *Essentielle* garantissant un capital en cas de décès ou de perte totale et irréversible d'autonomie (PTIA). Le contrat propose également une formule *Plus*, intégrant une avance de 20% du capital décès en cas de survenance d'une maladie redoutée. La garantie Critical Illness est unique et la prime est recalculée en cas de versement du capital anticipé dû à une maladie. Les maladies garanties sont : l'infarctus du myocarde, la chirurgie des artères coronaires, l'AVC, le cancer, les paralysies dûes aux accidents et les brûlures graves. La garantie CI cesse au plus tard la veille de l'échéance de cotisation qui suit le 65ème anniversaire de l'assuré. La colonne CI correspond à la différence des tarifs *Plus* et *Essentielle*.

Capital	25 000,00			45 000,00 €			100 000,00 €		
	Essentielle	Plus	CI	Essentielle	Plus	CI	Essentielle	Plus	CI
18-30 ans	58,80 €	61,20 €	2,40 €	100,80 €	105,60 €	4,80 €	219,60 €	231,60 €	12,00 €
31 ans	61,20 €	63,60 €	2,40 €	104,40 €	109,20 €	4,80 €	226,80 €	241,20 €	14,40 €
32 ans	62,40 €	66,00 €	3,60 €	108,00 €	114,00 €	6,00 €	236,40 €	249,60 €	13,20 €
33 ans	64,80 €	69,60 €	4,80 €	112,80 €	120,00 €	7,20 €	246,00 €	262,80 €	16,80 €
34 ans	68,40 €	72,00 €	3,60 €	117,60 €	126,00 €	8,40 €	256,80 €	273,60 €	16,80 €
35 ans	70,80 €	75,60 €	4,80 €	123,60 €	132,00 €	8,40 €	268,80 €	288,00 €	19,20 €
36 ans	74,40 €	79,20 €	4,80 €	129,60 €	138,00 €	8,40 €	282,00 €	303,60 €	21,60 €
37 ans	78,00 €	84,00 €	6,00 €	123,60 €	146,40 €	22,80 €	296,40 €	320,40 €	24,00 €
38 ans	81,60 €	87,60 €	6,00 €	142,80 €	153,60 €	10,80 €	310,80 €	338,40 €	27,60 €
39 ans	85,20 €	93,60 €	8,40 €	150,00 €	163,20 €	13,20 €	327,60 €	357,60 €	30,00 €
40 ans	90,00 €	97,20 €	7,20 €	157,20 €	172,80 €	15,60 €	344,40 €	378,00 €	33,60 €
41 ans	96,00 €	104,40 €	8,40 €	168,00 €	186,00 €	18,00 €	369,60 €	408,00 €	38,40 €
42 ans	103,20 €	112,80 €	9,60 €	182,40 €	200,40 €	18,00 €	399,60 €	441,60 €	42,00 €
43 ans	112,80 €	123,60 €	10,80 €	198,00 €	219,60 €	21,60 €	434,40 €	481,20 €	46,80 €
44 ans	121,20 €	134,40 €	13,20 €	213,60 €	237,60 €	24,00 €	470,40 €	524,40 €	54,00 €
45 ans	132,00 €	146,40 €	14,40 €	231,60 €	258,00 €	26,40 €	510,00 €	570,00 €	60,00 €
46 ans	142,80 €	157,20 €	14,40 €	250,80 €	280,80 €	30,00 €	553,20 €	619,20 €	66,00 €
47 ans	153,60 €	172,80 €	19,20 €	272,40 €	306,00 €	33,60 €	601,20 €	674,40 €	73,20 €
48 ans	166,80 €	187,20 €	20,40 €	295,20 €	333,60 €	38,40 €	651,60 €	736,80 €	85,20 €
49 ans	180,00 €	205,20 €	25,20 €	320,40 €	362,40 €	42,00 €	706,80 €	802,80 €	96,00 €
50 ans	195,60 €	222,00 €	26,40 €	346,80 €	396,00 €	49,20 €	766,80 €	874,80 €	108,00 €
51 ans	211,20 €	242,40 €	31,20 €	376,80 €	432,00 €	55,20 €	831,60 €	955,20 €	123,60 €
52 ans	229,20 €	262,80 €	33,60 €	408,00 €	470,40 €	62,40 €	902,40 €	1 040,40 €	138,00 €
53 ans	248,40 €	286,80 €	38,40 €	442,80 €	513,60 €	70,80 €	979,20 €	1 135,20 €	156,00 €
54 ans	268,80 €	313,20 €	44,40 €	480,00 €	559,20 €	79,20 €	1 062,00 €	1 237,20 €	175,20 €
55 ans	291,60 €	339,60 €	48,00 €	520,80 €	607,20 €	86,40 €	1 152,00 €	1 345,20 €	193,20 €
56 ans	316,80 €	369,60 €	52,80 €	564,00 €	661,20 €	97,20 €	1 249,20 €	1 462,80 €	213,60 €
57 ans	343,20 €	400,80 €	57,60 €	612,00 €	718,80 €	106,80 €	1 356,00 €	1 591,20 €	235,20 €
58 ans	372,00 €	435,60 €	63,60 €	663,60 €	780,00 €	116,40 €	1 471,20 €	1 728,00 €	256,80 €
59 ans	403,20 €	471,60 €	68,40 €	720,00 €	844,80 €	124,80 €	1 596,00 €	1 872,00 €	276,00 €
60 ans	436,80 €	511,20 €	74,40 €	781,20 €	914,40 €	133,20 €	1 731,60 €	2 029,20 €	297,60 €
61 ans	472,80 €	552,00 €	79,20 €	847,20 €	990,00 €	142,80 €	1 878,00 €	2 196,00 €	318,00 €
62 ans	513,60 €	596,40 €	82,80 €	919,20 €	1 070,40 €	151,20 €	2 037,60 €	2 374,80 €	337,20 €
63 ans	556,80 €	646,80 €	90,00 €	997,20 €	1 159,20 €	162,00 €	2 211,60 €	2 570,40 €	358,80 €
64 ans	603,60 €	698,40 €	94,80 €	1 081,20 €	1 254,00 €	172,80 €	2 398,80 €	2 784,00 €	385,20 €
65 ans	654,00 €	654,00 €	- €	1 172,40 €	1 172,40 €	- €	2 602,80 €	2 602,80 €	- €
66 ans	709,20 €	709,20 €	- €	1 272,00 €	1 272,00 €	- €	2 823,60 €	2 823,60 €	- €
67 ans	769,20 €	769,20 €	- €	1 380,00 €	1 380,00 €	- €	3 063,60 €	3 063,60 €	- €
68 ans	834,00 €	834,00 €	- €	1 497,60 €	1 497,60 €	- €	3 324,00 €	3 324,00 €	- €
69 ans	904,80 €	904,80 €	- €	1 624,80 €	1 624,80 €	- €	3 607,20 €	3 607,20 €	- €

FIGURE 1.4 – Primes commerciales des différentes formules proposées en fonction du capital garanti pour le produit Caisse d'Épargne

- Le produit Capital Diagnostic et Convalescence de CHUBB European Group SE garantit le versement d'un capital forfaitaire à l'assuré en cas de « 1er diagnostic » d'une Maladie Grave diagnostiquée et indemnisée, ainsi que le versement à l'assuré d'une indemnité journalière en cas d'Hospitalisation suite à la Maladie Grave diagnostiquée et indemnisée. Il garantit également le versement d'un forfait Convalescence à l'assuré suite à la Maladie Grave diagnostiquée et indemnisée, et pour les enfants assurés, le versement au souscripteur d'un capital forfaitaire en cas de « 1er diagnostic » d'une Maladie Grave diagnostiquée. Les maladies garanties sont les cancers, l'AVC et l'infarctus du myocarde. Le contrat peut être souscrit jusqu'aux 79 ans de

l'assuré.

	Cotisation mensuelle par assuré, selon sa tranche d'âge									
	18-29 ans	30-39 ans	40-44 ans	45-49 ans	50-54 ans	55-59 ans	60-64 ans	65-69 ans	70-74 ans	75-79 ans
<b>Option 1 : jusqu'à 10 000 € versés</b>	<input type="checkbox"/> 2,05 €	<input type="checkbox"/> 5,12 €	<input type="checkbox"/> 15,33 €	<input type="checkbox"/> 23,65 €	<input type="checkbox"/> 35,70 €	<input type="checkbox"/> 49,72 €	<input type="checkbox"/> 64,67 €	<input type="checkbox"/> 88,94 €	<input type="checkbox"/> 112,66 €	<input type="checkbox"/> 130,61 €
<b>Option 2 : jusqu'à 20 000 € versés</b>	<input type="checkbox"/> 4,10 €	<input type="checkbox"/> 10,23 €	<input type="checkbox"/> 30,68 €	<input type="checkbox"/> 47,30 €	<input type="checkbox"/> 71,38 €	<input type="checkbox"/> 99,43 €	<input type="checkbox"/> 129,34 €	<input type="checkbox"/> 177,86 €	<input type="checkbox"/> 225,34 €	<input type="checkbox"/> 261,24 €
<b>Option 3 : jusqu'à 30 000 € versés</b>	<input type="checkbox"/> 5,23 €	<input type="checkbox"/> 13,04 €	<input type="checkbox"/> 39,12 €	<input type="checkbox"/> 60,30 €	<input type="checkbox"/> 91,00 €	<input type="checkbox"/> 126,78 €	<input type="checkbox"/> 164,91 €	<input type="checkbox"/> 226,78 €	<input type="checkbox"/> 287,30 €	<input type="checkbox"/> 333,08 €

FIGURE 1.5 – Primes commerciales des différentes formules proposées en fonction du capital garanti par CHUBB

- Le produit CARCEPT PREV de KLESIA propose un contrat collectif décliné en trois formules : la première ne garantissant que les cancers, la deuxième garantit les cancers et l'AVC et la troisième ajoute les maladies coronaires, l'insuffisante rénale terminale, la greffe d'organe majeure et la sclérose en plaque. Le capital est de 5 000€ ou de 10 000€. Ce contrat étant collectif, une seule formule sera retenue par l'employeur en fonction de son coût et la cotisation est exprimée en pourcentage du salaire annuel brut (limité à 2 PASS, soit 82 272€)

	Formule 1		Formule 2		Formule 3	
<b>Capital Garanti</b>	5 000,00 €	10 000,00 €	5 000,00 €	10 000,00 €	5 000,00 €	10 000,00 €
<b>Prix en %SAB</b>	0,130%	0,260%	0,158%	0,301%	0,240%	0,460%

FIGURE 1.6 – Primes commerciales des différentes formules proposées en fonction du capital garanti par Klésia

- Le produit PREVI FAMILLE SERENITE de SURAVENIR (filiale d'assurance-vie et de prévoyance du Crédit Mutuel Arkéa) a été lancé en décembre 2022. Ce contrat de prévoyance individuelle a été décliné en deux formules, très similairement au produit de la Caisse d'Épargne : la formule *Essentielle* garantissant un capital allant de 20 000€ à 1 000 000€ en cas de décès ou de PTIA, et la formule *Optimale* qui, en plus de la garantie *Essentielle*, garantit un capital de 30% du capital décès (allant de 6 000€ à 300 000€) en cas de survenance d'une des maladies graves couvertes : Cancer, AVC, Infarctus du myocarde, Alzheimer et Parkinson.

Les tarifs de ce produit ne sont pas publics.

Les ALD entraînent souvent une hospitalisation et un suivi très long, engendrant des dépenses en santé conséquentes : pour un assuré aux Etats-Unis, un diagnostic de cancer engendre des frais de reste à charge moyen de l'ordre de 4.000\$ par patient. En France, le cancer et les différentes pathologies garanties sont dans la liste des 30 Affections de longue durée (ALD) dites "exonérantes". La Sécurité Sociale prend donc en charge l'intégralité du ticket modérateur des actes et médicaments afférents à la maladie. Cependant, le forfait journalier, la chambre particulière et le dépassement d'honoraire ne sont pas pris en charge par la Sécurité Sociale et restent à la charge de l'assuré en fonction de sa complémentaire santé. Concernant le maintien de salaire, celui-ci est pris en charge à 50% par la Sécurité Sociale pendant 1095 jours (3 ans en incapacité de travail) et après une pension si passage en invalidité.

En 2020, la loi de financement de la sécurité sociale crée un parcours de soin post traitement cancer. Ce parcours a pour vocation d'accompagner le patient sur trois différents postes en fonction de ses besoins : un soutien psychologique, des conseils diététiques et un accompagnement à la pratique d'une activité sportive. Cependant, ce parcours est très limité : seules certaines structures portent ce projet (essentiellement en Ile de France), l'indemnisation totale ne peut excéder 180€ par personne et il est non renouvelable.

On voit donc l'importance de garanties Critical Illness pour compléter ce dispositif.

### 1.2.3 Définition des garanties

Ces contrats peuvent être commercialisés en contrats individuels : les garanties sont alors choisies par l'assuré (dans la limite de la modularité acceptée par l'assureur), l'assureur va alors procéder à une tarification plus spécifique en introduisant un questionnaire médical dans la plupart des cas afin de préciser le risque de la souscription de ce contrat et un délai de carence sera souvent ajouté afin de réduire l'antisélection que pourrait générer un tel contrat. Ou bien en contrat collectif d'entreprise : dans ce cas, le chef d'entreprise choisit une garantie commune à l'ensemble de ses salariés (cela peut-être distingué par catégories socio-professionnelles) en complément du contrat de prévoyance d'entreprise.

On distingue trois différentes périodes dans la souscription d'un contrat de maladie redoutées :

- Une période de carence :

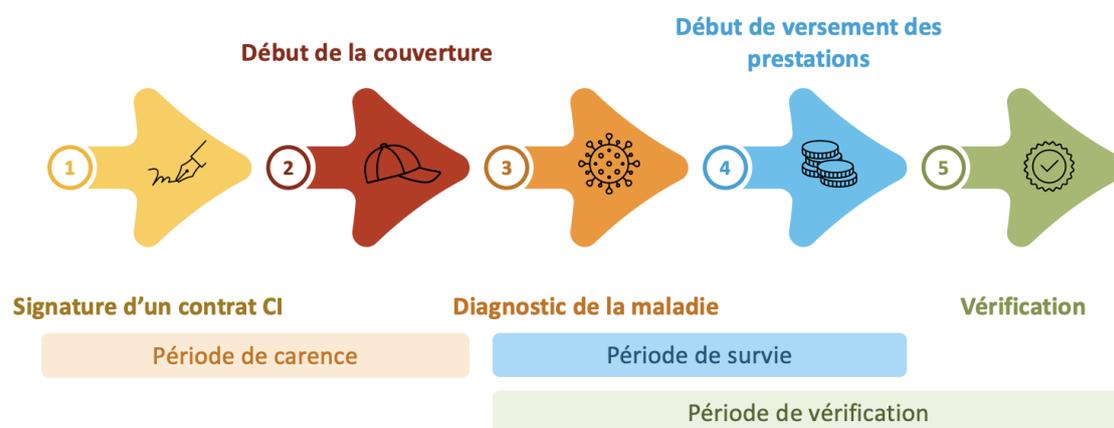
Elle est optionnelle et suit la signature du contrat, pendant cette période l'assuré ne peut pas recevoir de prestations en cas de survenance d'une maladie grave.

- Une période de survie :

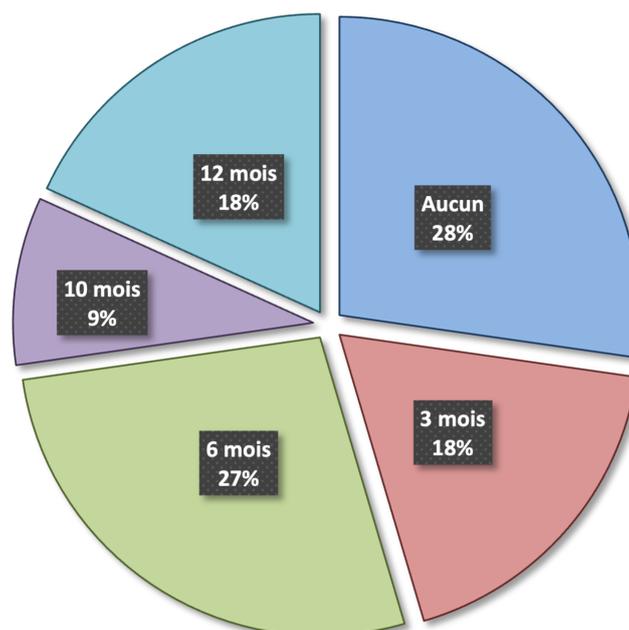
Souvent entre deux et trois semaines suivant le diagnostic, elle conditionne les prestations dues au titre de la maladie diagnostiquée à la survie de l'assuré durant cette période.

- Une période de vérification :

De l'ordre de quelques mois, elle permet de confirmer la légitimité de la prise en charge et ensuite ouvre droit au paiement des prestations.



### Délai de carence



On peut distinguer trois types de garanties "Maladies Redoutées" ou "Critical Illness" :

- **Accelerated CI** :

Garantit le paiement anticipé d'une partie du montant assuré en cas du diagnostic d'une des maladies garantie par le contrat. Le reste du montant assuré est versé lors du décès.

- **Additional CI** :

Garantit le paiement d'une prestation en cas du diagnostic d'une des maladies garantie par le contrat indépendamment du capital décès garanti.

- **Stand-alone CI** :

Ce type de contrat ne garantit pas le décès, il couvre le paiement d'une prestation unique en cas du diagnostic d'une des maladies garantie par le contrat et prend fin à la survenance de la pathologie.

Les avantages de proposer la garantie Maladies redoutées comme une accélération de la prestation décès sont les suivants :

- Le coût de la garantie Maladies redoutées est nettement inférieur à celui d'une garantie supplémentaire, ce qui augmente le potentiel de vente.
- L'incertitude sur les taux de prime demandés peut être réduite pour une garantie d'accélération où nous ne nous préoccupons que du surplus que cette garantie va générer sur la garantie décès.

Cependant, il y a aussi des avantages à offrir une garantie supplémentaire :

- La maladie redoutée peut être vendue en supplément de contrats qui ne contiennent pas de couverture décès.
- Après avoir souffert d'une maladie redoutée, l'assuré peut être inassurable ou faire l'objet d'une tarification élevée, de sorte que le maintien de l'ensemble de la couverture vie (le cas échéant) serait extrêmement difficile à assurer.
- Le revenu des primes est augmenté.

Deux problèmes peuvent se poser avec une prestation supplémentaire si le décès survient peu de temps après l'apparition d'une maladie redoutée. Premièrement, s'il y a une couverture décès attachée, le paiement des deux sinistres peut donner lieu à des prestations trop élevées puisqu'il n'y aura pas assez de temps pour que la prestation pour maladie redoutée soit utilisée pour l'un des objectifs suggérés dans l'introduction.

Deuxièmement, il pourrait être difficile de vérifier la validité d'une demande de prestation

pour maladie redoutée. Par exemple, en cas de décès résultant d'une crise cardiaque, il n'est pas toujours évident que cet événement réponde à la définition du contrat d'une maladie redoutée. Ces deux problèmes peuvent être résolus en subordonnant le versement de la prestation pour maladie redoutée à la condition que le demandeur survive à la maladie sur une période différée de 90 jours, par exemple. Lorsque le contrat comprend une couverture décès, il existe une alternative qui peut s'avérer plus intéressante sur le plan commercial : garantir le capital dû au titre de la maladie redoutée dès le diagnostic, en tant qu'accélération de la couverture, mais en rétablissant la totalité du capital décès après la période de différé.

Afin de limiter l'antisélection dans le cadre des contrats de maladies redoutées, il est généralement considéré comme souhaitable de limiter la prestation maximale autorisée dans le cadre d'un contrat. La police doit également permettre un paiement réduit lorsque le demandeur a d'autres contrats (y compris une couverture collective). Une indemnité maximale de maximum de 100 000 £ est couramment appliqué au Royaume-Uni, bien que certains produits permettent des montants plus élevés. La souscription, tant financière que médicale, doit être plus stricte si la somme assurée dépasse la norme du marché.

#### 1.2.4 Réassurance d'un contrat CI

La réassurance d'un produit comme les garanties Maladies graves et redoutées est très sollicitée car les réassureurs ont une expertise souvent plus poussée que les cédantes pour ces types de risques inhabituels. Un réassureur est bien souvent implanté dans le monde entier, par conséquent, il acquiert une connaissance du marché mondial et peut avoir accès à plus d'informations, comme par exemple des données étrangères, des portefeuilles de cédantes,... Et en plus de cela, l'appétence au risque du réassureur est parfois plus élevée qu'une cédante en raison de fonds propres plus importants qui permettent de souscrire des dossiers pour des risques mal maîtrisés ou méconnus tout en améliorant au fur et à mesure la maîtrise.

