Proposal to Extend Access to Loans for Serious Illnesses Using Open Data

Frédéric Planchet *,†[ID], **Édouard Debonneuil** *,‡ **and Marie Péju** §

ISFA—SAF Laboratory, Institut de Science Financière et d'Assurances (ISFA), Laboratoire SAF EA2429, University Lyon, Université Claude Bernard Lyon 1, F-69366 Lyon, France; peju.marie@gmail.com
* Correspondence: fred@planchet.net (F.P.); edebonneuil@yahoo.fr (É.D.)
† Frédéric Planchet is Professor at ISFA and Partner Actuary at PRIM'ACT.
‡ Edouard Debonneuil is manager of ActuRx.
§ Marie Péju is an independent *data scientist*.

Abstract: In France, access to a loan requires one to obtain loan insurance and the presence of a pathology in the applicant may be a reason for refusal. Improving knowledge of health risks and pooling risks are two methods of broadening access to loans. We attempt to analyse these possibilities using open data and risk pooling scenarios. We find that the removal of medical selection can be ensured if the current framework is adjusted. We also demonstrate how to use open data to estimate loan insurance premiums for a variety of diseases. We take two examples: breast cancer and type 1 diabetes. Broadening access to borrowing would be beneficial for patients and for the development of the economy associated with these projects.

Keywords: borrower insurance; mortality; serious illnesses

1. Introduction

Every year, borrowers are refused loans for the purchase of a property or the creation of a business because of their health status. The human impact is all the greater, as these patients did not choose to be ill. The impossibility of supplementing their financial resources with a loan to help them achieve a life goal can be perceived as discrimination. These refusals also have an economic impact beyond the loan itself, as these life projects would have contributed to the development of the economy.

Some of these refusals are due to insurers' insufficient knowledge of health risks. This refusal is the insurer's response to a risk of financial loss deemed too great due to the risks of death or temporary or permanent disability.

It may seem that insurers would have no issue in appreciating the risk. Insurers' actuaries generally know how to rate contracts on the basis of policyholder statistics. They also know how to deal with illnesses in various ways; in the underwriting process of protection insurance for products with death and disability benefits, premiums are increased notably when medical examinations reveal the presence of an illness. In some countries, such as in the United Kingdom, lower premiums are even applied for annuities in the presence of illness (Ridsdale 2012).

Still, for diseases that are currently refused, insurers do not have the corresponding statistics. It is not easy to find external information for reasons of health privacy. We have only partial access to data that enable us to construct statistics from such sources as health or life insurance data or national health data. Expert assessments of the risks associated with diseases are not unanimous; there are many diseases and associated risk factors, and scientific advances and changes in behaviour modify the risk levels. Faced with this lack of knowledge, insurers take a cautious approach (and they are obliged to do so to avoid bankruptcy and failure to honour their guarantees) and in practice have no choice but to

replace excessively high rates with a refusal to accept them in order to avoid a significant image risk (and they are obliged not to exceed a usury rate).

Furthermore, pooling high risks with low risks comes up against population behaviour. Within the existing framework, an insurer who creates a loan contract without medical selection, therefore being maximally inclusive with regard to pathologies, risks seeing an influx of patients who have been refused a loan elsewhere without this influx being compensated for by a sufficient influx of low-risk individuals; the experiment would come to nothing.

Thus, the existing framework does not favour the detailed knowledge of risks associated with pathologies or the pooling of high health risks with lower risks. Here we try to use open data and our knowledge of insurance to find ways out of the current situation. Throughout the paper we take breast cancer and diabetes as application examples. After an overview of mortality risks and the insurability of removing medical selection, we describe a generic approach to estimating loan insurance pure premiums for patients and apply it to breast cancer and type 1 diabetes. Throughout, we distinguish between mortality and temporary or permanent disability risks and provide elements for inclusive approaches. We hope that these elements can contribute to widening access to borrowing.

2. Overall Picture Based on French Mortality Benchmarks

For about fifty pathologies, the pathology sheets on the Ameli website (Ameli 2018a) provide information on the level of mortality in the general population. They give (at the top of their second page) mortality rates by age group. In Figure 1, we plot these mortality risks by age groups on a log scale for two pathologies, diabetes (Ameli 2018d) and breast cancer—both the risky active breast cancers (Ameli 2018b) and the less risky under (SEER 2020) breast cancers (Ameli 2018c). We also plot the mortality risks of various French populations not defined by pathology in order to compare them with the mortality risks of patients.
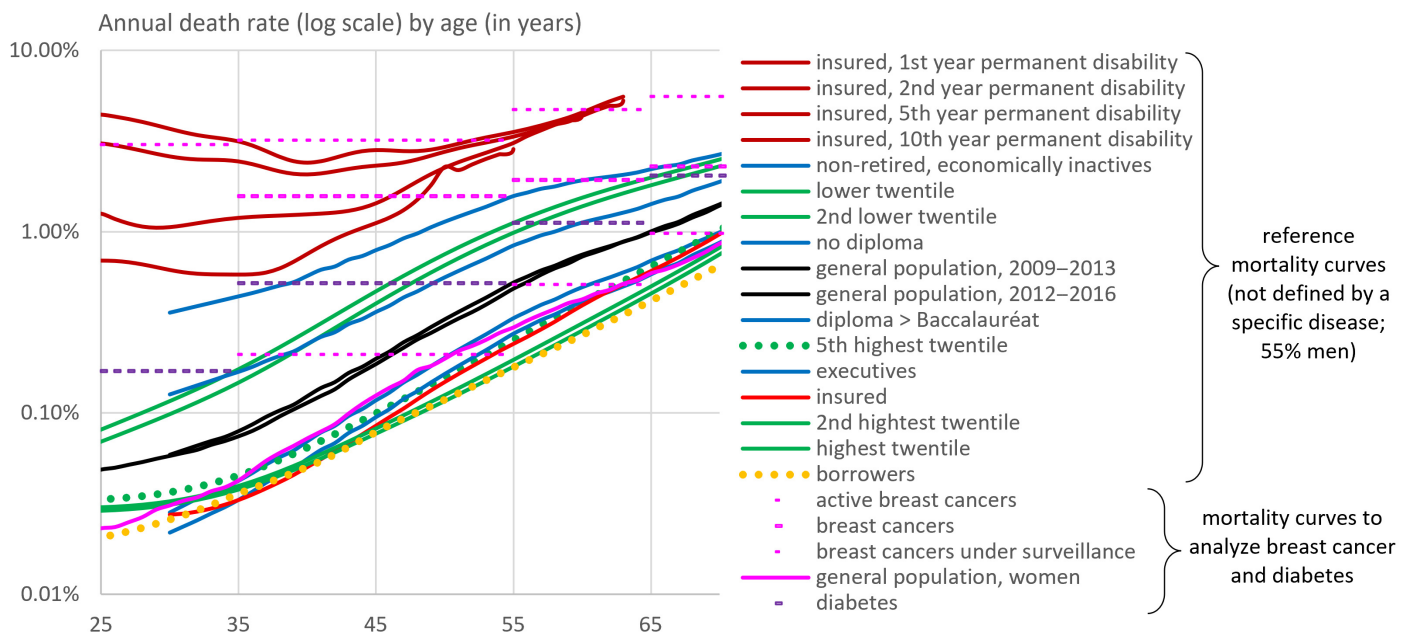