En France, les réassureurs fournissent l'essentiel du service de sélection médicale pour les demandes d'adhésion hors normes ou complexes, et particulier pour ce types de garanties. On distingue quatre grands types de contrats en réassurance :

- Le traité de réassurance obligatoire permet de réassurer la totalité d'un groupe de polices, voir d'un portefeuille sans distinction particulière.
- Le traité facultatif obligatoire (facob) permet à l'assureur de céder une partie des polices d'assurance facultativement, contrairement au réassureur qui a l'obligation d'accepter les contrats d'assurance cédés.
- La réassurance facultative permet de céder des polices d'assurance spécifiques sous des conditions particulières. Le réassureur n'a pas l'obligation de l'accepter comme la cédante ne doit pas obligatoirement les céder.
- Les titres financiers liés à l'assurance (Insured Linked Securities - ILS) sont des contrats de réassurance alternative. Ils sont définis comme des instruments financiers dont la valeur est corrélée avec la probabilité de réalisation des événements assurables résultant une perte à l'assureur. Dans le cadre de ce mémoire, nous étudierons des portefeuilles des assurances maladies redoutées qui sont cédés sous forme de traités.

Les différentes formes de réassurance sont :

- La réassurance proportionnelle :

Cette forme de réassurance répartit proportionnellement les gains et les pertes entre le réassureur et la cédante. Les deux types de traités sont le traité en *Quote-part* (ou *Quota share*, QP) où un pourcentage de rétention est défini pour les primes et les sinistres ; et le traité en *Excédent de plein* où un plein de rétention est fixé par classe de risque et les

sinistres et les primes sont réparties proportionnellement à ce plein (ou *line*).

— La réassurance non proportionnelle :

Cette forme de réassurance définit un seuil de sinistralité ou de perte pour la cédante à partir duquel le réassureur intervient. On peut distinguer les traités en *Excédent par sinistre* (ou *XS tête*), *Excédent par événement* (ou *XS Cat*) et *l'Excédent annuel* (ou *Stop Loss, SL*)

L'apport de ce service se traduit le plus souvent par une réassurance en *Quote-Part* avec un effet de durée : le réassureur restant engagé jusqu'à l'expiration naturelle des polices assumant ainsi les conséquences de la sélection médicale qu'il a préconisée. Plus rarement, on peut rencontrer des couvertures non proportionnelles (de type *XS tête* ou *XS Cat*) qui englobent dans leur périmètre ces garanties.

Au niveau des zones géographiques concernées, les pays du Moyen Orient, la Corée et d'autres pays asiatiques sollicitent régulièrement les réassureurs pour de tels produits, souvent en obligatoire pour des groupes ce qui permet de limiter l'antisélection.

## Chapitre 2

# Présentation des données

Dans ce chapitre, nous allons présenter les différentes données sur la population française utilisées dans cette étude. Pour tarifier nous allons utiliser les données en incidence des cancers, de l'AVC et de l'infarctus du myocarde. Ainsi que la survie à horizon 1 an, 3 ans et 5 ans en cas de diagnostic du cancer, et enfin la mortalité due à ces ALD et toutes causes de la population française. Les statistiques d'incidence et de mortalité des cancers que nous avons utilisées sont extraites d'une étude conduite par Santé Publique France, le Registres des cancers Francim, les Hopitaux Civils de Lyon, l'Institut National de la Santé et de la Recherche Médicale (INSERM) et l'Institut National du Cancer : *Données globales d'épidémiologie des cancers* (2020). Les statistiques de survie à horizon 1 an, 3 ans et 5 ans en cas de diagnostic du cancer sont issues d'une autre étude conduite par l'Institut de la Veille Sanitaire (InVS), le Registres des cancers Francim, les Hopitaux Civils de Lyon, et l'Institut National du Cancer, *Survie des personnes atteintes de cancer en France métropolitaine (1989-2018)* (2018). Les statistiques d'incidence des ALD que nous avons utilisées sont extraites du site de l'Assurance Maladie : *Nombre, incidence et caractéristiques des nouvelles exonérations pour ALD en 2019* (2020). Les statistiques de mortalité toute cause ainsi que l'effectif français par âge sont disponibles sur le site de l'INSEE.

### 2.1 Données en incidence

Ces données sont issues d'une projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017, étude conduite par Santé Publique France, le Registres des cancers Francim, les Hopitaux Civils de Lyon, l'Institut National de la Santé et de la Recherche Médicale (INSERM) et l'Institut National du Cancer, visant à estimer le nombre de diagnostics des cancer et de décès qu'ils occasionnent. Ces projections ont été réalisées à l'aide des données

de mortalité fournies par le Centre épidémiologique sur les causes médicales de décès et par l'INSERM sur la période 1975-2013. Les méthodes de projection utilisées sont : en prolongeant la tendance récente observée, en considérant comme constant le risque entre 2013 et 2017, ou en prolongeant la tendance moyenne entre 1980 et 2013. Compte tenu de la durée de projection (entre 2013 et 2017), c'est, dans la plupart des cas, la première méthode qui a été retenue. Toutefois dans quelques cas exceptionnels, les autres méthodes se révélaient être plus précises.

Tranche d'âge	Homme	Femme
[0; 14]	919	764
[15; 49]	14 472	25 874
[50; 64]	57 690	50 616
[65; 74]	75 247	46 442
[75; 84]	45 728	35 881
[85; ++]	19 965	26 028

TABLE 2.1 – Nombre de nouveaux cas en 2017 en France Métropolitaine par classe d'âge chez l'homme et la femme

Pour obtenir le taux d'incidence, nous avons besoin des données sur la population française pendant cette même année. Ces données sont disponibles sur le site de l'INSEE.

Tranche d'âge	Homme	Femme
[0; 14]	5 972 195	5 711 043
[15; 49]	13 886 417	14 006 207
[50; 64]	6 030 274	6 412 508
[65; 74]	3 092 968	3 494 540
[75; 84]	1 656 073	2 292 825
[85; ++]	648 902	1 435 181

TABLE 2.2 – Population par classe d'âge et par sexe en 2017 en France Métropolitaine

Grâce à ces informations, nous pouvons déduire les taux d'incidence de cancer par sexe et par tranche d'âge.

Tranche d'âge	Homme	Femme
[0; 14]	0,153880	0,133776
[15; 49]	1,042217	1,847324
[50; 64]	9,566729	7,893323
[65; 74]	24,328412	13,289875
[75; 84]	27,612309	15,649254
[85; ++]	30,767358	18,135692

TABLE 2.3 – Taux d'incidence pour 100 000 des cancers par classe d'âge et par sexe en 2017 en France Métropolitaine

Nous avons représenté les taux d'incidence pour les deux sexes sur le graphique ci-contre [2.1](#) :

Nous constatons un écart assez important entre les taux d'incidence homme et femme, notamment entre 65 et 84 ans, cela aura nécessairement un impact sur la tarification.

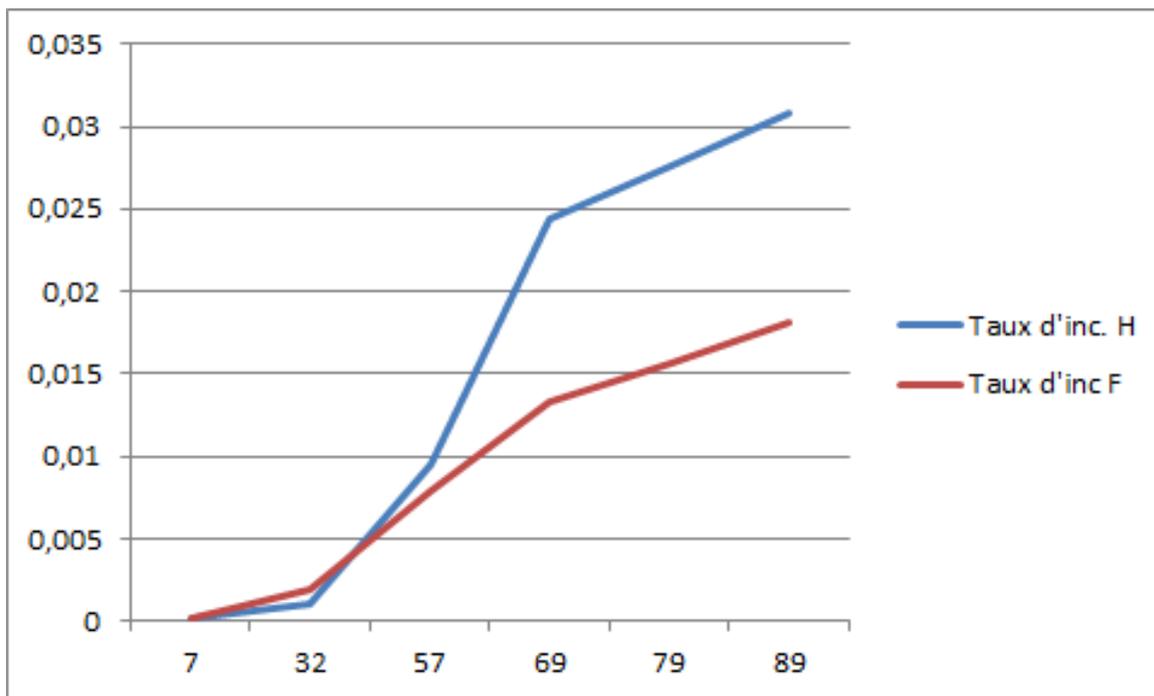


FIGURE 2.1 – Tracés des taux d'incidence du cancer en fonction de l'âge et du sexe.

De même concernant les taux d'incidence de l'AVC et de l'infarctus du myocarde, nous avons pu trouver sur le site de l'Assurance Maladie, une étude regroupant des données sur les effectifs des individus ayant contracté une de ces pathologies en 2019.

Tranche d'âge	Homme	Femme
[0; 4]	210	160
[5; 9]	90	60
[10; 14]	80	60
[15; 19]	100	100
[20; 24]	170	190
[25; 29]	200	230
[30; 34]	280	380
[35; 39]	510	480
[40; 44]	790	720
[45; 49]	1 530	1 190
[50; 54]	2 290	1 500
[55; 59]	3 190	1 770
[60; 64]	3 700	2 210
[65; 69]	4 300	2 870
[70; 74]	4 610	3 760
[75; 79]	3 630	3 630
[80; 84]	2 960	4 110
[85; 89]	1 990	3 840
[90; 94]	820	2 300
[95; ++]	200	760

TABLE 2.4 – Effectif des nouveaux diagnostics d'AVC par classe d'âge et par sexe en 2019 en France Métropolitaine

Que nous avons mis en relation avec les effectifs de la population française en 2019 donnés par l'INSEE :

Tranche d'âge	Homme	Femme
[0; 4]	1 822 919	1 749 149
[5; 9]	2 017 255	1 927 003
[10; 14]	2 043 994	1 952 698
[15; 19]	2 038 061	1 927 550
[20; 24]	1 828 640	1 784 178
[25; 29]	1 815 393	1 851 661
[30; 34]	1 917 425	2 018 490
[35; 39]	2 008 334	2 097 100
[40; 44]	1 966 705	2 008 973
[45; 49]	2 187 169	2 222 879
[50; 54]	2 118 506	2 190 341
[55; 59]	2 034 336	2 151 830
[60; 64]	1 888 253	2 074 228
[65; 69]	1 802 846	2 018 355
[70; 74]	1 481 970	1 713 914
[75; 79]	944 980	1 184 415
[80; 84]	744 505	1 092 432
[85; 89]	465 998	864 943
[90; 94]	182 981	469 596
[95; ++]	43 583	168 366

TABLE 2.5 – Population par classe d'âge et par sexe en 2019 en France Métropolitaine

Pour obtenir les taux d'incidence par classe d'âge et par sexe :

Nous avons tracé ces différents taux d'incidence :

Tranche d'âge	Homme	Femme
[0; 4]	0,115199853	0,087771316
[5; 9]	0,044615083	0,029743389
[10; 14]	0,039139058	0,030726718
[15; 19]	0,049066245	0,051879329
[20; 24]	0,092965264	0,106491617
[25; 29]	0,110168983	0,124212801
[30; 34]	0,146029180	0,188259541
[35; 39]	0,253941824	0,228887511
[40; 44]	0,401687086	0,358392074
[45; 49]	0,699534421	0,535341780
[50; 54]	1,080950443	0,684824874
[55; 59]	1,568079216	0,822555685
[60; 64]	1,959483184	1,065456642
[65; 69]	2,385117753	1,421950053
[70; 74]	3,110724239	2,193809024
[75; 79]	3,841351140	3,064804144
[80; 84]	3,975795999	3,762247902
[85; 89]	4,270404594	4,439598910
[90; 94]	4,481339593	4,897827068
[95; ++]	4,588945231	4,513975506

TABLE 2.6 – Taux d'incidence pour 100 000 des AVC par classe d'âge et par sexe en 2019 en France Métropolitaine

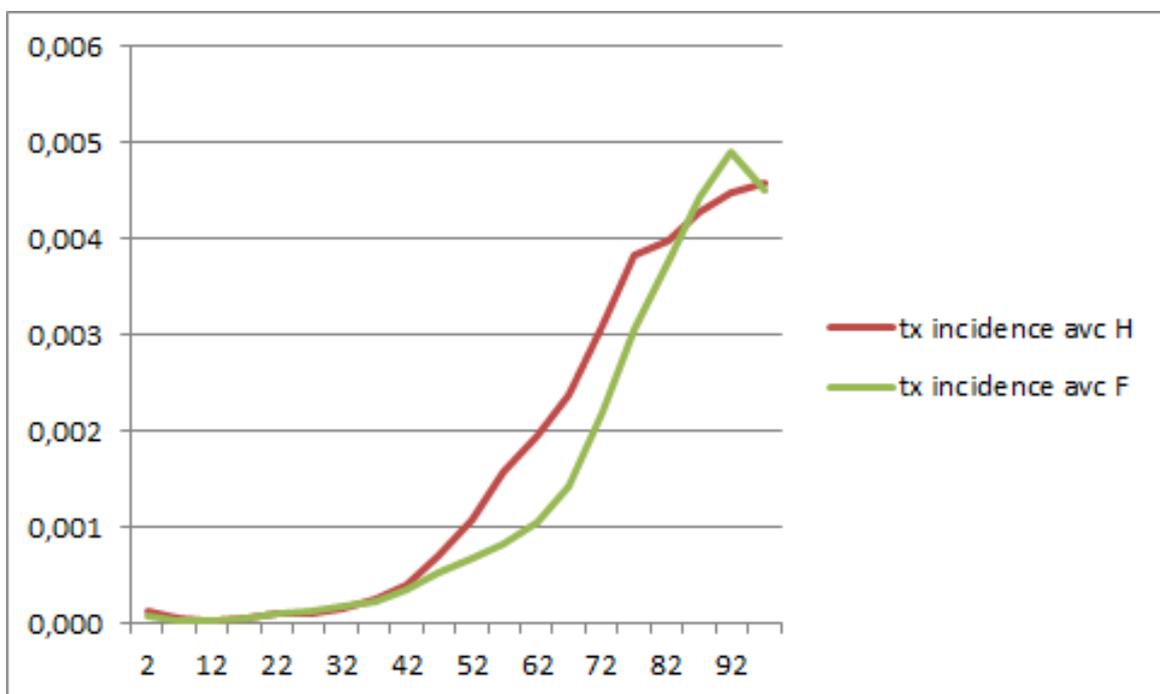


FIGURE 2.2 – Tracés des taux d'incidence de l'AVC en fonction de l'âge et du sexe.

Concernant l'infarctus du myocarde :

Tranche d'âge	Homme	Femme
[0; 4]	60	40
[5; 9]	10	10
[10; 14]	10	5
[15; 19]	10	10
[20; 24]	40	30
[25; 29]	100	150
[30; 34]	300	90
[35; 39]	810	210
[40; 44]	1 770	450
[45; 49]	4 060	1 090
[50; 54]	6 760	1 860
[55; 59]	9 050	2 660
[60; 64]	10 440	3 600
[65; 69]	11 300	4 300
[70; 74]	10 850	4 940
[75; 79]	6 870	4 040
[80; 84]	4 730	3 920
[85; 89]	2 600	3 240
[90; 94]	860	1 690
[95; ++]	190	570

TABLE 2.7 – Effectif des nouveaux diagnostics de l'infarctus du myocarde par classe d'âge et par sexe en 2019 en France Métropolitaine

Et nous avons obtenu de même les taux d'incidence par classe d'âge et par sexe :

Nous avons tracé ces différents taux d'incidence :

Nous pouvons donc aisément observer les positions relatives des points d'incidence en fonction du sexe pour ces différents risques, en effet les taux d'incidence masculins sont plus élevés que les taux d'incidence féminins.

Ces écarts peuvent être expliqués par les différences existantes entre eux, notamment physiologiques par exemple les hormones, la morphologie etc. Il existe aussi des différences comportementales entre eux, par rapport aux habitudes de consommation de santé et de mode de vie (alimentation, tabagisme, consommation d'alcool,...)

Ces écarts auront un impact sur la tarification finale.

Tranche d'âge	Homme	Femme
[0; 4]	0,032914244	0,021942829
[5; 9]	0,004957231	0,004957231
[10; 14]	0,0048923828	0,002560560
[15; 19]	0,004906624	0,005187933
[20; 24]	0,021874180	0,016814466
[25; 29]	0,055084491	0,027002783
[30; 34]	0,156459835	0,044587786
[35; 39]	0,403319368	0,100138286
[40; 44]	0,899982458	0,223995046
[45; 49]	1,856280882	0,490355076
[50; 54]	3,190927946	0,849182844
[55; 59]	4,448625989	1,236157131
[60; 64]	5,528920118	1,735585480
[65; 69]	6,267867583	2,130447815
[70; 74]	7,321335790	2,882291644
[75; 79]	7,269995132	3,410966595
[80; 84]	6,353214552	3,588324033
[85; 89]	5,579423088	3,745911580
[90; 94]	4,699941524	3,598838150
[95; ++]	4,359497969	3,385481629

TABLE 2.8 – Taux d'incidence pour 100 000 de l'infarctus du myocarde par classe d'âge et par sexe en 2019 en France Métropolitaine

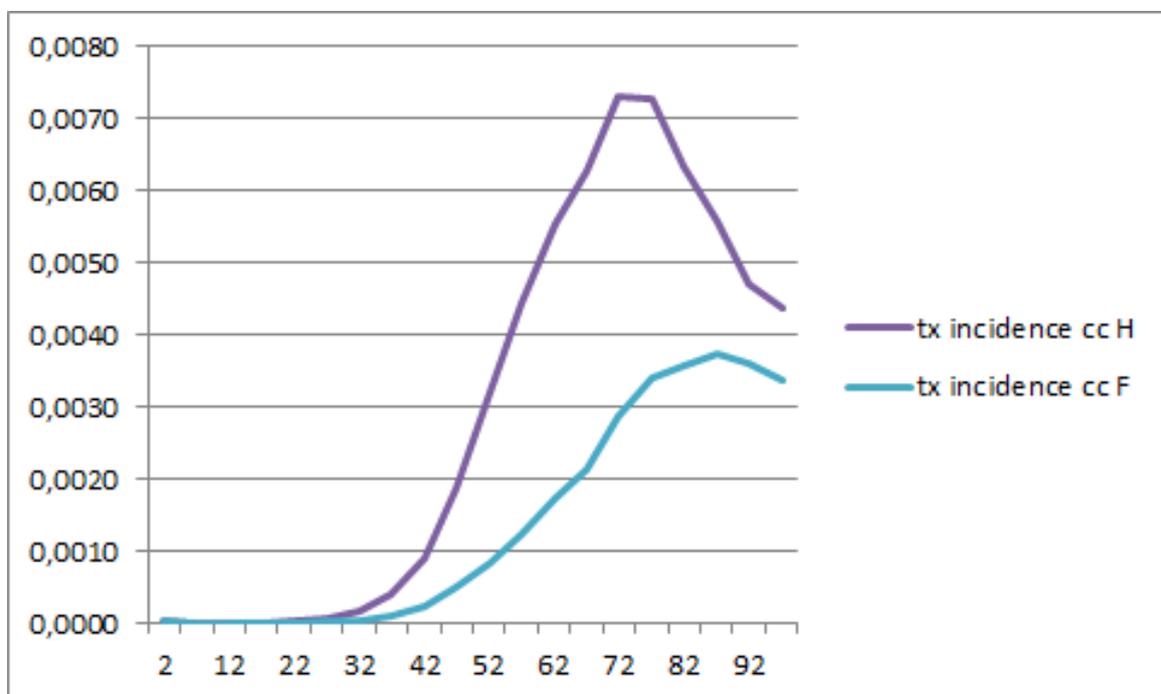


FIGURE 2.3 – Tracés des taux d'incidence de l'infarctus du myocarde en fonction de l'âge et du sexe.

## 2.2 Données de mortalité

Les données de mortalité toutes causes sont celles issues des tables de l'INSEE les plus récentes (2017-2019).

Tranche d'âge	Homme	Femme
moins d'1 an (a)	3,5	2,8
1 à 4 ans	0,3	0,2
5 à 9 ans	0,1	0,1
10 à 14 ans	0,1	0,1
15 à 19 ans	0,3	0,1
20 à 24 ans	0,6	0,2
25 à 29 ans	0,7	0,2
30 à 34 ans	0,8	0,3
35 à 39 ans	1,1	0,5
40 à 44 ans	1,6	0,8
45 à 49 ans	2,6	1,4
50 à 54 ans	4,2	2,3
55 à 59 ans	6,5	3,3
60 à 64 ans	10,4	4,9
65 à 69 ans	14,8	6,9
70 à 79 ans	24,5	12,7
80 à 89 ans	72,0	47,9
90 à 110 ans	215,0	172,0

TABLE 2.9 – Taux de mortalité pour 1 000 par sexe et âge en 2019

Concernant la mortalité due aux cancers, ces informations sont également fournies par l'étude portée sur l'incidence et de la mortalité par cancer en France métropolitaine en 2017. Cette étude nous fournit le nombre de décès par tranche d'âge et par sexe :

Tranche d'âge	Homme	Femme
[0; 14]	96	78
[15; 49]	2 757	2 892
[50; 64]	17 355	11 967
[65; 74]	24 625	14 517
[75; 84]	22 207	16 504
[85; ++]	17 060	20 245

TABLE 2.10 – Nombre de décès liés aus cancers en 2017 en France Métropolitaine par classe d'âge chez l'homme et la femme

Nous en déduisons donc le taux de mortalité des cancers :

Tranche d'âge	Homme	Femme
[0; 14]	0,10446	0,10209
[15; 49]	0,19051	0,11177
[50; 64]	0,30083	0,23643
[65; 74]	0,32726	0,31258
[75; 84]	0,48563	0,45996
[85; ++]	0,85449	0,77782

TABLE 2.11 – Taux de mortalité des cancers par classe d'âge et par sexe en 2017 en France Métropolitaine

Bien que nous ne connaissions pas les situations des individus décédés en 2017 d'un cancer (date d'incidence et durée de la pathologie), ces informations mises en relation avec celles des années passées nous donne une indication de l'impact qu'une telle pathologie a sur l'espérance de vie d'un individu.

Ces données ne sont pas disponibles concernant l'AVC et l'infarctus du myocarde.

Un aspect important à prendre en compte lors de l'utilisation de données publiques comme dans cette étude est l'effet de sélection du risque, en effet en considérant ces tables, on prend en compte l'intégralité de la population or lors d'une souscription d'une telle garantie, il peut y avoir une sélection médicale, permettant de confirmer que la personne ne présente pas un trop gros risque. En cas de souscription d'un contrat collectif, les assurés sont des salariés, ils ont donc un suivi organisé par la Médecine du travail, ce sont donc des populations dont le risque est mesuré de la même manière. Pour tarifier nos garanties au plus juste et pour rester dans le marché, il est donc nécessaire de procéder à un abattement des tables de mortalité. Cet abattement doit être mesuré et cela pourra constituer une étude à poursuivre.



## Chapitre 3

# Tarification d'une garantie Maladies graves et redoutées

### 3.1 Modèle markovien à plusieurs états

Dans cette partie, nous allons proposer une approche de tarification pour les produits *Stand Alone CI* et *Accelerated CI*, basée sur un modèle markovien à plusieurs états. La garantie *Stand Alone CI* indemnise un capital en cas de diagnostic d'une des maladies garanties par le contrat alors que la garantie *Accelerated CI* indemnise une avance du capital garanti en cas de diagnostic d'une des maladies garanties par le contrat et le restant du capital en cas de décès quelle qu'en soit la cause. Une police *Accelerated CI* peut donc être vue comme une garantie *Stand Alone CI* à laquelle on adjoint une garantie décès, [Gogola & Kopecká \(2017\)](#).

#### 3.1.1 Hypothèses du modèle

Nous pouvons donc définir les différents états que nous allons représenter sur la Figure 3.1 de la même manière que [Dash & Grimshaw \(1993\)](#), mais en utilisant un état différent par pathologie :

- $S$  : individu sain
- $C$  : individu diagnostiqué cancéreux
- $D_C$  : décès dû au cancer
- $A$  : individu victime d'un AVC
- $D_A$  : décès dû à un AVC
- $I$  : individu victime d'un infarctus du myocarde
- $D_I$  : décès dû à un infarctus du myocarde

- *D* : décès dû aux autres causes

**Définition 3.1** (Processus stochastique). Un processus stochastique  $Y = Y(t)_{t \in \mathcal{E}}$  est une famille de variables aléatoires  $Y_t$  indexée par un ensemble  $\mathcal{E}$ .

Dans notre étude,  $\mathcal{E} := \{S, C, A, I, D_C, D_A, D_I, D\}$  l'ensemble des états de l'individu assuré. On a bien que  $\{Y(t); t \geq 0\}$  est un processus stochastique en temps continu modélisant l'état de l'assuré.

**Définition 3.2** (Processus de Markov homogène). Un processus stochastique  $(Y_t)_{t \geq 0}$  à valeurs dans  $\mathcal{E}$  est un processus de Markov homogène si :

— (Propriété de Markov) : Pour tout  $t, s \geq 0$  nous avons presque sûrement

$$\mathbb{P}[Y_{t+s} = y | \mathcal{F}_t] = \mathbb{P}[Y_{t+s} = y | Y_t]$$

où  $\mathcal{F}_t := \sigma(X_s, s \in [0, t])$ .

— (Homogénéité) : Pour tout  $s \geq 0$  il existe une matrice  $P_s : E \times E \rightarrow \mathbb{R}$  telle que pour tout  $y \in E$  nous avons presque sûrement

$$\mathbb{P}[Y_{t+s} = y | Y_t] = P_s(Y_t, y)$$

Nous sommes bien ici dans ce cas, en effet, la probabilité d'être dans des états futurs est bien uniquement déterminé par l'état présent. Plus de détails concernant les propriétés markoviennes de ce modèle à plusieurs états sont renseignés dans [Dickson et al. \(2019\)](#).

**Définition 3.3** (Probabilité de transition). Soit  $\mathcal{E} := \{S, C, A, I, D_C, D_A, D_I, D\}$ . Alors pour tout  $i, j \in \mathcal{E}$  :

$${}_t p_x^{ij} = \mathbb{P}(Y(x+t) = i | Y(x) = j).$$

**Définition 3.4.** Soit  $x$  l'âge de l'individu. On définit l'intensité de transition entre l'état  $i$  et  $j$  par

$$\mu_{ij}(x) = \lim_{t \rightarrow 0_+} \frac{{}_tP_x^{ij} - \delta_{ij}}{t}, \quad i \neq j, \quad x \geq 0, \quad i, j \in \mathcal{E}.$$

Où  $\delta_{ij}$  représente le delta de Dirac, définit par :

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$

Dans notre modèle, il n'y aura pas de probabilité de transition entre l'état  $S$  et les états  $D_C$ ,  $D_A$  et  $D_I$  directement car il est nécessaire de transiter par les états  $C$ ,  $A$  et  $I$  dans un premier temps. De plus, comme le contrat se termine après le diagnostic de l'une des trois pathologies garanties, il n'y a pas non plus de transition entre les différents états de maladies graves et redoutées. Enfin, nous ne considérons pas dans notre modèle la guérison des différentes maladies donc la probabilité de transition des états malades à l'état sain est nulle dans notre étude.

Nous pouvons donc exprimer la matrice des probabilités de transition pour un assuré d'un âge  $x$  à un âge  $x + t$ , en considérant les états  $\mathcal{E} := \{S, C, A, I, D_C, D_A, D_I, D\}$ , comme :

$${}_tP_x = \begin{pmatrix} {}_tP_x^{SS} & {}_tP_x^{SC} & {}_tP_x^{SA} & {}_tP_x^{SI} & 0 & 0 & 0 & {}_tP_x^{SD} \\ 0 & {}_tP_x^{CC} & 0 & 0 & {}_tP_x^{CD_C} & 0 & 0 & {}_tP_x^{CD} \\ 0 & 0 & {}_tP_x^{AA} & 0 & 0 & {}_tP_x^{AD_A} & 0 & {}_tP_x^{AD} \\ 0 & 0 & 0 & {}_tP_x^{II} & 0 & 0 & {}_tP_x^{ID_I} & {}_tP_x^{ID} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Nous pouvons alors illustrer cette modélisation de l'état de santé d'un assuré par un schéma représentant les différents états et transitions entre ceux-ci de la sorte :

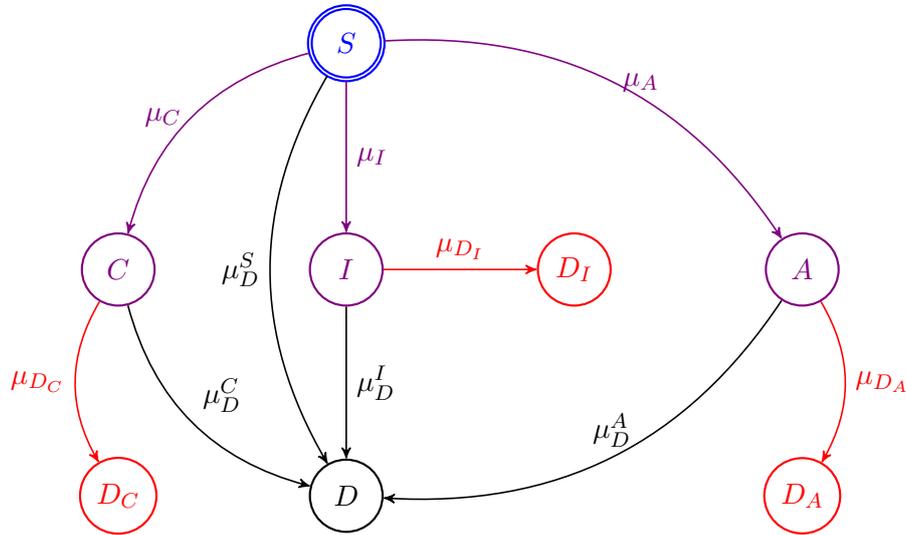


FIGURE 3.1 – Processus de Markov à plusieurs états, en fonction de l'âge

### 3.1.2 Processus de Markov en temps continu

Nous cherchons maintenant à obtenir l'expression des différents coefficients de cette matrice. Nous utilisons les équations de Kolmogorov *forward* et *backward* dans le cadre des processus de Markov qui permettent de relier les probabilités de transitions entre elles :

La probabilité d'aller de  $i$  en  $j$  à un âge  $x$  avec une durée  $t + u$  est notée  ${}_{t+u}p_x^{ij}$ . Pour cela, il est possible de passer par un état intermédiaire  $k$  pour ensuite rejoindre l'état  $j$ , plus formellement :

$$\begin{aligned}
{}_{t+u}p_x^{ij} &= \mathbb{P}(Y_{t+u} = j | Y_0 = i) \\
&= \sum_{k \in \mathcal{E}} \mathbb{P}(Y_{t+u} = j, Y_u = k | Y_0 = i) \\
&= \sum_{k \in \mathcal{E}} \mathbb{P}(Y_{t+u} = j | Y_u = k, Y_0 = i) \mathbb{P}(Y_u = k, Y_0 = i), \text{ d'après la propriété de Markov.} \\
&= \sum_{k \in \mathcal{E}} \mathbb{P}(Y_{t+u} = j | Y_u = k) \mathbb{P}(Y_u = k, Y_0 = i) \\
&= \sum_{k \in \mathcal{E}} \mathbb{P}(Y_t = j | Y_0 = k) \mathbb{P}(Y_u = k, Y_0 = i), \text{ d'après l'homogénéité du processus de Markov.} \\
&= \sum_{k \in \mathcal{E}} {}_t p_x^{kj} {}_u p_x^{ik}
\end{aligned}$$

En y introduisant les intensités de transition, on obtient :

$$\frac{\partial}{\partial t} {}_t p_x^{ij} = \sum_{k \in \mathcal{E}} {}_t p_x^{ik} \cdot \mu_{ij}(x + t)$$

Et on obtient, pour  $x, t \geq 0$  :

$${}_t p_x^{ii} = \exp\left(-\int_0^t \sum_{j \in \mathcal{E}} \mu_{ij}(x + s) \cdot (1 - \delta_{ij}) ds\right) \quad (3.1)$$

En particulier :

$${}_t p_x^{SS} = \exp\left(-\int_0^t [\mu_{SD}(x + s) + \mu_{SC}(x + s) + \mu_{SA}(x + s) + \mu_{SI}(x + s)] ds\right) \quad (3.2)$$

$${}_t p_x^{ii} = \exp\left(-\int_0^t (\mu_{iD_i}(x + s) + \mu_{iD}(x + s)) ds\right), \quad i = C, A, I. \quad (3.3)$$

De la même manière, on peut exprimer les probabilités de transitions entre différents états en fonction des intensités de transition, pour  $x, t \geq 0$  :

$${}_t p_x^{SD} = \int_0^t \mu_{SD}(x+s) \cdot {}_s p_x^{SS} ds \quad (3.4)$$

Et pour les états de maladies graves et redoutées :

$${}_t p_x^{Si} = \int_0^t \mu_{Si}(x+s) \cdot {}_s p_x^{SS} \cdot {}_{t-s} p_{x+s}^{ii} ds, \quad i = C, A, I. \quad (3.5)$$

$${}_t p_x^{iD_i} = \int_0^t \mu_{iD_i}(x+s) \cdot {}_s p_x^{ii} ds, \quad i = C, A, I. \quad (3.6)$$

Avec les équations de Chapman-Kolmogorov nous donnant les mêmes propriétés que pour les chaînes de Markov en temps discret, on peut écrire, pour  $t$  et  $u > 0$  :

$${}_{t+u} p_x^{ij} = \sum_{k \in \mathcal{E}} {}_t p_x^{kj} {}_u p_x^{ik}$$

Et il vient, matriciellement :

$${}_{t+u} P_x = {}_t P_{xu} P_x$$

### 3.1.3 Générateur infinitésimal

La dernière équation écrite peut se réécrire avec plusieurs "pas" :

$$P_{t+u+s} = P_t P_u P_s$$

Et, de plus, en considérant ces "pas" comme réguliers, il vient :

$$P_t = (P_{t/n})^n,$$

Dès lors, la matrice  $P_t$  est totalement déterminée par son comportement sur un intervalle de temps aussi petit soit-il. On définit le générateur infinitésimal comme la dérivée à droite de  $P_t$  en 0, que l'on note  $Q$ . On peut alors exprimer la dérivée de  $P_t$  à tout instant :

$$\begin{aligned} (P_t)' &= (P_{t/n})^n, \\ &= \frac{P_{t+h} - P_t}{h} \\ &= \frac{P_h P_t - P_t}{h} \\ &= \frac{P_h - P_0}{h} P_t, \\ &= \frac{P_h - I}{h} P_t \end{aligned}$$

En effet,  $P_0$  correspond à la matrice identité car à l'instant initial il n'y a pas de changement d'état.

Et en passant à la limite quand  $h$  tend vers 0, il vient :

$$(P_t)' = Q P_t$$

Et de la même manière :  $(P_t)' = P_t Q$ .

Cette équation différentielle matricielle, couplée à la condition initiale  $P_0 = I$  avec  $I$  la matrice identité, nous permet d'exprimer  $P_t$  de manière unique à l'aide de l'opérateur exponentiel pour les matrices :

$$P_t = e^{tQ} = \sum_{n \in \mathbb{N}} \frac{(tQ)^n}{n!}$$

Grace à cette relation, nous pouvons en déduire la matrice de transition sous réserve que  $Q$ , le générateur infinitésimal, soit diagonalisable.

On peut montrer que pour tout  $(x, y) \in \mathcal{E}^2, x \neq y, q(x, y) \geq 0$  et de plus,

$$q_x = q(x, x) = - \sum_{x \neq z} q(x, z)$$

Et de plus, nous pouvons écrire le comportement asymptotique au voisinage de 0, de  $P_t$  à l'aide du générateur infinitésimal :

$$\text{Pour } x \neq y, P_h(x, y) = hq(x, y) + o(h) \text{ et } P_h(x, x) = 1 + hq(x, x) + o(h)$$

### 3.1.4 Loi *phase-type*

Soit un processus de Markov  $Y = (Y_t)_{t \geq 0}$  sur une espace  $\mathcal{E}$  comportant un état absorbant, ici  $D$  le décès (en effet le temps de séjour dans cet état est infini), soit  $q_D = 0$ .

Le générateur infinitésimal de  $Y$  s'écrit de cette manière par bloc :

$$\left( \begin{array}{c|c} T & t \\ \hline 0 & 0 \end{array} \right)$$

Avec  $t$  et  $T$  respectivement le vecteur de taille  $\text{Card}(\mathcal{E})$  et la matrice carrée de dimension  $\text{Card}(\mathcal{E})$ , tels que  $t + T\mathbb{1}_{\mathcal{E}} = 0$  avec  $\mathbb{1}_{\mathcal{E}}$  le vecteur de taille  $\text{Card}(\mathcal{E})$  ne comportant que des 1. Ce qui nous permet de retrouver la propriété énoncée précédemment concernant la nullité de la somme des coefficients par ligne.

En fait le vecteur  $t$  correspond au vecteur des taux de transitions de chaque état de  $\mathcal{E}$  vers l'état absorbant.

On peut alors définir  $\tau$  le temps d'absorption (la durée de survie dans notre cas) comme :

$$\tau = \inf \{t \geq 0; Y_t = D\}$$

Dans ces conditions, la loi de  $\tau$  suit une loi *phase-type*, notée  $PH(\alpha, T)$ , avec  $\alpha$  la loi de  $Y$ , [Asmussen et al. \(2019\)](#).

On peut écrire la densité de la loi de  $\tau$  comme :

$$f_{\tau}(x) = \alpha \exp -(x - \mu)\beta g(\beta)$$

Avec :

$\alpha$  est la probabilité que le patient soit en vie après la première phase, appelé taux de transition initial ;

$\beta$  est la probabilité qu'un patient passe d'une phase  $i$  à la suivante, appelé taux de transition entre les phases ;

$\mu$  est l'espérance du temps de survie du patient, ce paramètre "décale" la loi de la durée de vie ;

$g$  est une densité de probabilité appelée fonction de forme dépendant de  $\beta$

Ce type de loi permet de modéliser la durée de vie des patients. Le nombre de phases représente la période de temps pendant laquelle le patient peut être en vie avant de passer à la phase suivante ou décéder, [Asmussen & Albrecher \(2010\)](#).

Pour pouvoir ajuster ces distributions, il est nécessaire d'obtenir des informations sur les patients atteints de pathologie (surtout le cancer dans notre étude car l'AVC et l'infarctus du myocarde ne présentent, en général, pas de temps de séjour long dans l'état de l'affection.) comme le stade de l'avancée de la pathologie, et la durée de vie depuis le diagnostic.

Nous pouvons ensuite en déduire une estimation des paramètres de cette loi à l'aide de la méthode du maximum de vraisemblance ou celle des moindres carrés. Ensuite, à l'aide des caractéristiques du patient, il est possible de prédire la durée de séjour dans un état donné et sa durée de vie.

### 3.1.5 Méthodologie de tarification

D'une manière très générale, lors de la souscription d'une garantie MGR, que ce soit pour la police *Stand Alone* et *Accelerated*, le capital en cas de diagnostic peut être fonction de la maladie contractée, on peut représenter les issues en fonction de la police et des états de l'assuré :

Nous allons maintenant établir des formules pour calculer la valeur de la prime unique nette pour les produits Maladies Graves et Redoutées grâce à notre modèle markovien à plusieurs états. Pour ce faire, nous avons besoin de définir un taux d'intérêt pour actualiser nos prestations. Pour plus de simplicité, nous le supposerons continu et constant :  $v_t = e^{-\delta t}$  pour  $t \geq 0$ .

Etat	Stand Alone	Accelerated
$S \rightarrow C$	$C_C$	$C_C$
$S \rightarrow A$	$C_A$	$C_A$
$S \rightarrow I$	$C_I$	$C_I$
$S \rightarrow D$	0	$C_D$

TABLE 3.1 – Capital perçu par l'assuré en fonction de sa police et son état

Nous pouvons alors écrire les formules des primes des différentes polices :

**Proposition 3.1** (Prime unique d'une garantie *Stand Alone*). Soit  $i \in \mathcal{I} := \{C, A, I\}$ . Soit  $B_i$  le capital versé dans le cas où l'individu est atteint d'une des maladies pour un contrat de  $n$  années. Alors

$$P_x^{SA} = \sum_{i \in \mathcal{I}} B_i \times \bar{A}_{x:n}^i$$

avec

$$\bar{A}_{x:n}^i = \int_0^n {}_t p_x^{SS} \times \mu_{Si}(x+t) \times e^{-\delta t} dt. \quad (3.7)$$

De même, la prime d'une garantie *Accelerated* s'écrit :

**Proposition 3.2** (Prime unique d'une garantie *Accelerated*). Soit  $i \in \mathcal{I} := \{C, A, I, D\}$ . Soit  $B_i$  le capital versé dans le cas où l'individu est décédé ou atteint d'une des maladies pour un contrat de  $n$  années. Alors

$$P_x^{Acc} = \sum_{i \in \mathcal{I}} B_i \times \bar{A}_{x:n}^i$$

avec, pour  $i \in \mathcal{I} := \{C, A, I\}$ .

$$\bar{A}_{x:n}^i = \int_0^n {}_t p_x^{SS} \times \mu_{Si}(x+t) \times e^{-\delta t} dt. \quad (3.8)$$

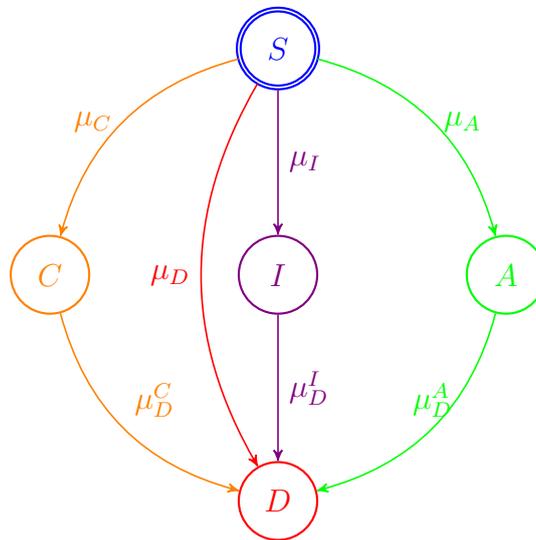
et,

$$\bar{A}_{x:n}^D = \int_0^n {}_t p_x^{SS} \times \mu_{SD}(x+t) \times e^{-\delta t} dt. \quad (3.9)$$

### 3.1.6 Résultats obtenus

Pour appliquer ce que nous avons vu ci-dessus, nous allons tarifer un contrat *Stand Alone* ainsi qu'un *Accelerated* garantissant un capital de 10 000€ en cas de diagnostic de l'une des trois pathologies étudiées dans le chapitre 1 et ce même capital en cas de décès pour la garantie *Accelerated*.