Figure 1. Annual mortality rate in France between the ages of 25 and 65 for different populations (55% men and 45% women)[1].

We can first observe the consistency of the risks of populations not defined by a pathology. The general French population had a similar risk (represented in black) in 2016 (noted general population in 2009–2013 (Blanpain 2016)) and 2018 (noted general population 2021–2016 (Blanpain 2018)) with a decreasing trend. When segmented in green by income twenties (Blanpain 2018) and in blue by diplomas (Blanpain 2016), the level of

mortality falls as one's socio-professional level increases. In red, the average risk of insured individuals, from a table known as "IA 2013" (Tomas and Planchet 2014), is around the line of executives and higher intellectual professionals (which is the lowest blue curve). In yellow-dotted lines, an average of actuarial tables in loan insurance[2], corresponding to the subscription to high incomes and an absence of severe pathology, demonstrates a risk lower than that in red of policyholders in general, which are, on average, less selected or even lower than the first twentieth in terms of income. The risk of death in case of disability, from tables known as "BCAC 2013" (Bagui 2013), is particularly high. Although it decreases in the early years (following an accident or illness, probably due to the death of those most at risk and a stabilization of health conditions for the others) before rising with age, it remains higher than for non-retired inactive individuals (disabled or not). We are used to the use of the latter curves and know that the order of magnitude is correct.

Secondly, the coherence observed between these reference curves allows us to reconstruct mentally and approximately what would be the mortality risks given by the Amelie pathology records if they were available in finer age brackets (the horizontal segments in Figure 1 would be inclined, similarly to the close reference curves, with approximately the same centres). We see that the risks of active breast cancer are of the order of those of disabled insured individuals in the years following the cause of disability and that the risks of the general population with breast cancer or diabetes are much higher than those of the borrower population, even in the case of breast cancer under surveillance. It is therefore necessary to study in greater detail the risks associated with different forms of disease and the transposition of risk to the socio-professional status of borrowers. This is conducted in Section 3.

Before moving to Section 3, however, it is interesting to note that Figure 1 helps us to appreciate what the risk of borrowers would be in the absence of medical selection. As the majority of loan applicants are executives, especially when weighted by amount, the risk should be close to the lowest blue curve. In the absence of medical selection, the risk should be similar to the red curve (tends to be higher due to this, tends to be lower due to the financial resources needed to borrow). Associating these professional categories with a twentieth, this would then correspond to the 5th twentieth of the French population, represented by the green-dotted line, i.e., an increase in risk of the order of 40% compared to current borrowers (yellow-dotted line)—or perhaps less, as patients with serious health conditions are likely to be less inclined to seek a loan and because fixed costs embedded into the insurance premiums absorb the risk (by definition they do not increase with the risk of borrowers). Given fixed insurance costs that do not increase with the risk of borrowers, commercial premiums that would keep current margins would be increased by less than 40% for all borrowers compared to current ones. However, given the almost 100-fold difference between these risks and those of people with active breast cancer, for example, this theoretical view's framework is insufficient and severe conditions are over-represented in one or a handful of insurers rather than being distributed evenly between insurers.

3. A Generic Method to Estimate Loan Insurance Premiums for Patients

We observe that if the framework is to be adjusted, it is important to refine our knowledge of the risks according to factors related to the pathology to minimize patient exclusion. In this section, we describe a method of finding aggregate open data in a way that estimates loan insurance premiums for a variety of pathologies. We do not claim that this is optimal; on the contrary, it sets the scene for potential improvements by other authors. We focus on France, but the approach is not limited to France.

3.1. Mortality Risk of Patients in the General Population According to Various Risk Factors

Finding mortality risks for specific diseases can typically be done with a Google or PubMed search on the disease and on keywords such as "survival", "age", "mortality rate", "death rate", "mortality", "death", "breast cancer", and "type 1 diabetes" (or their equivalents in French). As illustrated in Section 4 with breast cancer and diabetes, articles

are found that model mortality from national datasets, clinical trials, or epidemiologic studies, along with various risk factors.

We performed this for a number of diseases. Trade-offs have to be made between choosing recent articles or articles with more data and more risk factors. Otherwise article results must be combined to get the most out of different articles. Sometimes no adequate study is found in France or the desired country and studies from other countries are transposed while adjusting the results to macro statistics in France.

This step is relatively specific to each pathology.