Cependant en l'absence de données concernant la mortalité par cause (notamment les décès ayant pour cause l'AVC et la crise cardiaque), nous avons considéré un état "Décès" unique et non conditionné à sa cause, de cette manière nous pouvons représenter le schéma du modèle markovien sur lequel est basé notre étude :



Si on considère un contrat sur  $n$  années d'un assuré d'âge  $x$  garantissant 1€ en cas de diagnostic d'un cancer, l'expression de la prime devient :

$$P_x^{SA} = \int_0^n {}_t p_x^{SS} \times (\mu_{SC}(x+t) + \mu_{SIM}(x+t) + \mu_S(x+t)) \times e^{-\delta t} dt \quad (3.10)$$

### Modélisation de la probabilité de rester dans l'état Sain

Il nous faut donc modéliser  ${}_t p_x^{SS}$ , la probabilité de rester dans l'état sain. Pour ce faire, nous avons utilisé la table de l'INSEE donnant le nombre de survivants par âge 3.6 En toute rigueur il faudrait regarder non seulement la survie mais aussi le nombre de personnes diagnostiquées maladie grave et redoutées car ils ne sont plus dans l'état sain  $S$ , cependant cet effet peut être négligé selon Booth & Tickle (2008) et était complexe à implémenter en raison de la non concordance des données entre elles (les tranches d'âge étant différentes cela entraîne une perte significative d'informations), mais ce travail pourrait être affiné pour plus de précision.

Il nous faut donc modéliser la probabilité conditionnelle  ${}_t p_x^{SS} = \mathbb{P}(Y(x+t) = H \mid Y(x) = H)$  utilisée dans l'équation (3.9). En utilisant les tables de mortalité disponibles sur le site de l'INSEE (periode : 2017 - 2019, champ : France hors Mayotte), nous avons pu reconstruire empiriquement la fonction de survie  $S(x) = \mathbb{P}(T \geq x)$ .

**Proposition 3.3** (Loi de Weibull). Soit  $X$  une variable aléatoire à densité qui suit une loi de Weibull de paramètre d'échelle  $\lambda$  et de paramètre de forme  $\beta$  ( $X \sim \text{Weibull}(\lambda, \beta)$ ). Alors sa fonction de répartition est définie par

$$F(x; \beta, \lambda) = 1 - e^{-(x/\lambda)^\beta}.$$

Par dérivation, sa densité de probabilité est

$$f(x; \beta, \lambda) = \frac{\beta}{\lambda} \left(\frac{x}{\lambda}\right)^{\beta-1} e^{-(x/\lambda)^\beta}.$$

**Proposition 3.4** (Simulation d'une variable aléatoire). Soit  $X$  une variable aléatoire de fonction de répartition  $F(x) = \mathbb{P}(X \leq x)$ . Soit son inverse généralisé (fonction quantile)  $F^{-1}(u) := \inf\{x : F(x) \geq u\}$ . Alors

$$F^{-1}(U) \stackrel{\text{loi}}{=} X$$

avec  $U \sim \mathcal{U}([0, 1])$ .

Pour avoir une expression explicite, nous ajustons une loi de Weibull sur nos données. Pour ce faire, nous utiliserons la proposition 3.4 afin de simuler des tirages suivant la fonction de répartition  $F := 1 - S$ .

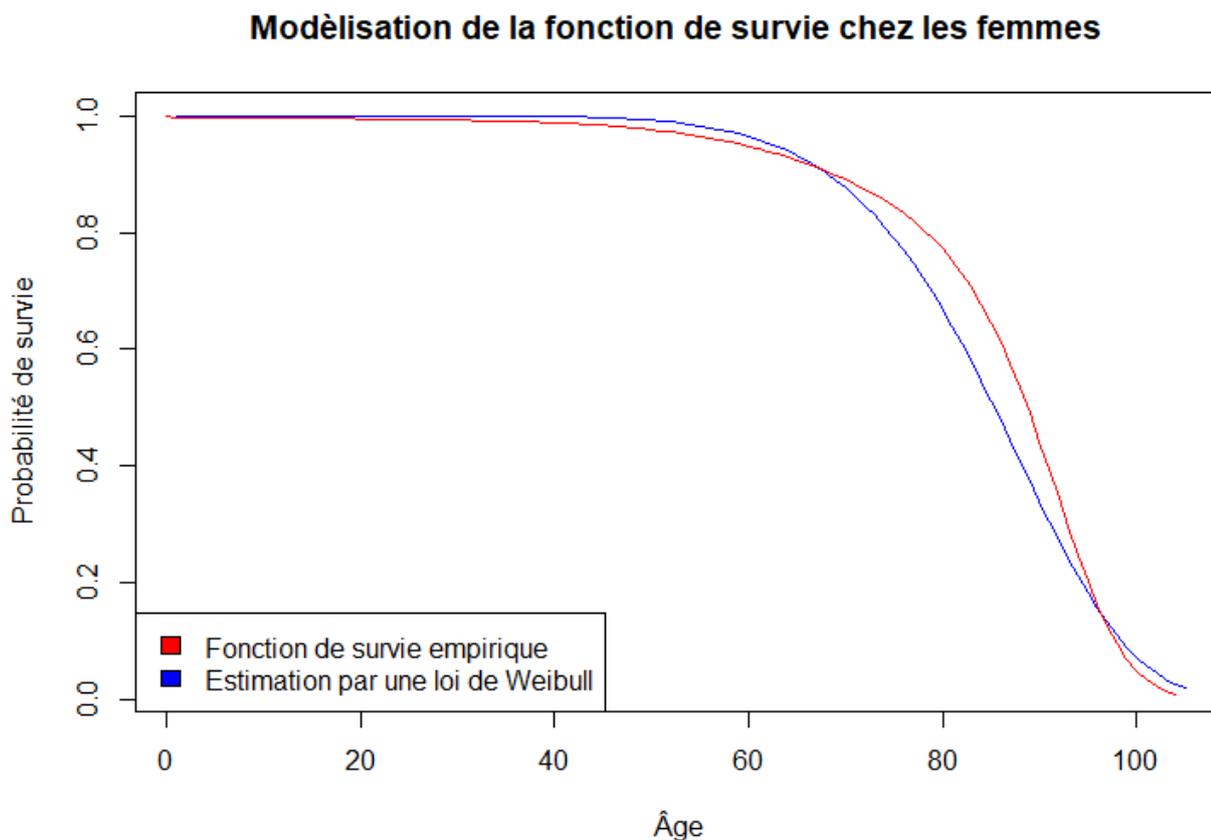


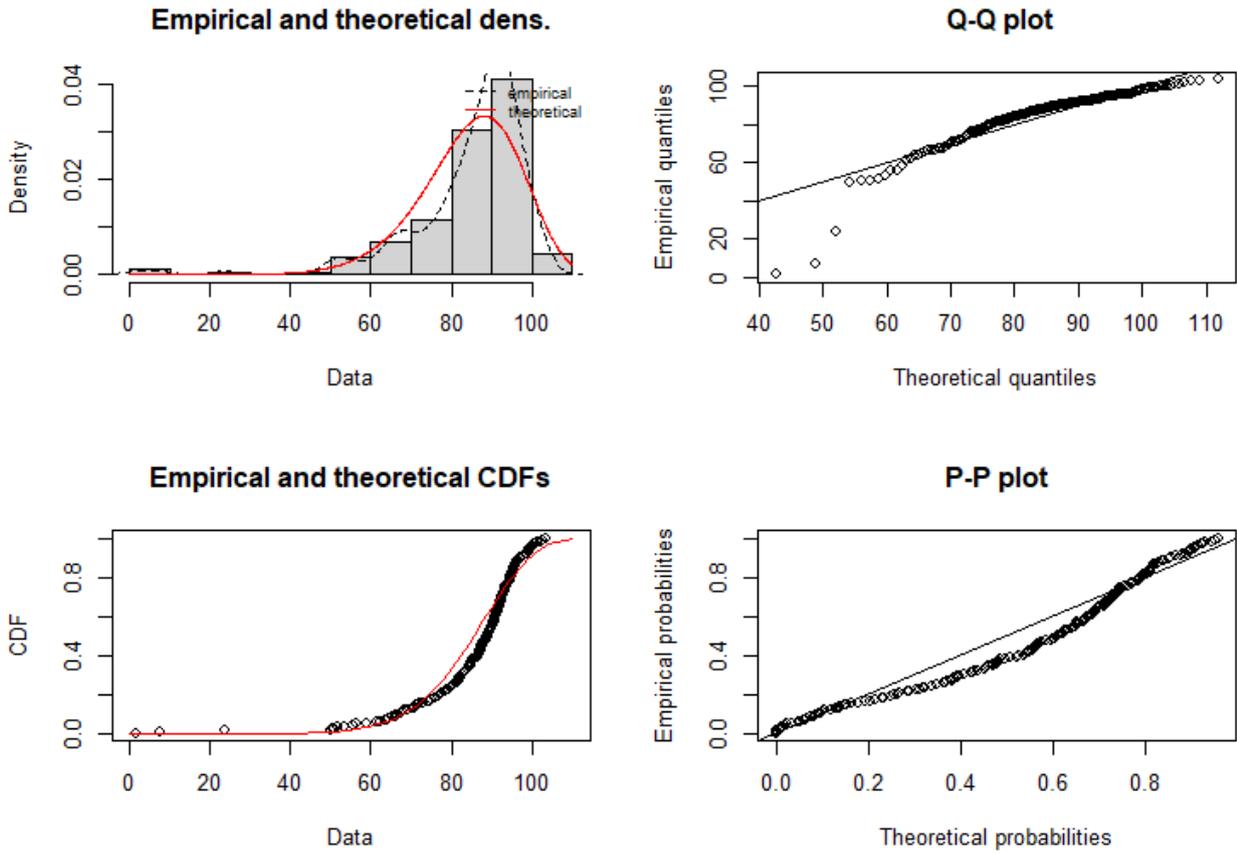
FIGURE 3.2 – La fonction de survie empirique et la fonction de survie de la loi de Weibull ajustée à nos données pour le sexe féminin.

Concernant la fonction empirique, il s'agit d'un tracé linéaire entre chaque point empirique. Comme nous avons un point par année, cette fonction est très fidèle à la réalité et correspond à l'hypothèse classique du taux de décès uniforme au sein d'une même année. Cette représentation nous permet une intégration très facile numériquement, il s'agit d'une somme d'aires de trapèzes.

Grace à la bibliothèque `fitdistrplus` dans R , nous avons obtenu les estimation des paramètres de la loi de Weibull : le paramètre d'échelle  $\lambda \approx 89,61409$  et le paramètre de forme  $\beta \approx 8,058507$ , [Wilson \(1994\)](#).

La figure 3.2 montre l'évolution empirique et paramétrique de  ${}_t p_x$  pour  $x = 0$ . Nous avons tracé les graphiques permettant d'analyser l'ajustement de la courbe de Weibull à nos données :

On voit donc grâce à ces graphiques que l'ajustement de la courbe de Weibull à nos données est relativement précise : le *Q-Q plot* semble linéaire sauf pour les premières données ce qui est



cohérent car à ces âges la mortalité est très faible donc les quantiles aux premiers âges ne sont pas parfaitement estimés. Et concernant le *P-P plot*, on constate que les probabilités de survie sont un peu surestimées, nous en tiendrons compte par la suite.

Trivialement on peut déduire que

$$\forall x > 0, \forall t \geq 0 \quad {}_t p_x^{SS} = \frac{{}_{t+x} p_0^{SS}}{{}_t p_0^{SS}}.$$

Comme

$${}_t p_0 = S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^\beta\right)$$

avec  $S$  la fonction de survie d'une loi de Weibull, on a

$$\begin{aligned} {}_t p_x &= \frac{{}_{t+x} p_0}{{}_t p_0} = \frac{\exp\left(-\left(\frac{t+x}{\lambda}\right)^\beta\right)}{\exp\left(-\left(\frac{t}{\lambda}\right)^\beta\right)} \\ &= \exp\left(-\lambda^{-\beta} \cdot (t^\beta - (t+x)^\beta)\right) \end{aligned}$$

De même pour les hommes :

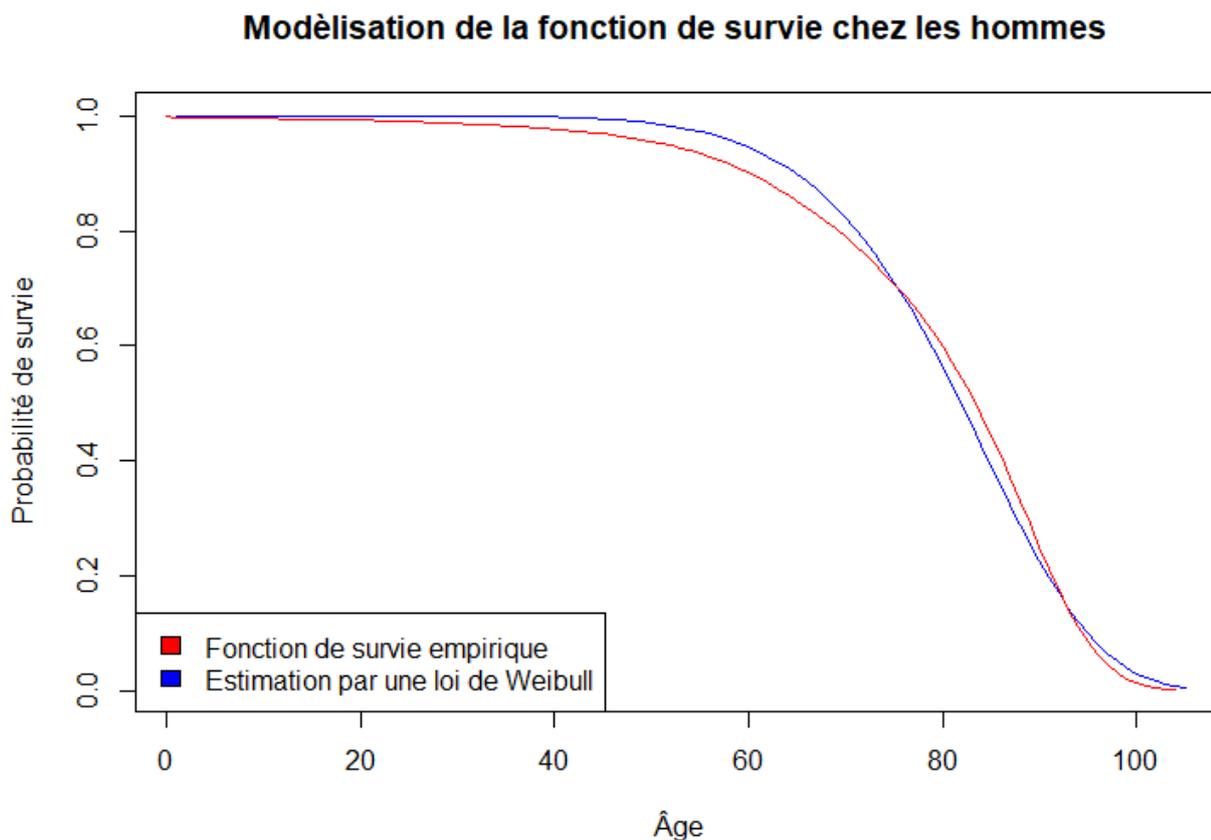


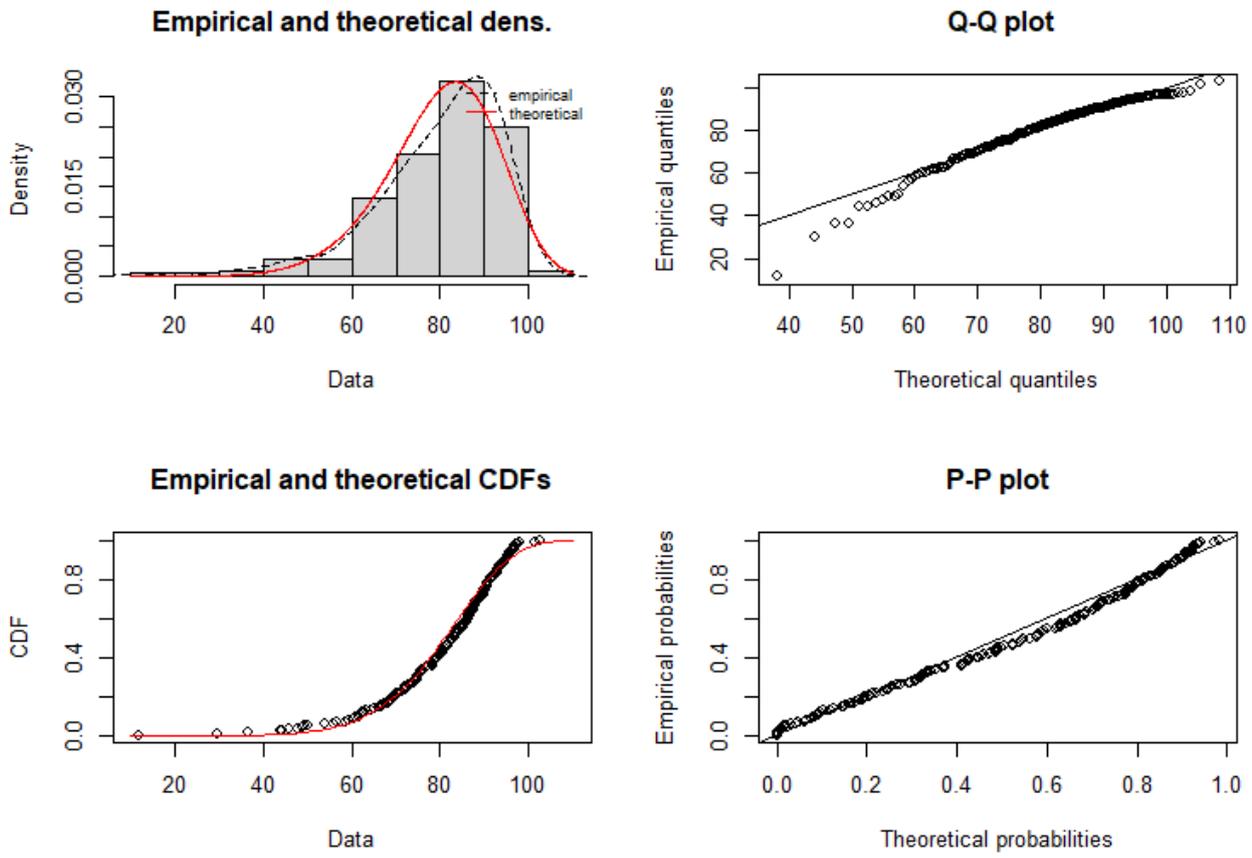
FIGURE 3.3 – La fonction de survie empirique et la fonction de survie de la loi de Weibull ajustée à nos données pour le sexe masculin.

De la même manière, nous avons obtenu les estimations des paramètres de la loi de Weibull : le paramètre d'échelle  $\lambda \approx 85,71229$  et le paramètre de forme  $\beta \approx 8,08158$ .

La figure 3.3 montre l'évolution empirique et paramétrique de  ${}_t p_x$  pour  $x = 0$ .

Nous avons tracé les graphiques permettant d'analyser l'ajustement de la courbe de Weibull à nos données :

Ici, en comparaison avec la modélisation réalisée pour les femmes, nous constatons un ajustement plus précis : le *Q-Q plot* est bien linéaire à partir de 60 ans et le *P-P plot* montre une surestimation beaucoup moins prononcée des probabilités de survie qu'il n'en était pour les femmes.