3.2. Transposition to Borrowers

The mortality risks obtained so far correspond to patients in the general population, not to patients who have the typical wealth and lifestyle of borrowers. However, as observed in Figure 1, the mortality of current policyholders and insured borrowers is, on average, 2 to 2.5 times lower than that of the general population. Thus, do loan applicants with a medical condition have the normal risk of patients with that condition or about half the risk?

3.2.1. Theoretical Considerations and Definition of a Relative Risk Multiplier

It is logical that mortality in the event of illness is lower among loan applicants than the general population: they are, on average, more educated about health and therefore get screened earlier, have better living conditions and fewer co-morbidities and therefore recover more easily from health problems. Thus, the more the survival of a disease depends on these elements, the higher the survival of the insured persons compared to the general population.

A mathematical perspective provides an a priori idea of quantification for certain diseases. Insured populations are typically half as likely to die from all causes. This translates into a reference ratio for causes of death: the death rate for a given cause (death rate relative to the size of the population, not the diagnosed population) is 2 + x times less than that of the general population, where x is centred in some way around zero (sometimes positive, sometimes negative, depending on the pathology).

In particular, if a condition is very common, x is likely to be small in absolute terms (because if x were large, the other causes of death would have to have a large x on average in the other direction for all-cause mortality to be what it is). Mortality from this condition is itself decomposed as the product of incidence and mortality, if diagnosed. Here, depending on the disease considered, knowledge of the qualitative link between incidence and socio-professional category gives an idea of post-diagnosis mortality. Thus, in the case of breast cancer, the death rate (relative to the population size) of insured women can be expected to be about half that of women in the general population. Considering that screening is much more frequent and early in the higher socio-professional status, the incidence in the diagnostic sense and not in relation to a stage of the pathology should tend to be greater in insured women: we can expect the mortality rate in the case of breast cancer to be less than half that in the general population.

Cancers are special in that their early detection radically changes the associated survival—hence the screening approaches for breast cancer, melanoma, or colon cancer. For other diseases, we expect a lower incidence in the insured population than in the general population due to better health on average. However, the insured population tends to see the doctor earlier, so there is a tendency not to be very far from the incidence in the general population. Thus, the factor two of all-cause mortality would be found more in terms of mortality in the case of pathology than in terms of incidence.

Thus, the consensus is that the relative risk between mortality in the presence or absence of a specific pathology—the "relative mortality risk" due to that pathology—should be on average a little larger for insured individuals than for the general population, and smaller in the case of pathologies, such as cancers, for which insured individuals are screened particularly early. This is why we define the "relative risk multiplier" as the ratio

between these two relative risks, which should, therefore, on average, be a little higher than 1 but lower than 2, and probably lower than 1 for breast cancer.

3.2.2. Mathematical Definition of the Multiplier and How to Obtain it

Let us consider four populations: A', B', C', and D'. The prime is, here, to indicate that they are chosen to compute a risk multiplier ρ' that does not exactly correspond to any one insurance loan. Let us use the same letters to represent their annual mortality risk: $q_x^{A'}$, $q_x^{B'}$, $q_x^{C'}$, and $q_x^{D'}$. Table 1 shows how to choose the definition of these populations to obtain a relative mortality multiplier by combining their mortality risks.

Table 1. Risk multipliers.

	Population with a Pathology	Healthy Population (*)	Relative Mortality Risk Due to Disease
Higher socio-professional status	$q_x^{A'}$	$q_x^{B'}$	$q_x^{A'}/q_x^{B'}$
General population	$q_x^{C'}$	$q_x^{D'}$	$q_x^{C'}/q_x^{D'}$

(*) Here, the word "healthy" is an abuse of language to indicate that the population has not been selected for a particular pathology.

The relative risk multiplier ρ is then defined by:

$$\rho' = \frac{q_x^{A'}/q_x^{B'}}{q_x^{C'}/q_x^{D'}}$$

It is very difficult to find $q_x^{A'}/q_x^{B'}$ on the Internet as it requires articles that focus on a pathology in high socioeconomic populations, whereas articles that focus on a pathology tend to also focus on social inequalities and more particularly on low socioeconomic populations. However, the multiplier can be rewritten as follows:

$$\rho' = \frac{q_x^{A'}/q_x^{C'}}{q_x^{B'}/q_x^{D'}}$$

This might seem a pure change of notation, but it leads us to search for other types of articles in the scientific literature that are easier to find, as social inequalities are an important issue for social and health policies. It is what permits one to use open data to estimate the risks of borrowers with a disease without having to guess what ρ' might be. $q_x^{A'}/q_x^{C'}$ is found in articles that study the impact of socioeconomic status on survival outcome in the case of a disease. $q_x^{B'}/q_x^{D'}$ is found in articles that study the impact of socioeconomic status on mortality in the general population; it is important to find the same definition of socioeconomic status for the two ratios to be comparable (diplomas, occupations, salaries, etc.).

The relative risk multiplier is therefore computed with the definitions of high socioeconomic status and the pathology found in such articles. Let us remove primes when defining the relative risk multiplier that would be based on borrowers with a specific definition of the pathology related to the insurance contract and underwriting process:

$$\rho = \frac{q_x^{A}/q_x^{C}}{q_x^{B}/q_x^{D}}$$

Of note, D and D' represent the same general population. The mortality rate of loan applicants with a medical condition is then derived from the definition of ρ:

$$q_x^{A} = q_x^{B} \times \frac{q_x^{C}}{q_x^{D}} \times \rho$$

In this formula,

- $q_x{}^B$ (annual mortality rate of borrowers) and $q_x{}^D$ (annual mortality rate in the general population) are functions of age and possibly gender. $q_x{}^B$ is an average of placeholder borrower tables calculated by the laboratory; an insurer can apply its own experience table;
- $q_x{}^C$ (annual mortality rate of persons with the pathology) depends on age and possibly gender but also on disease-related variables, such as how long ago the diagnosis was made and elements of disease severity;
- $q_x{}^A$ (annual mortality rate of borrowers with the pathology) has all these variables;
- ρ is theoretically a function of age, wealth, history of disease, and any other characteristic associated with the loan applicant.