### Modélisation du taux de transition de l'état sain vers les différentes pathologies

Dans l'équation (3.12) pour pouvoir obtenir un tarif, nous cherchons à modéliser  $\mu_{SC}, \mu_{SI}$  et  $\mu_{SA}$  définis dans 3.4. En utilisant les données de l'étude sur les incidences et la mortalité des cancers (2.1), nous avons effectué une régression polynomiale de degré 6 sans intercept afin de modéliser le taux de l'incidence en fonction de l'âge de l'individu. En effet, pour une modélisation la plus précise, nous avons supposé que le taux d'incidence des pathologies est supposé nul à la naissance, en effet cela n'aura pas d'impact sur le tarif car nous ne tarifierons pas de garanties secondaires comme par exemple un capital en cas de diagnostic d'un cancer pour un ayant droit qui, dans ce cas, pourrait être un nouveau-né. Nous tarifierons dans notre cas des garanties sur une population d'au moins 30 ans. Pour les autres points empiriques, nous avons déterminé un âge en pondérant par les effectifs des différents âges au sein d'une même classe d'âge.

Nous avons aussi représenté des fonctions en escalier pour les trois risques avec les taux d'incidence par classe d'âge déterminés permettant d'approcher notre tarif d'une différente

manière.

En observant les courbes 3.4 et 3.5, on se rend compte que cette approche est très peu précise mais peut constituer une aide au tarif en tenant compte du fait que, notamment pour le cancer où nous n'avons que peu de tranches d'âge, un saut important est présent vers 45 ans ce qui va augmenter le tarif d'une manière importante.

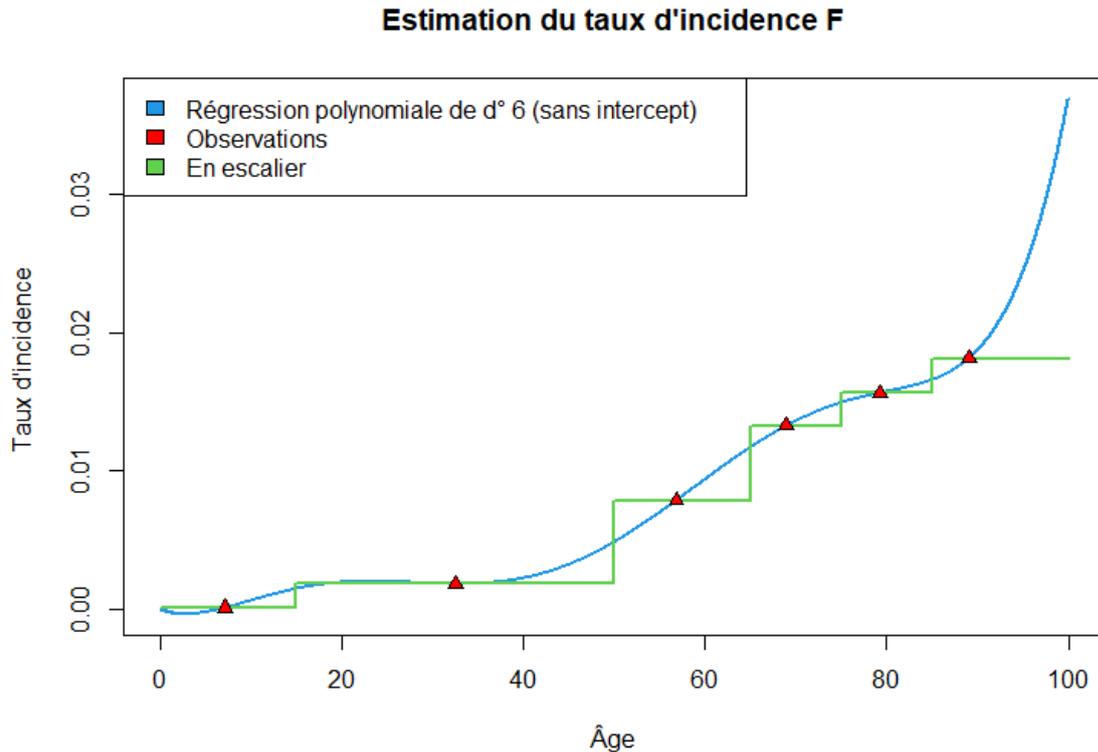


FIGURE 3.4 – Tracés des différents modèles de l'incidence du cancer pour le sexe féminin.

Nous pouvons constater une différence significative entre les deux sexes, pour le sexe féminin, la regression polynomiale devient largement croissante et cela risque d'impacter le tarif notamment pour des contrats viager.

Nous avons donc choisi de rajouter artificiellement un point à l'âge 100 afin de ne pas tendre vers un tarif prohibitif. Pour la valeur de ce point, après plusieurs essais, nous avons opté pour la valeur moyenne des deux derniers taux. Dans les faits, l'incidence du cancer baisse à ces âges avancés mais reste élevée et en multipliant cette estimation par la probabilité très faible de survie à ces âges, l'hypothèse réalisée a peu d'impact sur la tarification.

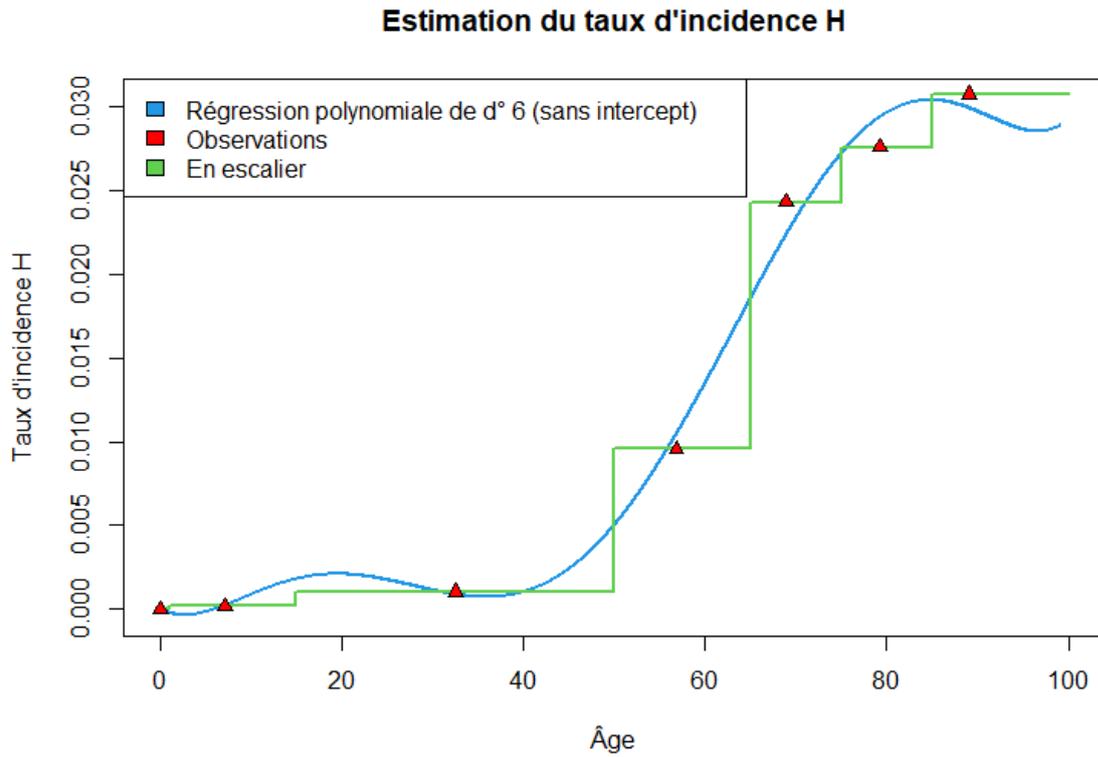


FIGURE 3.5 – Tracés des différents modèles de l'incidence du cancer pour le sexe masculin.

Nous constatons que l'ajout de ce point a permis d'obtenir une régression dont les valeurs aux âges les plus avancés ne dépassent pas la valeur en 90 ce qui paraît plus cohérent que précédemment.

Cette régression polynomiale nous permet de obtenir une expression explicite du taux d'incidence en fonction de l'âge de l'individu :

Soit  $x$  l'âge de l'individu et soit  $\mu_{SC}(x)$  le taux d'incidence. Par la régression polynomiale, nous avons obtenu :

$$\mu_{SC}(x) \approx \sum_{i=1}^6 \alpha_i x^i \quad (3.11)$$

— Pour les femmes :

$$\alpha = (1,76 \cdot 10^{-5}; 8,77 \cdot 10^{-7}; -6,31 \cdot 10^{-8}; 3,38 \cdot 10^{-9}; -4,02 \cdot 10^{-11}; 1,34 \cdot 10^{-13})$$

— Pour les hommes :

$$\alpha = (-2,95 \cdot 10^{-4}; 7,05 \cdot 10^{-5}; -4,08 \cdot 10^{-6}; 9,50 \cdot 10^{-8}; -9,31 \cdot 10^{-10}; 3,24 \cdot 10^{-12})$$

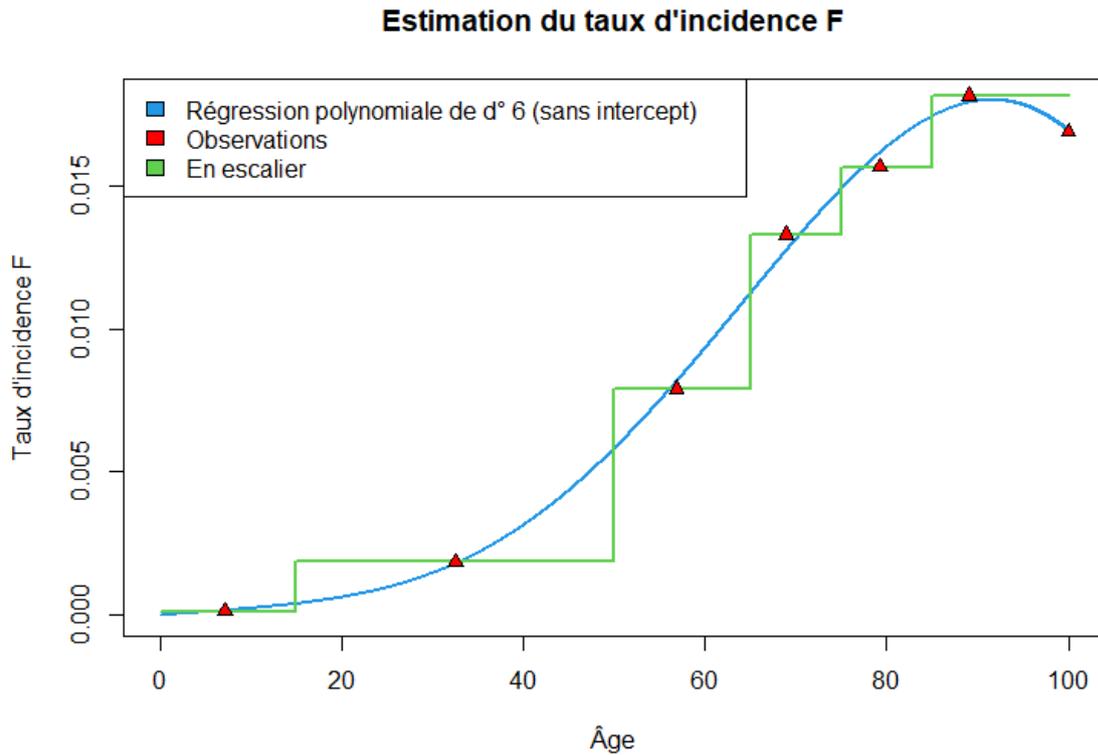


FIGURE 3.6 – Tracés des différents modèles de l'incidence du cancer pour le sexe féminin en ajoutant un point à l'âge 100.

Nous avons donc une expression explicite de l'estimation de  $\mu_{SC}$  en fonction de l'âge de l'individu.

De la même manière, nous avons modélisé les incidences de l'infarctus du myocarde et de l'AVC en fonction de l'âge pour les deux sexes.

De la même manière que pour le cancer, la régression polynomiale obtenue nous permet d'obtenir une expression explicite des courbes modélisant les différentes incidences que nous recherchons. Ainsi, nous allons pouvoir dans la suite intégrer ces expressions pour en déduire la valeur des primes correspondant à ces garanties.

Nous constatons que contrairement au cancer, nous avons plus de tranches d'âge différentes concernant les incidences d'infarctus du myocarde et de l'AVC. Cela va donc nous donner un tarif plus lisse en fonction de l'âge car ces fonctions en escalier possèdent plus d'informations et nous n'aurons pas les sauts que nous allons sûrement observer pour une garantie s'appliquant

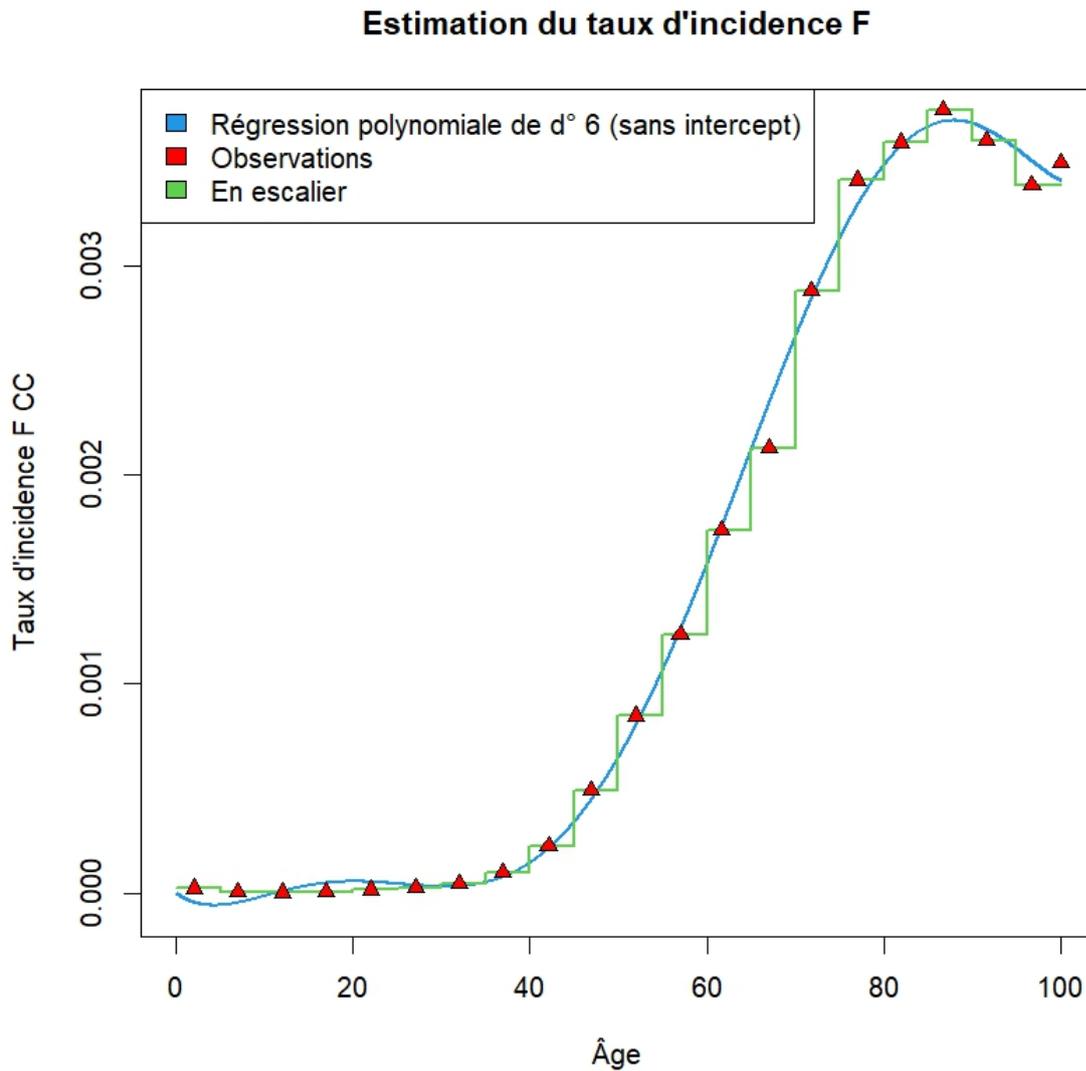


FIGURE 3.7 – Tracés des différents modèles de l'incidence de l'infarctus du myocarde pour le sexe féminin

uniquement aux cancers.

Nous pouvons désormais passer au calcul de la prime selon les modélisations choisies.

### Comparaisons des primes obtenues

Dans cette étude, nous avons utilisé un modèle markovien à plusieurs états pour pouvoir calculer la prime de contrats "classiques" de maladies graves et redoutées. Nous avons décrit les états possibles d'un assuré et nous avons estimé les probabilités et les intensités de transition entre un état non absorbant et un autre ou bien de rester dans ce même état.

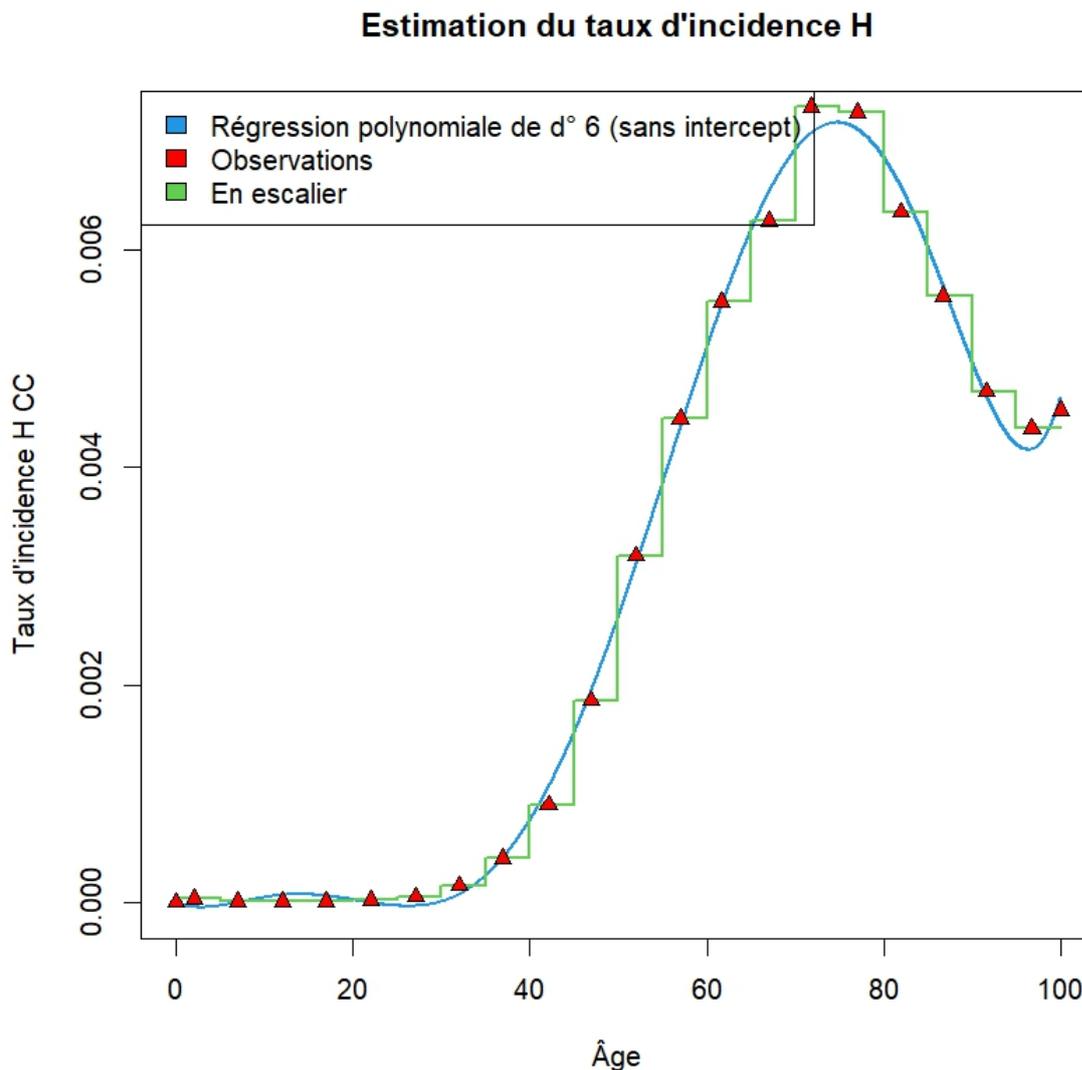


FIGURE 3.8 – Tracés des différents modèles de l'incidence de l'infarctus du myocarde pour le sexe masculin

Dans ce modèle inspiré de [Baione & Levantesi \(2014\)](#), nous ne considérons pas un état *Critical Illness*, mais plusieurs en fonction du nombre de maladies garanties, cela nous permet d'allouer un capital différent en fonction de la pathologie et nous pouvons utiliser plus de données pour déterminer les différentes probabilités de transition.

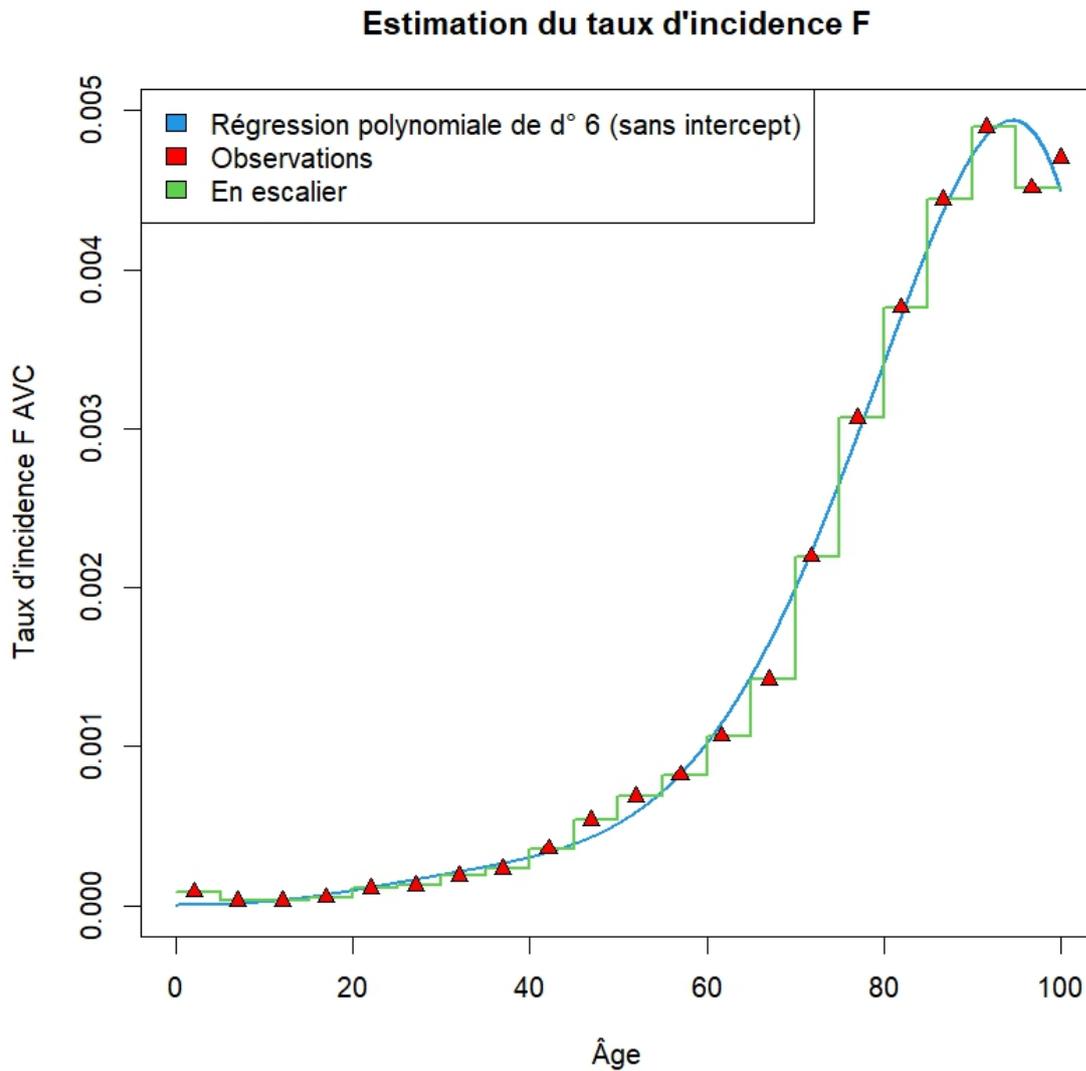


FIGURE 3.9 – Tracés des différents modèles de l'incidence de l'AVC pour le sexe féminin

Nous allons maintenant présenter les différentes primes obtenues en fonction de la méthode utilisée et de la police choisie. Concernant les méthodes, nous en avons mis quatre en place : la probabilité de rester dans l'état Sain,  ${}_t p_x^{SS}$ , a été modélisée de deux manières :

- Dans un premier temps, nous avons ajusté une loi de Weibull sur nos données empiriques, permettant d'obtenir une expression explicite que nous pouvons intégrer.
- Dans un second temps, nous avons tracé une fonction empirique en reliant nos différents points empiriques.

De plus, le taux d'incidence des pathologies a lui aussi été modélisé de deux manières

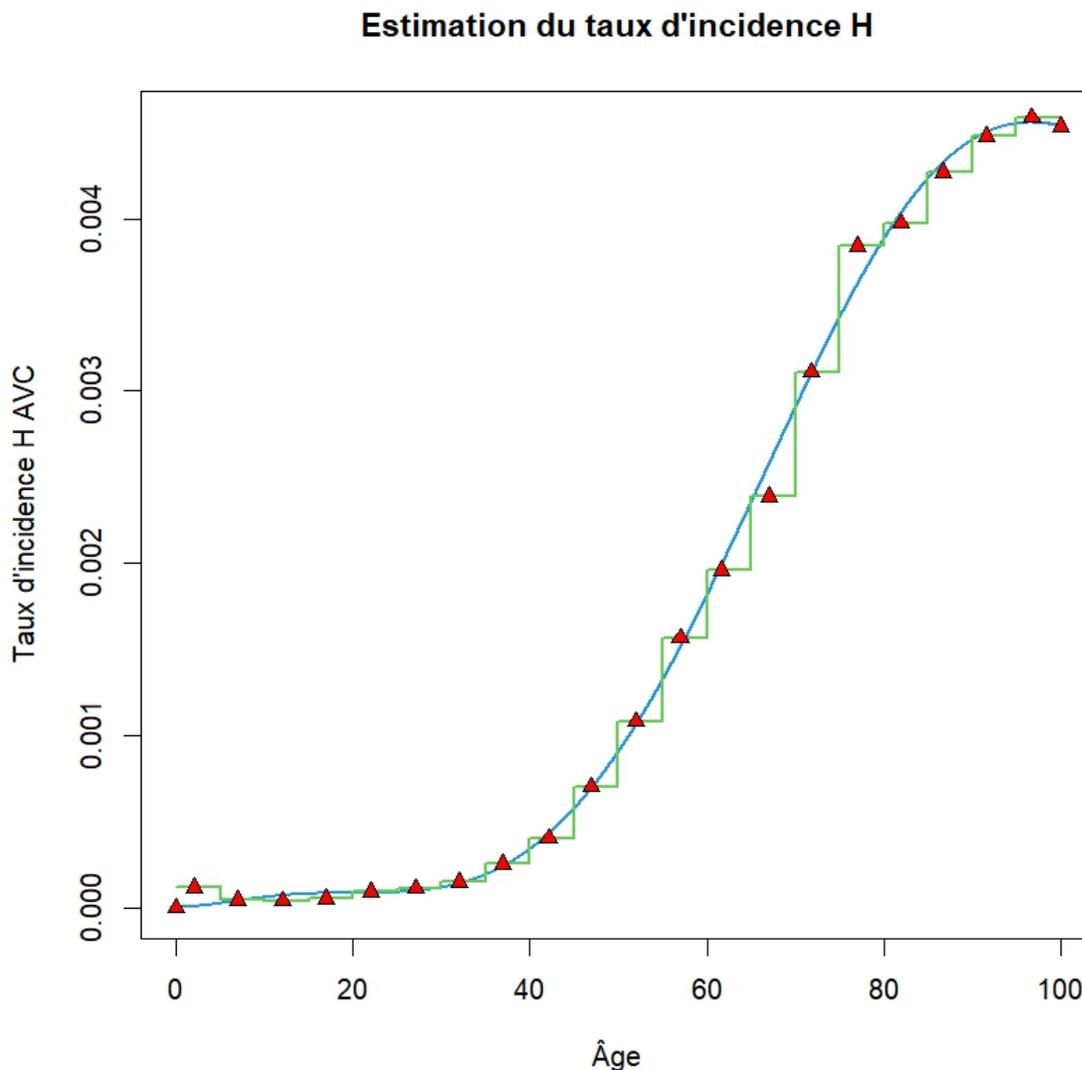


FIGURE 3.10 – Tracés des différents modèles de l'incidence de l'AVC pour le sexe masculin

différentes :

- Dans un premier temps, nous avons effectué une régression polynomiale de degré 6 sur nos données empiriques, permettant d'obtenir de même une expression explicite que nous pouvons intégrer.
- Dans un second temps, nous avons tracé une fonction empirique en escalier en associant à chaque âge à l'intérieur d'une même tranche d'âge le taux d'incidence correspondant à cette tranche d'âge.

Dans cette étude, nous allons considérer des policeS *Stand Alone* garantissant le capital de 10

000€ en cas de diagnostic d'une des trois pathologies garanties donc les cancers (dans la limite de ceux garantis : ici nous n'avons exclu que le mélanome cutané, c'est une pratique courante concernant ces garanties) quels qu'ils soient, l'infarctus du myocarde et l'AVC, et ce jusqu'à la centième année de l'assuré. Le taux d'intérêt continu :  $v_t = e^{-\delta t}$  pour  $t \geq 0$  est supposé constant avec  $\delta = 5\%$  ce qui correspond à peu près à l'environnement actuel de remontée des taux.

En reprenant la formule de la prime :

$$P_x^{SA} = \int_0^n {}_t p_x^{SS} \times (\mu_{SC}(x+t) + \mu_{SA}(x+t) + \mu_{SI}(x+t)) \times e^{-\delta t} dt \quad (3.12)$$

Nous pouvons, dans un premier temps, en déduire les primes nettes uniques d'un contrat garantissant uniquement les cancers et dans un second temps ajouter les infarctus du myocarde et les AVC, car le format de données étant différent dû aux tranches d'âge, il sera plus aisé d'analyser les résultats de cette manière.

TABLE 3.2 – Primes nettes uniques en fonction de l'âge de souscription pour une garantie cancer

age	méthode 1		méthode 2		méthode 3		méthode 4	
	H	F	H	F	H	F	H	F
35	123,57	396,56	125,87	395,00	166,55	313,10	158,74	334,85
40	167,33	553,36	172,58	552,27	174,55	324,34	169,25	324,98
45	366,54	736,90	384,58	738,56	181,12	333,54	177,80	312,40
50	704,03	933,21	757,77	943,49	1234,84	1132,14	1045,90	1073,04
55	1109,36	1120,00	1239,73	1150,58	1344,61	1212,48	1203,21	1180,26
60	1475,70	1267,47	1742,22	1337,29	1442,98	1280,40	1340,67	1266,46
65	1692,72	1341,77	2163,73	1475,53	2836,91	1745,19	2219,35	1586,98

Concernant les deux premières méthodes, c'est-à-dire, en utilisant la regression polynomiale pour la probabilité de transition et en modélisant la probabilité  ${}_t p_x^{SS}$  avec une loi de Weibull ou une approche empirique, on constate que les résultats obtenus sont proches, la méthode 2 (approche empirique de la probabilité de rester dans l'état sain) donne des prix un peu plus élevés que la méthode 1 (approche par la loi de Weibull) ce qui est en accord avec l'analyse des *P-P plot* de la 3.2, qui montrait une légère surestimation de la probabilité de survie.

En comparant les primes par sexe, on constate un écart assez important entre les deux, cela s'explique par l'écart que nous avons observé entre les taux d'incidence des cancers.

Dans un second temps, nous pouvons remarquer que les résultats des méthodes 3 et 4 ne sont pas satisfaisants : nous constatons un écart de tarif trop important entre les âges 45 et 50

et au contraire un écart trop faible entre les âges 35, 40 et 45. En effet cela s'explique par la modélisation en escalier, en fonction de la position de la tranche d'âge par rapport aux différents sauts de la fonction en escalier, le tarif n'évolue que peu ou trop. Nous voyons que le saut que nous avons mentionné entre 45 et 50 ans a ici un impact non négligeable sur le tarif : en effet entre 45 et 50 la prime pure augmente de près de 700% !

Cette méthode nous donne des informations mais elle ne peut pas servir à commercialiser un tarif en l'état pour des problèmes d'antisélection. Il faut donc utiliser une méthode plus "lisse" comme les deux premières.

Nous allons maintenant intégrer les risques liés à l'infarctus du myocarde et de l'AVC, en espérant que concernant les méthodes 3 et 4 nous aurons des primes un peu plus lissées en fonction de l'âge.

Nous nous plaçons toujours dans le cadre d'une garantie *Stand Alone* garantissant un capital de 10 000€ en cas de diagnostic de l'une des trois pathologies étudiées.

TABLE 3.3 – Primes nettes uniques en fonction de l'âge de souscription pour une garantie *Stand Alone*

age	méthode 1		méthode 2		méthode 3		méthode 4	
	H	F	H	F	H	F	H	F
35	198,51	449,81	202,21	448,05	295,35	392,94	289,94	394,49
40	345,38	632,05	356,20	630,81	392,57	426,59	380,64	427,44
45	691,74	859,52	725,79	861,46	575,01	486,94	548,03	485,85
50	1198,24	1119,74	1289,71	1132,07	1993,19	1385,64	1569,81	1260,41
55	1759,77	1386,36	1966,57	1424,21	2190,26	1528,72	1959,93	1488,09
60	2230,92	1618,84	2633,84	1708,01	2295,46	1677,49	2234,09	1706,00
65	2471,01	1766,10	3158,60	1942,16	3845,93	2211,68	3008,72	2011,18

Nous constatons ici, que similairement à la garantie cancer, les méthodes 1 et 2 sont très similaires en terme de résultats : on observe pour les deux le même écart entre les deux sexes que ce soit à la hausse ou à la baisse. Pour les jeunes âges, le tarif est beaucoup plus élevé pour les femmes que pour les hommes, ceci peut s'expliquer par la plus forte mortalité des hommes à ces âges donc le  ${}_t p_x^{SS}$  est alors plus important pour les femmes que pour les hommes ce qui tend à expliquer cet écart. Et à l'inverse, aux âges plus avancés, ce sont les hommes qui sont tarifés plus cher ce qui est dû aux taux d'incidence tous plus élevés pour les hommes que pour les femmes concernant les pathologies pour une mortalité similaire.

Sur ces deux graphiques, la mention  $f_i$  et  $h_i$  correspond à la méthode de calcul utilisée. Nous pouvons donc bien constater sur ces différents graphiques que les couples de méthodes 1;2 et

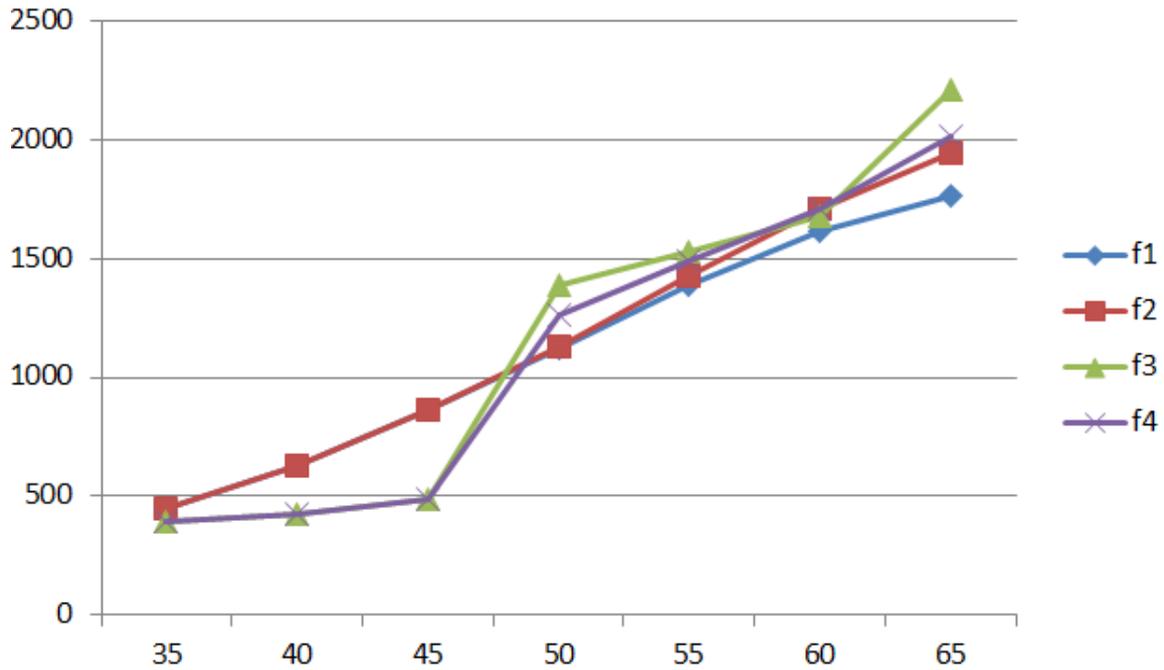


FIGURE 3.11 – Tracés des différentes Primes Unique Nettes en fonction de l’âge de souscription pour le sexe féminin

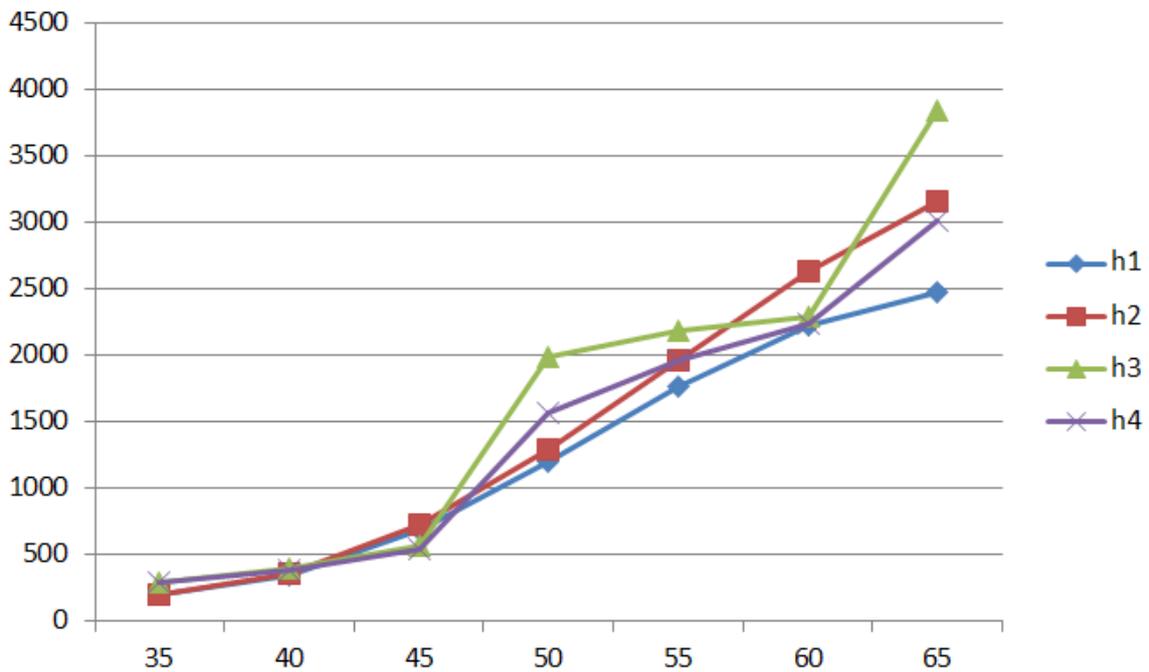


FIGURE 3.12 – Tracés des différentes Primes Unique Nettes en fonction de l’âge de souscription pour le sexe masculin

3;4 sont similaires entre eux. Il semblerait que ce soit la modélisation des taux d’incidence qui ait le plus d’impact en comparaison avec la modélisation de la survie des individus.

Les méthodes sont cohérentes entre elles : on observe les mêmes tendances pour les deux sexes. Les deux premières méthodes semblent tout de même à privilégier, en effet les différences de prix par âge sont plus régulières ce qui tend à rendre le tarif plus attractif tout en se protégeant contre l'anti-sélection.

De plus, sur un portefeuille ayant des assurés d'âge et de sexe régulièrement distribués, la mutualisation entre-eux sera de meilleure qualité.

### Impact du taux d'intérêt

Après l'analyse des primes selon différentes approches, nous étudions dans cette partie l'impact des taux d'intérêt sur les primes des polices d'assurance. Dans l'hypothèse de base, une force d'intérêt constante de 5% est utilisée pour la tarification.

Nous testons des scénarios dans lesquels la force d'intérêt constante ( $= 5\%$ ) est augmentée et diminuée, respectivement de 10% et de 20% dans le scénario de base. Cette étude semble pertinente compte tenu du contexte actuel de remontée et d'évolution des taux, les impacts tarifaires de ceux-ci ne sont pas toujours évidents à mesurer en prévoyance.

D'après l'analyse ci-dessous et étant donné les similitudes dans l'intégration du taux d'intérêt, nous étudierons cet impact sur la première méthode uniquement.