We make the assumption $\rho \approx \rho'$; we assume that the multiplier established with populations A′, B′, C′, and D′ is sufficient to rate contracts if it is a mortality adjustment. To validate this hypothesis, for a given pathology it can be useful to establish ρ' for various countries to ensure that it is robust with similar values despite different environments. Additionally, if ρ' were greater than 150% or less than 75%, for example, this assumption would need to be adjusted, for example by considering different categories of the pathology and different categories of high socio-professional status (at the cost of difficult internet searches).

In the case of breast cancer, we obtain a multiplier ρ of 83% for French data. For diabetes, we obtain a multiplier of 115% for France, 108% for Korea, 111% for the United States, and 123% for Scotland. It should be noted that these multipliers are in line with the most likely expectations expressed at the beginning of the section: multipliers slightly below 1 when early diagnosis in the higher socio-professional status makes a difference to the risks and slightly above 1 for frequent pathologies.

At this stage, it is therefore possible for insurers to establish mortality tables for an inclusive or partially inclusive approach for various conditions, bearing in mind that mortality risk is the most important guarantee in loan contracts.

3.2.3. Temporary Disability

As for mortality, for each disease the first step is to look for sources of data or studies that can be used to model the transition to temporary disability, or at least to work stoppage, and the continuation of temporary disability. This search is not easy, as the available data mentioning a specific disease in connection with work stoppage are often very macroscopic indicators. Luckily, in terms of modelling, mortality is the most important guarantee of loan insurance, thus approximations are more acceptable than when modelling mortality risk.

A solution is to rely on the American open data database made available by the National Center for Health Statistics (NHIS 2016). As this is an American database, the risks of temporary disability are lower and the prevalence of certain pathologies, such as diabetes, is higher. It is then a question of comparing the macroscopic indicators to appreciate the differences, modelling the risk in the United States, and adjusting it to the French macroscopic indicators within the framework of the understanding obtained when comparing the indicators.

This American database was compiled from a questionnaire sent each year between 2010 and 2018 to different, randomly selected individuals. It consists of 284,809 individuals, 55% of whom are women. A total of 3.3% of the women were diagnosed with breast cancer and the mean and median ages of these women are 68 and 70 years, respectively. The mean age at diagnosis is 56 years. It is possible to test different combinations of variables and thus select the most relevant ones.

With this database, we can model the annual probability i_x to get into work stoppage of more than d days, and the average duration of work stoppage that follows d_x. The risk is then the product of both ($i_x d_x$). Similar to mortality, we can define populations A, B, C, and D and their associated risks $i_x{}^A$, $i_x{}^B$, ..., $d_x{}^D$, and we get the formula that defines a risk multiplier ρ:

$$i_x{}^A d_x{}^A = i_x{}^B d_x{}^B \times \frac{i_x{}^C d_x{}^C}{i_x{}^D d_x{}^D} \times \rho$$

Here, A and B still represent French borrowers, but C and D represent American workers, with a specific pathology in the case of A and C. $i_x^C d_x^C$ and $i_x^D d_x^D$ are modelled based on the American database, $i_x^B d_x^B$ comes from knowledge of French insurers and we make the assumption that the risk multiplier is the same as for mortality: $\rho \approx \rho'$ (a higher mortality is associated with a more severe disease and greater risk of disability).

For applications for breast cancer and diabetes we decided to model incidence (i_x^C and i_x^D) and duration (d_x^C and d_x^D) separately to facilitate the modelling in a Markovian approach.

1. For incidence, we considered a logistic regression giving the annual probability of a work stoppage greater than three days. We chose a logistic regression because this annual incidence is a probability is between 0 and 1, but other models could have been chosen. The three-day threshold was chosen by expert judgement to differentiate temporary disability from pure work stoppage.
2. For duration, we considered a gamma regression, giving the duration of a work stoppage greater than 3 days. We chose a gamma regression because it leads to durations that are positive, but other models could have been chosen.

The use of regressions improves the understanding of the relationship between temporary disability and risk factors. It is important to look for the most relevant explanatory variables for the condition under study.

i_x^B is an average of placeholder borrower tables calculated by the laboratory using the same approach as for mortality. An insurer can use its own experience table. For d_x^B we used the BCAC 2013 tables (Bagui 2013).

The approach we described here is limited in terms of risk factors compared to the risk factors that are typically found in articles related to mortality risks. As cases that lead to higher mortality risk likely also lead to higher disability risks, we should include mortality-related risk factors with some proportionality rules and interpolations on the risk factors already modelled for disability risk.

3.2.4. Permanent Disability

As with temporary disability, it is particularly difficult to find permanent disability statistics related to diseases but we found an approach for modelling temporary disability with open data.

As shown in Table 2, a French study (Cuerq et al. 2008) provides the probabilities P_3 and P_{10} of becoming disabled 3 years and 10 years after diagnosis for various pathologies. It is possible to interpolate and extrapolate these data to model permanent disability risk as a function of time.

Table 2. Percentage of patients in permanent disability 3 and 10 years after the start of long-term sickness in France.

	Rate (%)	
Long-Term Sickness	**3 Years Later**	**10 Years Later**
Multiple sclerosis	14.0	23.4
Incapacitating stroke	19.8	21.9
Severe active rheumatoid arthritis	10.1	7.1
Chronic arteriopathies with ischemic manifestations	10.6	17.0
Coronary artery disease	9.9	15.1
Heart failure, severe heart disease	9.6	14.3
Severe chronic kidney disease and nephrotic syndrome	7.6	13.2
Severe forms of neurological conditions, severe epilepsy	8.8	13.0
Long-term psychiatric conditions	9.9	13.0
Severe chronic respiratory failure	9.0	12.6
Severe ankylosing spondylitis	8.0	12.5
Malignant tumors	8.4	10.8
Severe high blood pressure	5.1	9.8
Chronic active liver disease and cirrhosis	5.9	8.6
Type 1 and 2 diabetes	3.3	7.6
Crohn's disease and active ulcerative colitis	2.2	4.7
Severe primary immunodeficiency, AIDS	1.9	3.6

Let us denote the annual probability of entering permanent disability as p_x. Here, also, let us consider four populations, A, B, C, and D, and let us reuse the mortality risk multiplier ρ (that transposes relative risk to policy-holders) so that the following can be written:

$$p_x{}^A = p_x{}^B \times \frac{p_x{}^C}{p_x{}^D} \times \rho$$

where A and B represent policy-holders, C and D the population of workers, and A and C are identified as having a specific pathology.