TABLE 3.4 – Evolution du tarif en fonction de l'évolution du taux d'intérêt pour le sexe masculin

	-20%	-10%	10%	20%
35	16,76%	7,84%	-6,93%	-13,08%
40	15,63%	7,35%	-6,55%	-12,41%
45	14,43%	6,82%	-6,13%	-11,66%
50	13,18%	6,26%	-5,68%	-10,84%
55	11,87%	5,67%	-5,19%	-9,95%
60	10,55%	5,06%	-4,67%	-9,00%
65	9,22%	4,45%	-4,15%	-8,02%

TABLE 3.5 – Evolution du tarif en fonction de l'évolution du taux d'intérêt pour le sexe féminin

	-20	-10	10	20
35	17,74%	8,27%	-7,26%	-13,66%
40	16,69%	7,82%	-6,92%	-13,06%
45	15,55%	7,32%	-6,53%	-12,38%
50	14,32%	6,77%	-6,10%	-11,61%
55	13,01%	6,19%	-5,62%	-10,75%
60	11,64%	5,56%	-5,11%	-9,80%
65	10,24%	4,92%	-4,56%	-8,79%

D'après les résultats présentés dans les tableaux 3.4 et 3.5, nous constatons qu'avec une baisse de 20 % du taux d'intérêt, les taux de prime augmentent d'environ 14 % en général pour toutes les polices, tant pour les hommes que pour les femmes, et pour les âges de souscription de 35 à 55 ans ; tandis que les taux de prime augmentent d'environ 6,5 % avec une baisse de 10 % du taux d'intérêt. En comparaison avec une baisse du taux d'intérêt, si le taux d'intérêt

augmente de 10 %, les taux de prime diminuent d'environ 6 %, ce qui est légèrement inférieur à l'augmentation observée dans le scénario de "-10 %". De même, les taux de prime diminuent d'environ 12 % avec une augmentation de 20 % du taux d'intérêt. De plus, il est observé que les taux de prime pour les hommes ont changé davantage que ceux pour les femmes aux âges plus jeunes pour tous les types de polices.

## 3.2 Modèle bayésien avec lois bêta binomiale

Ce paragraphe a pour but d'expliquer et d'appliquer les méthodes d'inférence statistique classique et bayésienne pour estimer la probabilité des diagnostics des différentes maladies redoutées pour les hommes et les femmes et pour différentes classes d'âge. Les probabilités d'événements estimées sont ensuite utilisées pour fixer les primes de risque dans des groupes homogènes par sexe et par âge. Et ainsi ajuster le tarif déterminé dans la première partie à l'aide des résultats obtenus par l'assureur.

### 3.2.1 Hypothèses du modèle

Considérons un ensemble de  $n$  assurés, le contrat garantit une indemnisation à l'assuré à qui l'on diagnostique l'une des maladies prise en charge par le contrat. On peut donc exprimer la distribution de ces diagnostics à l'aide d'une loi binomiale  $B(n; \theta)$  de paramètres  $n$  et  $\theta$ , où  $\theta$  est la probabilité de l'événement : dans notre étude cela correspond au diagnostic de l'une des maladies couvertes par le contrat.

L'approche classique de cette estimation considère les paramètres comme fixes mais inconnus et non comme des variables. On peut utiliser la méthode du maximum de vraisemblance pour obtenir des estimateurs :

$$\hat{\theta}_n = \arg \max_{\theta \in [0,1]} L_\theta(X_1, X_2, \dots, X_n)$$

La méthode du maximum de vraisemblance fournit des estimateurs qui sont en général assez satisfaisants car ils ont, sous certaines conditions, des propriétés recherchées telles que :

— la cohérence :

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

— l'efficacité asymptotique :

Pour définir le caractère efficace asymptotiquement d'un estimateur, définissons dans un premier temps la vraisemblance et l'information de Fisher d'un estimateur. Considérons un estimateur  $\theta$  et un échantillon de taille  $n$   $(x_i)_{i \in [1;n]}$ , on définit la fonction de vraisemblance  $L(x_1, x_2, \dots, x_n)$  comme :  $L_\theta = \mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ .

On définit l'information de Fisher comme l'espérance du carré de la dérivée partielle de la log-vraisemblance par rapport à  $\theta$  :

$$I_n(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]$$

L'inégalité de Cramer-Rao donne une borne inférieure pour la variance d'un estimateur  $\hat{\theta}$  sans biais estimant le paramètre  $\theta$  (i.e.  $\mathbb{E}(\hat{\theta}) = \theta$ ) :

$$\mathbb{V}(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

Finalement, nous pouvons définir le caractère efficace d'un estimateur : un estimateur sans biais est efficace si sa variance est minimale : avec l'inégalité de Cramer-Rao, cela se formalise comme :

$$\mathbb{V}(\hat{\theta}) = \frac{1}{I_n(\theta)}$$

Le caractère efficace étant souvent trop restrictif, on peut être amené à définir l'efficacité asymptotique, i.e. pour un grand échantillon sa variance est équivalente à cette borne inférieure, dite de Cramer-Rao :

$$\mathbb{V}(\hat{\theta}) \underset{n \rightarrow +\infty}{\sim} \frac{1}{I_n(\theta)}$$

Ainsi, la méthode du maximum de vraisemblance est la plus fréquemment utilisée. Le principe du maximum de vraisemblance nous dit que nous devrions utiliser comme estimation la valeur qui maximise la probabilité de l'événement observé. La différence essentielle de l'approche par statistiques inférentielles avec l'approche bayésienne est que les paramètres sont maintenant traités comme des variables aléatoires et qu'ils ont donc des distributions de probabilité que nous allons chercher à estimer.

Supposons que l'on observe un échantillon  $x = (x_1, x_2, \dots, x_n)$  d'une population ayant pour densité sachant  $\theta$  :  $f_\theta(x) = f(x|\theta)$ .

$$\text{Nous pouvons écrire : } f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

Le théorème de Bayes énonce que pour toute hypothèse  $H$  et tout événement  $A$  non négligeable, on peut écrire une relation entre les probabilités conditionnelles :

$$\mathbb{P}(H|B) = \frac{\mathbb{P}(B|H) \cdot \mathbb{P}(H)}{\mathbb{P}(B)}$$

Ce que nous pouvons réécrire en terme de densité :

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{\int_{\theta} f(x|\theta) f(\theta) d\theta}$$

Ce qui nous permet de déterminer la distribution a posteriori et de nous rendre compte que :

$$f(\theta|x) \propto f(x|\theta) \cdot f(\theta)$$

qui peut s'interpréter comme : *distribution a posteriori*  $\propto$  *vraisemblance* \* *distribution a priori*.

La distribution a posteriori contient toutes les informations disponibles sur  $\theta$  et doit donc être utilisée pour le calcul des primes.

L'approche Bayésienne que nous allons maintenant détailler, cherche à déterminer la distribution de  $\theta$ . Dans cette étude, nous ferons l'hypothèse d'une distribution de type *bêta* à deux paramètres,  $\alpha$  et  $\beta$  deux réels, dont la densité vaut :

$$\forall 0 \leq \theta \leq 1, f(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

Avec  $\Gamma$  la fonction *gamma* définie par :

$$\forall z > 0, \Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

On peut également réécrire cette expression à l'aide de la fonction *bêta* à deux paramètres également, définie par :

$$\forall z > 0, B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1 - t)^{\beta-1} dt.$$

En effet, considérons le produit  $\Gamma(\alpha)\Gamma(\beta)$ , on a :

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt \int_0^{+\infty} u^{\beta-1} e^{-u} du = \iint_{\mathbb{R}^+} t^{\alpha-1} u^{\beta-1} e^{t+u} dt du$$

Introduisons le changement de variable suivant :  $\begin{cases} r = t + u \\ rw = t \end{cases}$ , ou encore :  $\begin{cases} r = t + u \\ w = \frac{t}{t+u} \end{cases}$ .

On définit alors  $\Phi$   $\mathcal{C}^1$ - difféomorphisme permettant de réaliser le changement de variable :

$$\begin{aligned} \Phi : \mathbb{R}^+ \times [0; 1] &\longrightarrow (\mathbb{R}^+)^2 \\ (t, u) &\longmapsto (\Phi_1(r, w), \Phi_2(r, w)) \end{aligned}$$

Avec,  $\Phi_1(r, w) = rw$  et  $\Phi_2(r, w) = r(1 - w)$ .

En effet,  $\Phi$  est bijective,  $\mathcal{C}^1$  et  $\Phi^{-1}$  est aussi  $\mathcal{C}^1$ .

Notre produit d'intégrale devient donc, à l'aide du théorème de changement de variable :

$$\Gamma(\alpha)\Gamma(\beta) = \iint_{\mathbb{R}_+^2} f(\Phi(r, w)) |\det(J(\Phi)(r, w))| dr dw$$

Déterminons la matrice jacobienne de  $\Phi$  :

$$J(\Phi)(r, w) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial r} & \frac{\partial \Phi_1}{\partial w} \\ \frac{\partial \Phi_2}{\partial r} & \frac{\partial \Phi_2}{\partial w} \end{pmatrix} = \begin{pmatrix} w & r \\ 1-w & -r \end{pmatrix}$$

Et,

$$|\det(J(\Phi)(r, w))| = |-rw - r(1-w)| = |-r| = r$$

Soit :

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^1 \int_{\mathbb{R}_+} e^{-(rw+(1-w)r)} (rw)^{\alpha-1} (r(1-w))^{\beta-1} r dr dw = \int_0^1 w^{\alpha-1} (1-w)^{\beta-1} dw \int_{\mathbb{R}_+} e^{-r} r^{\alpha-1} r^\beta dr$$

En conclusion,

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^1 w^{\alpha-1} (1-w)^{\beta-1} dw \int_{\mathbb{R}_+} e^{-r} r^{\alpha+\beta-1} dr = B(\alpha, \beta)\Gamma(\alpha + \beta)$$

Nous pouvons alors réécrire l'expression de la densité d'une loi *beta* :

$$\forall 0 \leq \theta \leq 1, f(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

L'utilisation de cette distribution se justifie car le paramètre  $\theta$  se situe entre 0 et 1 et cette distribution modélise un paramètre entre 0 et 1.

L'un des intérêts de la loi *Bêta* est qu'elle s'interprète aisément en terme de succès ou d'échecs, ce qui nous est particulièrement utile ici pour estimer une proportion de population diagnostiquée atteint d'une maladie garantie par le contrat.

Elle permet également d'établir un intervalle de confiance autour de probabilités déterminées par une loi binomiale, tout en prenant en compte des observations déjà réalisées.

Nous pouvons alors, dans notre cas, exprimer la distribution a posteriori d'après (3.2.1) :

$$f(\theta|x) = \frac{f(x|\theta) \cdot f(\theta)}{\int_{\theta} f(x|\theta) f(\theta) d\theta}$$

$$f(\theta|x) = \frac{\binom{n}{x} \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}}{B(\alpha, \beta) \int_{\theta} f(x|\theta) f(\theta) d\theta}$$

Et :

$$f(\theta|x) = \frac{\theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}}{B(\alpha+x, \beta+n-x)}$$

Nous remarquons alors que la densité a posteriori est la densité d'une loi *bêta* avec les paramètres suivants :

$$\alpha' = \alpha + x \quad \text{et} \quad \beta' = \beta + n - x$$

Nous pouvons donc chercher à obtenir un estimateur bayésien de  $\theta$  convenable. A l'aide de la théorie de crédibilité, nous avons un critère de décision : le critère de l'erreur quadratique moyenne ou MSE (Mean Squared Error).

Celui-ci énonce qu'un estimateur  $\hat{\theta}$  de  $\theta$  est au moins aussi bon qu'un estimateur  $\tilde{\theta}$  si :

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \leq \mathbb{E}[(\tilde{\theta} - \theta)^2]$$

Ce qui est équivalent à écrire que le meilleur estimateur de  $\theta$  est la fonction des observations  $g(x)$  minimisant l'erreur quadratique moyenne :

$$\tilde{\theta} = \arg \min_{\theta} \mathbb{E}[(g(x) - \theta)^2]$$

Nous allons maintenant montrer que  $\tilde{\theta} = \mathbb{E}(\theta|X)$  est le meilleur estimateur de  $\theta$ .

En effet, considérons  $\hat{\theta}$  un estimateur de  $\theta$  et  $\tilde{\theta} = \mathbb{E}(\theta|X)$  l'espérance a posteriori de  $\theta$ .

On a :

$$\begin{aligned} \mathbb{E}[(\hat{\theta} - \theta)^2] &= \mathbb{E}[\mathbb{E}[(\hat{\theta} - \theta)^2|X]] = \mathbb{E}[\mathbb{E}[(\hat{\theta} - \tilde{\theta} + \tilde{\theta} - \theta)^2|X]] \\ \mathbb{E}[(\hat{\theta} - \theta)^2] &= \mathbb{E}[\mathbb{E}[(\hat{\theta} - \tilde{\theta})^2|X]] + 2\mathbb{E}[\mathbb{E}[(\hat{\theta} - \tilde{\theta})(\tilde{\theta} - \theta)|X]] + \mathbb{E}[\mathbb{E}[(\tilde{\theta} - \theta)^2|X]] \end{aligned}$$

Si l'on s'intéresse en particulier à  $\mathbb{E}[(\hat{\theta} - \tilde{\theta})(\tilde{\theta} - \theta)|X]$  :

$$\mathbb{E}[(\hat{\theta} - \tilde{\theta})(\tilde{\theta} - \theta)|X] = \mathbb{E}[\hat{\theta}\tilde{\theta} - \hat{\theta}\theta - \tilde{\theta}^2 + \theta\tilde{\theta}|X] = \tilde{\theta}\mathbb{E}[\hat{\theta}|X] - \mathbb{E}[(\hat{\theta}\theta|X)] - \tilde{\theta}^2 + \tilde{\theta}\mathbb{E}[\theta|X]$$

Finalement, comme  $\tilde{\theta} = \mathbb{E}(\theta|X)$  :

$$\mathbb{E}[(\hat{\theta} - \tilde{\theta})(\tilde{\theta} - \theta)|X] = 0$$

D'où :

$$\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[(\hat{\theta} - \tilde{\theta})^2] + \mathbb{E}[(\tilde{\theta} - \theta)^2]$$

Ainsi comme un carré est positif et avec la positivité de l'espérance,  $\mathbb{E}[(\hat{\theta} - \tilde{\theta})^2] \leq 0$ , il vient :

$$\mathbb{E}[(\tilde{\theta} - \theta)^2] \geq \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Ce qui prouve que  $\tilde{\theta} = \mathbb{E}(\theta|X)$  est le meilleur estimateur au sens de l'erreur quadratique moyenne.

Nous pouvons alors déterminer le meilleur estimateur de  $\theta$  dans notre exemple,  $\tilde{\theta} = \mathbb{E}(\theta|X)$ .

Et, nous avons démontré que la distribution de  $\theta|X$  est une loi *bêta*.

Calculons alors l'espérance d'une variable  $X$  suivant une loi *bêta* :

$$\begin{aligned}\mathbb{E}(X) &= \int_0^1 x f_X(x) dx = \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \\ \mathbb{E}(X) &= \int_0^1 \frac{x^{\alpha+1-1}(1-x)^{\beta-1}}{B(\alpha+1, \beta)} dx \times \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} = \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)}\end{aligned}$$

En effet, nous reconnaissons sous l'intégrale la densité d'une loi *bêta* donc son intégrale sur son domaine de définition, ici  $[0; 1]$ , vaut 1.

Nous pouvons poursuivre le calcul en utilisant l'expression de la fonction *bêta* à l'aide de la fonction *Gamma* démontrée précédemment :

$$\begin{aligned}\mathbb{E}(X) &= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)\Gamma(\alpha)} \\ \mathbb{E}(X) &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} = \frac{\alpha}{\alpha+\beta}\end{aligned}$$

Dans notre cas, nous avons montré que  $\theta|X$  suit une loi *bêta* de paramètres  $(\alpha+x, \beta+n-x)$ , nous pouvons donc en déduire l'espérance de  $\theta|X$  :

$$\mathbb{E}(\theta|X) = \frac{\alpha+x}{\alpha+x+\beta+n-x} = \frac{\alpha+x}{\alpha+\beta+n}$$

Ce qui peut se réécrire de la forme :

$$\mathbb{E}(\theta|X) = \frac{n}{\alpha+\beta+n} \left( \frac{x}{n} - \frac{\alpha}{\alpha+\beta} \right) + \frac{\alpha}{\alpha+\beta} = \frac{n}{\alpha+\beta+n} \frac{x}{n} + \left( 1 - \frac{n}{\alpha+\beta+n} \right) \frac{\alpha}{\alpha+\beta} = Z.M + (1-Z).\mu$$

Avec :

- $Z = \frac{n}{\alpha+\beta+n}$  le facteur de crédibilité qui dépend de  $n$ ,
- $M = \frac{x}{n}$  la moyenne des observations,
- $\mu = \frac{\alpha}{\alpha+\beta}$  l'espérance de la distribution a priori.

Si l'on étudie le facteur de crédibilité  $Z$ , ou plus particulièrement la suite  $(Z_n)_{n \in \mathbb{N}}$  :  
 On voit que trivialement  $Z_0 = 0$ , en effet si l'on ne possède aucune observation, l'estimateur devient l'espérance de la distribution a priori en l'absence d'informations supplémentaires.  
 Si l'on écrit  $Z_n$  de la forme :

$$Z_n = \frac{n}{n(1 + \frac{\alpha}{n} + \frac{\beta}{n})} = \frac{1}{1 + \frac{\alpha}{n} + \frac{\beta}{n}}$$

Et on remarque aisément que :  $\lim_{n \rightarrow +\infty} Z_n = 1$ , autrement dit, l'estimateur de  $\theta$  tends vers la moyenne des observations quand celles-ci sont de plus en plus nombreuses.

En pratique, il est possible de pas avoir d'information a priori, dans notre cas, si nous ne connaissons pas la distribution a priori de  $\theta$ , une solution qui peut être envisagée est de considérer que la variable  $\theta$  suit une loi uniforme sur  $[0; 1]$  par exemple (on sait que la loi uniforme sur cet intervalle est un cas particulier de la loi *bêta* où  $\alpha = \beta = 1$ ).

Dans ce cas précis, on a  $Z = 0$  et  $M = \mathbb{E}(X) = \frac{1}{1+1} = 0.5$ , en réalité cette estimation est trop élevée pour convenir à la probabilité de diagnostic d'une maladie redoutée. Pour pallier à cet surestimation, il conviendra dans la suite de considérer un intervalle  $[\theta_{min}; \theta_{max}]$  pour borner notre estimateur.

### 3.2.2 Modèle de risque individuel

Dans cette partie, l'on considère un portefeuille d'une compagnie d'assurance composé d'un nombre fixe  $n$  de polices indépendantes.

L'information recherchée par une compagnie d'assurances est le montant total de sinistres à indemniser sur une période déterminée, par exemple un exercice comptable.

Pour ce faire, nous allons modéliser la charge totale de sinistres comme la somme des sinistres de chaque police individuelle, d'où l'appellation "modèle de risque individuel".

Nous notons  $S_n$  la somme des sinistres du portefeuille, qui peut s'écrire de la forme :

$$S_n = \sum_n Y_i$$

Avec  $(Y_i)_{i \in [1; n]}$  correspondant au  $i^{me}$  risque et  $n$  le nombre de risques en portefeuille.

Nous allons maintenant établir les relations permettant d'estimer les paramètres  $\alpha$  et  $\beta$ . Pour ce faire, il est nécessaire de déterminer une fraction d'échantillonnage  $p$  permettant d'utiliser les

résultats précédents.

Pour cela, il faut que :

- $\mathbb{E}(p) = \theta$
- $\mathbb{V}(p) = \frac{\theta(1-\theta)}{n}$

Comme nous l'avons mentionné plus haut, nous avons fixé comme critère de comparaison l'erreur quadratique moyenne ou MSE.

Or, pour un estimateur  $\hat{\theta}$  :

$$\begin{aligned} MSE(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\ &= \mathbb{V}(\hat{\theta} - \theta) + [\mathbb{E}(\hat{\theta} - \theta)]^2 \\ &= \mathbb{V}(\hat{\theta}) + (\text{Biais}(\hat{\theta}))^2 \end{aligned}$$

Et la fraction d'échantillonnage  $p$  doit être sans biais, donc :

$$MSE(p) = \mathbb{V}(p) = \frac{\theta(1-\theta)}{n}$$

La moyenne de l'estimation ponctuelle bayésienne s'écrit de cette forme :  $\hat{\theta} = \frac{\alpha+x}{\alpha+\beta+n}$  et :

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}\left(\frac{\alpha+x}{\alpha+\beta+n}\right) = \frac{\alpha+n\theta}{\alpha+\beta+n}$$

On constate donc qu'il s'agit clairement d'une estimation biaisée, en effet, l'expression de l'espérance est différente du paramètre estimé  $\theta$ .

Si l'on s'intéresse à la variance de cet estimateur :

$$\begin{aligned} \mathbb{V}(\hat{\theta}) &= \mathbb{V}\left(\frac{\alpha+x}{\alpha+\beta+n}\right) = \frac{\mathbb{V}(\alpha) + \mathbb{V}(x)}{(\alpha+\beta+n)^2} \\ &= \frac{1}{(\alpha+\beta+n)^2} (0 + n\theta(1-\theta)) \\ &= \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} \end{aligned}$$

Et, pour tout  $\alpha$  et  $\beta$  positifs,

$$\mathbb{V}(\hat{\theta}) = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} < \frac{\theta(1-\theta)}{n} = \mathbb{V}(p),$$

L'estimation ponctuelle bayésienne a moins de variance que l'estimation ponctuelle classique.

En utilisant les précédentes relations, l'erreur quadratique moyenne de l'estimateur bayésien  $\hat{\theta}$  peut s'exprimer comme suit :

$$\begin{aligned} MSE(\hat{\theta}) &= \mathbb{V}(\hat{\theta}) + (\text{Biais}(\hat{\theta}))^2 \\ &= \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} + \left(\frac{\alpha + n\theta}{\alpha + \beta + n} - \theta\right)^2 \\ &= \frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} + \left(\frac{\alpha - (\alpha + \beta)\theta}{\alpha + \beta + n}\right)^2 \end{aligned}$$

Nous allons chercher à déterminer sous quelles conditions cette quantité est inférieure à l'erreur quadratique moyenne de l'estimation ponctuelle classique.

Cette condition peut s'écrire de cette manière :

$$\frac{n\theta(1-\theta)}{(\alpha + \beta + n)^2} + \left(\frac{\alpha - (\alpha + \beta)\theta}{\alpha + \beta + n}\right)^2 < \frac{\theta(1-\theta)}{n}$$

Nous allons dans la suite chercher des conditions sur les variables  $\alpha$  et  $\beta$ , car dans notre cas ce sont les seules que nous pouvons influencer pertinemment.

La modélisation a priori du paramètre  $\theta$  que nous estimons avec une loi *Beta* n'est pas aussi simple qu'il n'y paraît à première vue, en effet, bien qu'elle soit très flexible de part ses deux paramètres, elle est très sensible à une variation de ces paramètres.

En connaissant la moyenne et la variance de la distribution, nous pourrions utiliser la méthode des moments pour exprimer les paramètres  $\alpha$  et  $\beta$  en fonction des estimateurs de la moyenne et de la variance :

Pour ce faire, nous devons au préalable déterminer la variance d'une distribution *Beta* : Nous pouvons utiliser la relation :  $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ , pour ce faire il nous reste à déterminer  $\mathbb{E}(X^2)$  de la même manière que nous avons déterminé  $\mathbb{E}(X)$ .

$$\begin{aligned} \mathbb{E}(X^2) &= \int_0^1 x^2 f_X(x) dx = \int_0^1 x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \\ &= \int_0^1 \frac{x^{\alpha+2-1}(1-x)^{\beta-1}}{B(\alpha+2, \beta)} dx \times \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} = \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)\Gamma(\alpha)} \\ &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha+2+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \end{aligned}$$

Maintenant, si l'on pose  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  la moyenne empirique et  $v = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$ , nous pouvons résoudre le système suivant pour déterminer des estimateurs d' $\alpha$  et  $\beta$  :

$$\begin{cases} \bar{x} = \frac{\alpha}{\alpha + \beta} \\ v = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \end{cases}$$

Pour ce faire, on peut remarquer qu'avec la première équation on obtient aisément  $\beta$  en fonction d' $\alpha$  :  $\beta = \frac{1 - \bar{x}}{\bar{x}} \alpha$ .

Nous pouvons maintenant substituer  $\beta$  dans la deuxième équation :

$$\frac{1 - \bar{x}}{\bar{x}} \alpha^2 = v \left(1 + \frac{1 - \bar{x}}{\bar{x}}\right)^2 \alpha^2 \left[\left(1 + \frac{1 - \bar{x}}{\bar{x}}\right) \alpha + 1\right]$$

Comme  $\alpha > 0$ , nous pouvons diviser cette expression par  $\alpha^2$ , et obtenir :

$$\frac{1 - \bar{x}}{\bar{x}} = v \left(1 + \frac{1 - \bar{x}}{\bar{x}}\right)^2 \left[\left(1 + \frac{1 - \bar{x}}{\bar{x}}\right) \alpha + 1\right]$$

Et ainsi isoler le dernier  $\alpha$  :

$$\alpha = \frac{\frac{1 - \bar{x}}{\bar{x}} - v \left(1 + \frac{1 - \bar{x}}{\bar{x}}\right)^2}{v \left(1 + \frac{1 - \bar{x}}{\bar{x}}\right)^3} = \bar{x} \left[ \frac{\bar{x}(1 - \bar{x})}{v} - 1 \right]$$

Et enfin avec la relation que nous avons mis en évidence plus haut, on obtient  $\beta$  :

$$\beta = \frac{1}{1 - \bar{x}} \left[ \frac{\bar{x}(1 - \bar{x})}{v} - 1 \right] = \frac{\bar{x}}{v} - \frac{1}{1 - \bar{x}}$$

Cependant, en pratique s'il est aisé de déterminer précisément une valeur moyenne, la volatilité de la distribution est, elle, plus difficile à estimer, et comme nous avons deux paramètres pour la distribution considérée, nous avons besoin de deux relations pour les estimer, en effet, il existe une infinité de couples  $(\alpha, \beta)$  conduisant à une valeur moyenne similaire de cette distribution *beta*.

Nous avons alors essayé de trouver une solution optimale à ce problème.

Dans un premier temps, nous avons cherché s'il existe une distribution a priori (i.e. un ou plusieurs couples  $(\alpha, \beta)$ ) qui vérifierait l'inégalité recherchée.

Intuitivement, nous pouvons essayer avec les couples  $(\alpha, \beta)$  tels que la moyenne de la distribution a priori soit le paramètre recherché  $\theta$  :  $\frac{\alpha}{\alpha + \beta} = \theta$  i.e. comme le paramètre recherché est non

trivialement nul,  $\beta = \alpha(\frac{1}{\theta} - 1)$ , l'erreur quadratique moyenne devient alors :

$$\begin{aligned}
 MSE(\hat{\theta}) &= \frac{n\theta(1-\theta)}{(\alpha + \alpha(\frac{1}{\theta} - 1) + n)^2} + \left[ \frac{\alpha - (\alpha + (\alpha(\frac{1}{\theta} - 1))\theta)}{\alpha + \alpha(\frac{1}{\theta} - 1) + n} \right]^2 \\
 &= \frac{n\theta(1-\theta)}{(\frac{\alpha}{\theta} + n)^2} \\
 &= \frac{n\theta^3(1-\theta)}{(\alpha + n\theta)^2} \\
 &= \frac{n\theta^3(1-\theta)}{n^2(\frac{\alpha}{n} + \theta)^2} \\
 &= \frac{n\theta(1-\theta)}{n} \cdot \frac{\theta^2}{(\frac{\alpha}{n} + \theta)^2} < \frac{n\theta(1-\theta)}{n} \text{ car } \alpha, n > 0
 \end{aligned}$$

Nous avons donc montré qu'une distribution a priori dont la moyenne est égale au paramètre estimé satisfait l'inégalité  $MSE(\hat{\theta}) < MSE(p)$ .

Bien sur cette distribution est purement théorique, mais cela permet de vérifier qu'il existe une distribution vérifiant la propriété recherchée.

Nous recherchons maintenant des distributions a priori qui, en plus de satisfaire l'inéquation désirée, ont une espérance différente du paramètre à estimer.

Pour simplifier les calculs dans la suite, nous introduisons le couple de variables suivant :

$$(s, q) : \begin{cases} s = \frac{\alpha}{\alpha + \beta} \in ]0; 1[ \text{ correspondant à l'espérance de la distribution a priori,} \\ q = \frac{\alpha}{s} \in \mathbb{R}_+ \end{cases}$$

Ce qui nous permet d'écrire les relations suivantes :

$$\begin{aligned}
 \alpha &= qs \\
 \beta &= q(1-s) \\
 \alpha + \beta &= q
 \end{aligned}$$

Et l'inégalité recherchée devient alors :

$$\begin{aligned}
& \frac{n\theta(1-\theta)}{(q+n)^2} + \left(\frac{qs - q\theta}{q+n}\right)^2 < \frac{\theta(1-\theta)}{n} \\
\Rightarrow n\theta(1-\theta) + q^2(s-\theta)^2 & < \frac{\theta(1-\theta)}{n}(q+n)^2 \\
\Rightarrow q^2(s-\theta)^2 < \theta(1-\theta) & \left[\frac{(q+n)^2}{n} - n\right] \\
\Rightarrow (s-\theta)^2 < \theta(1-\theta) & \left[\frac{q^2 + 2nq + n^2 - n^2}{nq^2}\right] \\
\Rightarrow (s-\theta)^2 < \theta(1-\theta) & \frac{q+2n}{nq} \\
\Rightarrow |s-\theta| < \sqrt{\theta(1-\theta)} & \frac{q+2n}{nq}
\end{aligned}$$

En se remémorant que  $s$  correspond à l'espérance de la distribution a priori, cette inégalité donne un intervalle où doit se trouver cette espérance pour que l'inégalité soit vérifiée.

Cet intervalle s'écrit :

$$\left[\theta - \sqrt{\theta(1-\theta)} \frac{q+2n}{nq}; \theta + \sqrt{\theta(1-\theta)} \frac{q+2n}{nq}\right]$$

Cela va ainsi nous permettre de pouvoir définir l'intervalle  $[\theta_{min}; \theta_{max}]$  que nous avons mentionné dans la partie précédente

Si nous résolvons l'égalité au lieu de l'inégalité, nous obtenons les points extrêmes de cet intervalle, où les erreurs quadratiques moyennes des deux estimations (bayésienne et classique) sont égales.

Nous pouvons ainsi obtenir un paramètre transformé de la distribution a priori  $q$  en égalisant les deux expressions :

$$(s-\theta)^2 = \theta(1-\theta) \frac{q+2n}{nq}$$

Ainsi nous pouvons écrire :

$$q = \frac{2n\theta(1-\theta)}{n(s-\theta)^2 - \theta(1-\theta)}$$

### 3.2.3 Calcul de l'estimation ponctuelle bayésienne

Comme nous l'avons expliqué précédemment, il est difficile de déterminer une distribution a priori qui estimerait parfaitement le paramètre à estimer. En effet en pratique il est ardu d'estimer la volatilité de ce paramètre, il est beaucoup plus aisé de déterminer un intervalle dans lequel le paramètre se trouve presque sûrement.

L'objectif est alors de se trouver dans un intervalle dans lequel l'estimation bayésienne est plus précise que l'estimation classique.

Si l'on suppose que  $\theta$  se trouve dans l'intervalle  $[\theta_{min}; \theta_{max}]$ , nous prendrons comme valeur moyenne a priori la médiane de cet intervalle :

$$s = \frac{\theta_{min} + \theta_{max}}{2}$$

Ensuite, on détermine la valeur de  $q$  avec la relation déterminée au paragraphe précédent :

$$q = \frac{2n\theta(1-\theta)}{n(s-\theta)^2 - \theta(1-\theta)}$$

en utilisant à la place de  $\theta$  une borne de l'intervalle  $[\theta_{min}; \theta_{max}]$ , en pratique nous prendrons celle qui est la plus éloignée de 0,5.

Ce qui nous permet par la suite de pouvoir déterminer  $\alpha$  et  $\beta$  avec les relations écrites précédemment :

$$\alpha = qs$$

$$\beta = q(1-s)$$

Et enfin, nous pouvons en déduire l'estimateur de  $\theta$  à l'aide de la distribution a priori ainsi définie.

$$\mathbb{E}(\theta|X) = \frac{\alpha + x}{\alpha + \beta + n}$$

Nous avons donc montré comment déterminer un tarif à l'aide de ce modèle. Ce procédé vise à permettre d'intégrer les résultats des années passées dans le tarif. Nous pouvons ensuite utiliser ce modèle de crédibilité pour mettre en relation le tarif empirique obtenu dans la première partie et les données collectées par l'assureur au cours du pilotage de son contrat.

Des modèles d'arbres de décisions peuvent également être implémentés permettant d'entraîner les données obtenues. Ce modèle permet de diviser votre population assurée en sous-groupes homogènes en termes de risque. En pratique on ajuste un arbre de décision en utilisant toutes les variables explicatives disponibles, évalue les performances de l'arbre sur l'ensemble de test à l'aide du RMSE, et utilise l'arbre pour déterminer le tarif optimal pour chaque sous-groupe identifié par l'arbre.

# Conclusion

Pour conclure, nous avons vu dans un premier temps, que ces types de garanties sont novatrices et en plein essor. Cela constitue une sécurité supplémentaire pour l'assuré concernant un risque de plus en plus courant, surtout après la crise sanitaire qui a accentué le besoin de garanties Prévoyance et a créé une prise de conscience aussi sur l'après-traitement. Concernant l'assureur, cela constitue une réelle opportunité, composée de besoins nouveaux avec des perspectives en plein évolution comme la prévention. Il conviendra d'intégrer que cette prévention a un coût qui n'est pas à négliger et qui peut réduire la rentabilité du contrat même s'il est plus attractif. Cependant la couverture prévoyance existante étant déjà importante on peut se demander sous quelles formes ce type de produit a-t-il le plus d'avenir.

Ces produits tendent à évoluer, on peut imaginer des nouvelles garanties au-delà d'une simple somme unique versée, comme une rente ou de nombreux services secondaires pouvant être mis en place pour améliorer la situation de l'assuré et peut-être le personnaliser. Et même au niveau des pathologies : certains cancers ne sont pas aussi meurtriers que d'autres, on pourrait donc adapter le produit en ce sens. Nous constatons aussi que si le système de Santé et Retraite en France est un système avec l'un des plus haut taux de remboursement en Europe et dans le monde, cependant au niveau de la Prévoyance la situation est différente, le financement est très faible comme pour la dépendance, la défiscalisation sur ce types de garanties en collectif n'est pas possible ce qui le rend d'autant moins attractif.

Concernant les résultats obtenus dans notre étude, ils restent à fiabiliser et à être discutés. Certaines modélisations peuvent être améliorés et des modèles plus poussés peuvent être utilisés. Nous avons choisi dans cette étude d'utiliser des données en incidence. Nous aurions pu nous inspirer du travail de [Baione & Levantesi \(2014\)](#) qui estime les intensités de transition à l'aide de données de prévalence, comme l'étudiant également Oudin et Tosetti concernant le risque de dépendance dans [OUDIN & TOSETTI \(2002\)](#).

Dans ce projet, les contrats d'assurance des maladies graves et redoutées sont tarifés à l'aide de modèles à plusieurs états Markoviens. Tout d'abord, nous décrivons les produits d'assurance maladie grave disponibles sur le marché français et les comparons avec les produits offerts dans d'autres pays.

Ensuite, nous concevons plusieurs polices d'assurance de maladies graves typiques à étudier dans ce projet. La couverture de base du cancer, de l'accident vasculaire cérébral et de la crise cardiaque est prise en compte dans la police Stand Alone. Nous proposons également une police Full Accelerated qui comprends une prestation de décès. Un modèle à plusieurs états Markovien est introduit pour décrire l'état de santé dans notre conception et les probabilités et intensités de transition requises sont calculées pour la tarification des plans d'assurance maladie grave.

Enfin nous avons étudié de quelle manière la théorie de la crédibilité bayésienne peut nous permettre d'intégrer de l'expérience dans notre tarification. En effet, en ajustant une loi beta-binomiale, il est relativement aisé de montrer les relations entre distribution *a priori* et distribution *a posteriori* ce qui nous permet d'obtenir un tarif ajusté à nos observations.

# Bibliographie

- Asmussen, S. & Albrecher, H. (2010), *Ruin probabilities*, Vol. 14, World scientific.
- Asmussen, S., Laub, P. J. & Yang, H. (2019), ‘Phase-type models in life insurance : Fitting and valuation of equity-linked benefits’, *Risks* **7**(1), 17.
- Baione, F. & Levantesi, S. (2014), ‘A health insurance pricing model based on prevalence rates : Application to critical illness insurance’, *Insurance : Mathematics and Economics* **58**, 174–184.
- Booth, H. & Tickle, L. (2008), ‘Mortality modelling and forecasting : A review of methods’, *Annals of actuarial science* **3**(1-2), 3–43.
- Brett, P. & du Toit, J. (2007), A critical table : Pricing critical illness in the uk on a new insured lives table, in ‘Iaahs Colloquium, Cape Town, South Africa, May’, pp. 14–18.
- Dash, A. & Grimshaw, D. (1993), ‘Dread disease cover—an actuarial perspective’, *Journal of the Staple Inn Actuarial Society* **33**(1), 149–193.
- Dickson, D. C., Hardy, M. R. & Waters, H. R. (2019), *Actuarial mathematics for life contingent risks*, Cambridge University Press.
- Données globales d’épidémiologie des cancers* (2020). <https://www.e-cancer.fr/Professionnels-de-sante/Les-chiffres-du-cancer-en-France/Epidemiologie-des-cancers/Donnees-globales>, Last accessed on 2022-08-30.
- Dujardin, J.-J. & Cambou, J.-P. (2005), ‘Épidémiologie de l’infarctus du myocarde’, *EMC - Cardiologie-Angéiologie* **2**(4), 375–387.  
**URL:** <https://www.sciencedirect.com/science/article/pii/S1762613705000242>
- Fiche produit Maladies Redoutées-Carcept* (2022). [https://www.klesia.fr/sites/default/files/media/documents/2020/09/Fiche\\_Produit\\_Maladies\\_Redoutees\\_Carcept.pdf](https://www.klesia.fr/sites/default/files/media/documents/2020/09/Fiche_Produit_Maladies_Redoutees_Carcept.pdf), Last accessed on 2022-08-30.

- Gogola, J. & Kopecká, L. (2017), Multiple state models for critical illness policy, *in* ‘European Financial Systems 2017 : proceedings of the 14th International Scientific Conference’, Masarykova univerzita.
- Li, O. (2021), ‘The gen re dread disease survey – key takeaways for the hong kong market’. <https://www.genre.com/knowledge/blog/2021/december/the-gen-re-dread-disease-survey-key-takeaways-for-the-hong-kong-market-en>, Last accessed on 2022-08-30.
- Lim, S. (2022), ‘The gen re dread disease survey – key takeaways for the singapore market [infographic]’. <https://www.genre.com/knowledge/blog/2022/july/the-gen-re-dread-disease-survey-key-takeaways-for-the-singapore-market-en>, Last accessed on 2022-08-30.
- Nombre, incidence et caractéristiques des nouvelles exonérations pour ALD en 2019* (2020). <https://assurance-maladie.ameli.fr/etudes-et-donnees/incidence-nouvelles-exonerations-ald-2019>, Last accessed on 2022-08-30.
- LOUDIN, A. & TOSETTI, A. (2002), “‘éléments tarifaires en assurance dépendance’, *Bulletin Français d’Actuariat* .
- Survie des personnes atteintes de cancer en France métropolitaine (1989-2018)* (2018). <https://www.e-cancer.fr/Expertises-et-publications/Les-donnees-sur-les-cancers/Survie-des-personnes-atteintes-de-cancer-en-France-metropolitaine>, Last accessed on 2022-08-30.
- Wilson, D. L. (1994), ‘The analysis of survival (mortality) data : fitting gompertz, weibull, and logistic functions’, *Mechanisms of ageing and development* **74**(1-2), 15–33.

# Annexes

Age	Femme $S(x)$	Homme $S(x)$	Age	Femme $S(x)$	Homme $S(x)$
0	100000	100000	51	97447	95144
1	99666	99595	52	97230	94761
2	99643	99568	53	97000	94339
3	99630	99548	54	96753	93875
4	99618	99535	55	96486	93378
5	99610	99524	56	96194	92835
6	99602	99515	57	95877	92234
7	99595	99507	58	95533	91591
8	99590	99499	59	95174	90877
9	99585	99492	60	94777	90104
10	99578	99486	61	94348	89268
11	99573	99478	62	93898	88352
12	99567	99470	63	93422	87367
13	99560	99462	64	92920	86320
14	99551	99452	65	92386	85241
15	99543	99439	66	91812	84096
16	99532	99420	67	91209	82890
17	99519	99396	68	90565	81647
18	99507	99368	69	89876	80331
19	99487	99329	70	89145	78947
20	99471	99281	71	88364	77494
21	99452	99228	72	87507	75965
22	99431	99169	73	86575	74362
23	99412	99113	74	85572	72653
24	99391	99054	75	84474	70851
25	99372	98994	76	83270	68941
26	99349	98931	77	81969	66904
27	99327	98862	78	80525	64732
28	99303	98798	79	78938	62396
29	99278	98728	80	77161	59879
30	99247	98656	81	75159	57235
31	99219	98583	82	72911	54362
32	99188	98502	83	70444	51315
33	99156	98417	84	67680	48033
34	99119	98330	85	64597	44534
35	99078	98238	86	61200	40878
36	99035	98137	87	57449	37060
37	98986	98030	88	53337	33115
38	98931	97913	89	48928	29113
39	98873	97793	90	44236	25123
40	98810	97661	91	39354	21248
41	98741	97524	92	34399	17605
42	98662	97375	93	29530	14221
43	98578	97213	94	24760	11231
44	98479	97031	95	20254	8642
45	98374	96834	96	16138	6424
46	98257	96612	97	12497	4621
47	98126	96373	98	9421	3206
48	97980	96107	99	6858	2169
49	97820	95821	100	4874	1412
50	97645	95497	101	3337	934

TABLE 3.6 – Table de mortalité par sexe, 2017-2019, INSEE