Using constant values of $p_x{}^C$ between 0 and 3 years following diagnosis, noting that $q'_x{}^C$ is the annual death rate of people from population C when in permanent disability, and considering the average values of $p_x{}^C$, $q_x{}^C$ and $q'_x{}^C$ over the first three years post-diagnostic, P_3 is the probability of entering permanent disability in the first, second, or third year without dying:

$$P_3 = p_x{}^C \left(1 - q'_x{}^C\right)^2 + \left(1 - q_x{}^C - p_x{}^C\right) p_x{}^C \left(1 - q'_x{}^C\right) + \left(1 - q_x{}^C - p_x{}^C\right)^2 p_x{}^C$$

Considering that $p_x{}^C$, $q_x{}^C$, and $q'_x{}^C$ are small and that the product of more than two of these numbers is negligible compared to 1, the probability of ending up with permanent disability after 3 years is:

$$P_3 \approx p_x{}^C \left(3 - 3q'_x{}^C - 3q_x{}^C - 3p_x{}^C\right).$$

As a result, $p_x{}^C \approx P_3/3$ in the first three years following diagnosis and this estimation is slightly prudent. A simpler approach is to neglect mortality, as we see that it is slightly prudent with respect to $p_x{}^C$. Neglecting mortality, the probability of remaining out of permanent disability is:

$$1 - P_3 \approx \left(1 - p_x{}^C\right)^3$$

As a result, $p_x{}^C \approx 1 - (1 - P_3)^3$ is a refined approximation that is still slightly prudent overall from diagnosis to 3 years later.

Similarly, if we neglect mortality up to 10 years following diagnosis, the probability of remaining out of permanent disability starting from the percentage $1 - P_3$ of people 3 years after diagnosis is:

$$\frac{1 - P_{10}}{1 - P_3} \approx \left(1 - p_x{}^C\right)^7$$

As a result, $p_x{}^C \approx 1 - \left(\frac{1 - P_{10}}{1 - P_3}\right)^{1/7}$ is an estimation that is slightly prudent from year 3 to 10 after diagnosis. We use it for any year following the first three years post-diagnosis, and this might be prudent because, as diseases stabilise, the risk of entering long-term disability is lower.

For a given pathology, we have just defined $p_x{}^C$ as two numbers that represent weighted means over age, gender, and disease risk factors. This is why we define $p_x{}^D$ similarly. Concretely, we multiply the probability of entering disability during at least one month (Kusnick-Joinville et al. 2006) by the probability to remain in disability and enter permanent disability based on the BCAC 2013 tables as performed by (Bagui 2013), averaged over working ages. $p_x{}^B$ is an average of placeholder borrower tables calculated by the laboratory using the same approach as mortality.

We detailed how to compute the annual probability of entering permanent disability, $p_x{}^A$. We also need to define the annual probability of dying when in permanent disability, $q'_x{}^A$. In real life, it is greater than the death rate when not in permanent disability, but to evaluate loan insurance premiums it does not have great importance and $q_x{}^A$ already represents the weighted average of the two death rates, thus $q_x{}^A$ can be used as well.

3.2.5. Premium Calculation

A Markov model is a simple approximation for assembling mortality and temporary or permanent disability risks[3]. A four-state Markov model is shown in Figure 2.
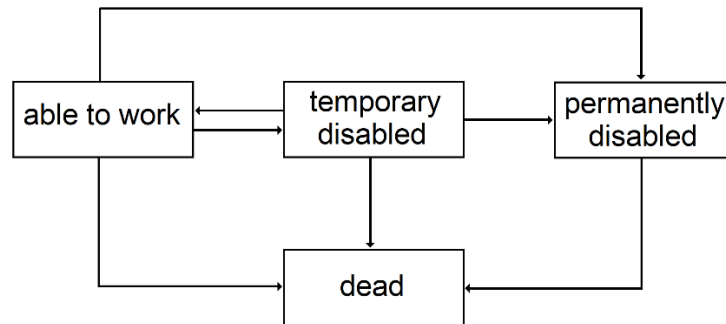

Figure 2. Diagram of a Markov model with four states.

Each arrow corresponds to a probability; the probability of death from one of the three other states, probability of going into temporary disability, of going into permanent disability before the end of the three years in temporary disability (after three years the transition is automatic), and the probability of returning to the work state. These probabilities would typically be functions of age, pathology characteristics, and length of time in the state or length of time in the diagnosis in a semi-Markovian framework.

In practice, it is difficult to find data for each arrow. A three-state Markov model shall thus be used, as described in Figure 3, with steps of one year.
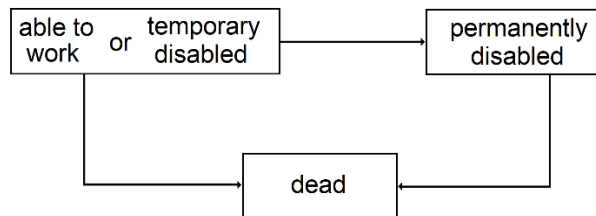

Figure 3. Diagram of a Markov model with three states.

At t = 0, the proportion of borrowers is 1 in the "Work or Disability" state and 0 in the other two states. It is therefore represented by a vector (1; 0; 0). At each subsequent time step, the proportion of borrowers in each state is the product of, on the one hand, the transition matrix made up of the annual transition probabilities and, on the other hand, the vector of the proportion of borrowers in the states of the previous time step.

Note that the matrix is as follows (in Table 3) and that its numerical value is different for each date (t = 0, 1, etc. over the duration of the contract) depending on the age and the initial characteristics of the pathology.

Table 3. Transition matrix.

$1-q_x{}^A-p_x{}^A$	$p_x{}^A$	$q_x{}^A$
0	$1-q_x{}^A$	$q_x{}^A$
0	0	1

The pure premium is the sum of the amounts the insurer must pay to the borrowers in the year for each state multiplied by the proportion of borrowers in the state. The simplest example is that of a loan covered by a borrower's insurance policy for death only, as it

does not depend on parameters related to disability. Each year $t \geq 0$ the proportion S_t of patients is alive:

$$S_0 = 1 \quad S_t = \prod_{x=x_0}^{x=x_0+(t-1)} 1 - q_x{}^A$$

who have borrowed an amount C at age x_0 for a term T with an interest rate r. In the case of death, the insurer must take over the payment of the remaining balance, or capital remaining due (CRD):

$$CRD_t = \left[\frac{C}{T}\left(1+r\right)^T\right](T-t)$$

assuming in this loan that the same amount, in square brackets, is to be repaid every year. Then, the pure premium P^A is:

$$P^A = \sum_{t=0}^{T} S_t \, q_{x_0+(t-1)}{}^A \, CRD_t$$

Note that in a two-state model (alive, dead) S_t is obtained by reading the proportion in the first state; in the three-state model (work-or-disability, disability, and death) S_t is the sum of the proportions of the first two states.

The inclusion of charges (e.g., distribution, claims handling, and reinsurance costs) would result in a commercial premium that is higher than the pure premium.

This pure premium P^A for a borrower with a medical condition can be compared to the pure premium P^B for a borrower under usual conditions by replacing A with B in the transition matrix. The additional premium $P^A - P^B$ due to the pathology can be defined as a percentage p of the usual premium or a permillage m:

$$p = \frac{P^A - P^B}{P^B} 100$$

$$m = \left(P^A - P^B\right)1000$$

In practice, not all expenses are proportional to the risk: there are fixed costs. The fixed costs do not increase with risks; therefore, surcharges in real life should be lower than the one we computed here.

Examples are given below for breast cancer and diabetes.

4. Application to Two Diseases: Breast Cancer and Diabetes

We apply the modelling to breast cancer and type 1 diabetes.

4.1. Breast Cancer

4.1.1. Mortality by Risk Factor

Dabakuyo et al. (2008) modelled mortality in women with active invasive breast cancer. This study was based on data from 3831 French women with breast cancer. The data were collected between 1982 and 1997, with a median follow-up time of 9 years, on women aged 19 to 99 at diagnosis (mean and median age of 61 years). They model risk ratios (HR) based on 2615 women and the following risk factors:

- Age at diagnosis;
- Oestrogen receptor function;
- TNM stage;
- SBR grade.

The impact of these risk factors was calculated with a Cox model using the following reference population:

- Age between 45 and 59;
- Stage T of 1;

- Stage N of 0;
- Stage M of 0;
- Oestrogen receptor dysfunction;
- SBR grade of 1.

We do not provide the results of the model here; the goal is to illustrate the type of information one finds to model loan insurance risks.

Such a model of mortality risk must be transposed from the general population to the insured population using a multiplier defined as $\rho' = \frac{q_x^{A'}/q_x^{C'}}{q_x^{B'}/q_x^{D'}}$.

To obtain $q_x^{A'}/q_x^{C'}$, a ratio of mortality risks of women with breast cancer but different socio-professional groups, we relied on the study by (Gentil-Brevet et al. 2008). This study is based on a sample of 1150 French women with invasive breast cancer diagnosed between 1995 and 1997 who were followed up with to 2006. It is important to carefully note the distribution of socio-professional status in the study in order then to compute $q_x^{B'}/q_x^{D'}$ accordingly:

- 42.5% were in the higher socio-professional status (SPS+) group defined as "executives, middle professional group, and clerical employees". Their survival 5 and 7 years post-diagnosis was $(1 - q_x^{A'})^5 = 88.2\%$ and $(1 - q_x^{A'})^7 = 83.0\%$, thus on average $q_x^{A'} = 1 - 88.2\%^{\frac{1}{5}}$ and $q_x^{A'} = 1 - 83.0\%^{\frac{1}{7}}$;
- 16.6% did not have a specified SPS;
- 40.9% were in the lower socio-professional status (SPS−) group defined as "famers, artisans, manual workers, unemployed"; Their survival 5 and 7 years post diagnosis was 77.4% and 69.4%.

The general population being approximately represented by as many SPS+ as SPS−, we computed $q_x^{C'}$ as the average of the mortality rate of the two populations: $q_x^{C'} = \left[1 - 88.2\%^{\frac{1}{5}} + 1 - 77.4\%^{\frac{1}{5}}\right]/2$ over 5 years and $q_x^{C'} = \left[1 - 83.0\%^{\frac{1}{5}} + 1 - 69.4\%^{\frac{1}{5}}\right]/2$.

This leads to $\frac{q_x^{A'}}{q_x^{C'}} = 66.4\%$ from diagnosis to 5 years later and $\frac{q_x^{A'}}{q_x^{C'}} = 68.1\%$ from diagnosis to 7 years later. Thus, we have two possible values.

To obtain $q_x^{B'}/q_x^{D'}$, a ratio of mortality risks depending on socio-professional status, we used mortality tables by sex and standard of living quintile (Blanpain 2018). As in the article by (Gentil-Brevet et al. 2008), we separated the socio-professional status into two groups (SPS+ and SPS−). We used the same weighting as the article for the population affected by breast cancer: 25% of women are under 50 years of age, approximately 50% are between 50 and 70 years of age and 25% are between 71 and 85 years of age. We obtained the mortality rates of 686 and 860 per 100,000 persons, respectively, for the SPS+ and SPS− populations: $\frac{q_x^{B'}}{q_x^{D'}} = \frac{686}{860} = 80\%$.

Alternatively, $q_x^{B'}/q_x^{D'}$ can be obtained with mortality tables by sex and the socio-professional status provided by (Blanpain 2016). It needs to be combined with the distribution of socio-professional status among females, such as those provided by (Maruani and Meron 2012) to make weighted average mortality rates that have similar socio-professional statutes, as in (Gentil-Brevet et al. 2008). We do not detail the calculation, as it is long and does not bring much to this article—we obtained $\frac{q_x^{B'}}{q_x^{D'}} = 85\%$.

The multiplier is then one of four different values depending on the choice of the numerator and denominator: $\rho' = \frac{66.4\,\%}{80\,\%} = 83\%$ or $\frac{68.1\,\%}{80\,\%} = 85.1\%$ or $\frac{66.4\,\%}{85\,\%} = 78.1\%$ or $\frac{68.1\,\%}{85\,\%} = 81.1\%$. We chose the first value, as it is close to the average of the four values and slightly higher—thus closer to more prudent premiums (additionally, while building these values we felt it had the best match of assumptions between the numerator and denominator).

4.1.2. Short-Term Disability

Using the American database, incidence and duration were modelled among women as follows:

- A logistic regression (glm function in R, using "family=binomial(logit)") yields the following average incidence rate:

$$i_x = \frac{1}{1 + e^{-(\alpha + \beta x + \gamma \, 1_{breast \; cancer})}}$$

where $1_{breast \; cancer}$ is 1 or 0 depending on whether the person was diagnosed with breast cancer or not. We then have $i_x{}^D = \frac{1}{1 + e^{-(\alpha + \beta x)}}$ and $i_x{}^C = \frac{1}{1 + e^{-(\alpha + \gamma + \beta x)}}$. Our regression found that $\beta = -0.019$ (more frequent work stoppage at younger ages) and $\gamma = 0.53$ (effect of cancer). α was adjusted to match the average frequency of work stoppage in France (Kusnick-Joinville et al. 2006);

- A gamma regression (glm function in R, using "family = Gamman(link = 'log')") yields the following average duration:

$$d_x = e^{-(\alpha + \beta x + \gamma \, 1_{breast \; cancer})}$$

We then have $d_x{}^D = e^{-(\alpha + \beta x)}$ and $d_x{}^C = e^{-(\alpha + \gamma + \beta x)}$. Here, $\beta = 0.018$ (longer work stoppage with age) and $\gamma = 0.73$ (effect of cancer). α was adjusted to match the average frequency of work stoppage in France (Kusnick-Joinville et al. 2006).

4.1.3. Long-Term Disability

In Table 2, we do not have data for breast cancer, only for cancers ("tumours"). We still considered the latter, which is prudent, as breast cancer has a similar prognostic value to most cancers.

- $p_x{}^C = 1 - (1 - 0.084)^{1/3} = 2.88\%$ for t between 0 and 3 years.
- $p_x{}^C = 1 - \left(\frac{1 - 0.108}{1 - 0.084} \right)^{1/7} = 0.38\%$ for t beyond 3 years.

4.1.4. Results

Table 4 illustrates additional premiums, in the last column, depending on various breast cancer cases described in the other columns. These are the m values we obtain for women who are 42 years old at the time of application for a 10-year loan and were diagnosed 3 years ago with a non-metastatic breast cancer.

The additional commercial premium may be smaller than the additional pure premium due to fixed costs (and commercial decisions).

It is possible to pool some cases, as shown in Table 5, to provide loan access to a maximum number of patients. The weighting used to calculate the average excess premiums is the distribution of such cancers found in SEER[4], a free application that provides very precise statistics on cancers in the United States.

Figures 4 and 5 show the additional disability pure premium (this time expressed as a percentage p) as a function of the age of the borrower at the time of application. Here, all severities of breast cancers have been pooled.

Of note, it would not have made much sense to express additional mortality premiums as a percentage p. This is because it would have been particularly high for young women (who have a particularly low mortality risk in the absence of breast cancer)—according to the model we developed, we observe that Tables 3 and 4 are roughly valid for all ages. Additionally, it would not have made much sense to express additional disability premiums as a permillage m. This is because long-term disability is limited at higher ages.

Table 4. Detailed additional pure mortality premiums.

T Stage	N Stage	M Stage	SBR Grade	Oestrogen Receptor Function	m
1	0	0	1	Positive	2.2
2	0	0	1	Positive	6.4
3	0	0	1	Positive	8.9
1	1	0	1	Positive	5.9
2	1	0	1	Positive	14.6
3	1	0	1	Positive	19.7
1	0	0	2	Positive	5.8
2	0	0	2	Positive	14.4
3	0	0	2	Positive	19.4
1	1	0	2	Positive	13.2
2	1	0	2	Positive	31.0
3	1	0	2	Positive	41.3
1	0	0	3	Positive	8.3
2	0	0	3	Positive	20.0
3	0	0	3	Positive	26.8
1	1	0	3	Positive	18.4
2	1	0	3	Positive	42.4
3	1	0	3	Positive	56.2
1	0	0	1	Negative	3.2
2	0	0	1	Negative	8.5
3	0	0	1	Negative	11.7
1	1	0	1	Negative	7.8
2	1	0	1	Negative	18.9
3	1	0	1	Negative	25.4
1	0	0	2	Negative	7.7
2	0	0	2	Negative	18.7
3	0	0	2	Negative	25.1
1	1	0	2	Negative	17.2
2	1	0	2	Negative	39.7
3	1	0	2	Negative	52.7
1	0	0	3	Negative	10.9
2	0	0	3	Negative	25.8
3	0	0	3	Negative	34.4
1	1	0	3	Negative	23.8
2	1	0	3	Negative	54.1
3	1	0	3	Negative	71.3

Table 5. Pooled additional pure mortality premiums.

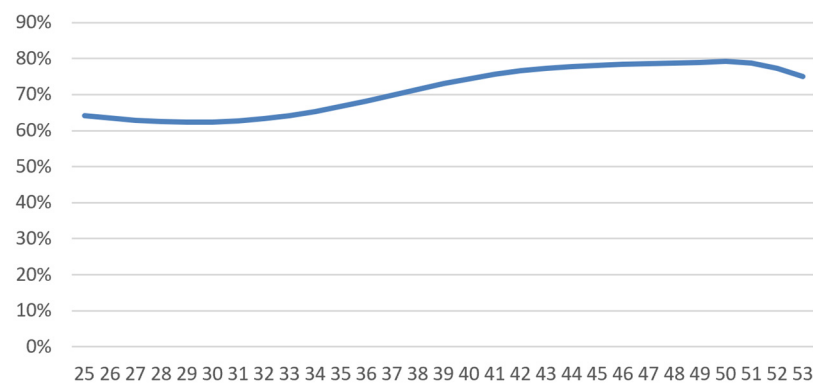| T Stage	N Stage	M Stage	SBR Grade	Oestrogen Receptor Function	p
-	-	0	-	-	13.5
1	-	0	-	-	5.8
2	-	0	-	-	21.1
3	-	0	-	-	31.7
4	-	0	-	-	36.7
-	1	0	-	-	7.2
-	0	0	-	-	24.9
-	-	0	1	-	3.5
-	-	0	2	-	11.5
-	-	0	3	-	21.1
-	-	0	-	Negative	21.4
-	-	0	-	Positive	11.3

Figure 4. Additional pure premium *p* for short-term disability as a function of age.
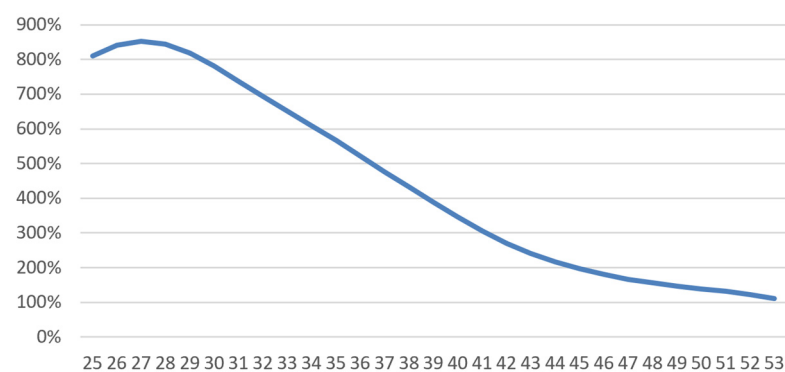

Figure 5. Additional pure premium *p* for long-term disability as a function of age.

4.2. Diabetes

There is much open data about diabetes in France but little about type 1 diabetes. Thus, to take a more complex example and appreciate the limits of what open data brings, we present the modelling of type 1 diabetes.

4.2.1. Mortality by Type of Diabetes

We failed to identify French articles that finely modelled type 1 diabetes mortality, thus we searched for articles from neighbouring countries. We selected a Danish study that investigated the evolution of relative and absolute mortality rates in 4821 type 1 diabetic patients followed between 2002 and 2011 (Jørgensen et al. 2013). Of these, 54% were male, the median age was 44 years, and the median follow-up time was 18 years.

Standard-mortality ratios (SMR) are calculated relative to the general population by sex, current age, age at diagnosis, and the presence or not of kidney disease, as diabetic nephropathy is a complication that can progress to renal failure. They are presented in Figure 6.

Unsurprisingly, mortality is higher for those diagnosed at a young age and for those with kidney disease. Given the high risk associated with kidney disease, in what follows we decided to limit the mortality risk to other cases of type 1 diabetes.

To model the mortality rates of diabetics in France, we apply the Danish SMR to the mortality rates of the general French population. The model must then be transposed from the general population to the insured population using a multiplier defined as
$$\rho' = \frac{q_x^{A'}/q_x^{C'}}{q_x^{B'}/q_x^{D'}}.$$
We did not find articles to produce this multiplier specifically for type 1 diabetes in France, so we decided to simply estimate it for diabetes (essentially type 2).
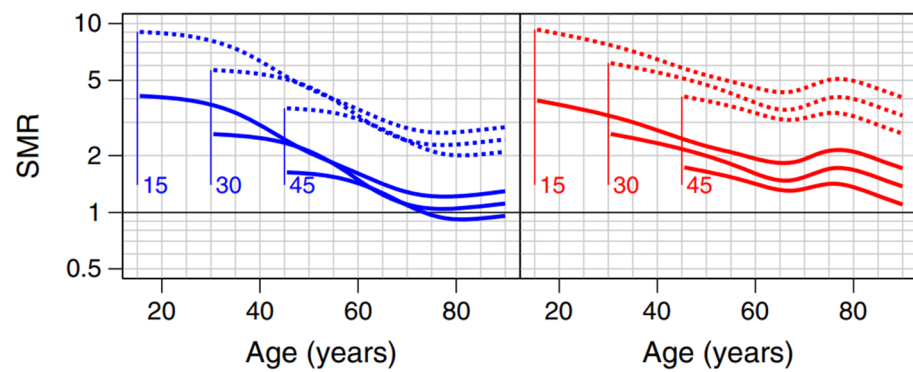
Figure 6. SMR by age at diagnosis and current age[5].

To obtain $q_x^{A'}/q_x^{C'}$, a ratio of mortality risks with diabetics from different socio-professional groups, we relied on a French article (Piffaretti et al. 2016). They studied the mortality of 7218 individuals with type 2 diabetes who answered a questionnaire in 2001 and 2007 and were followed until 2013. Figure 7 shows the relative risks according to their socio-professional category.
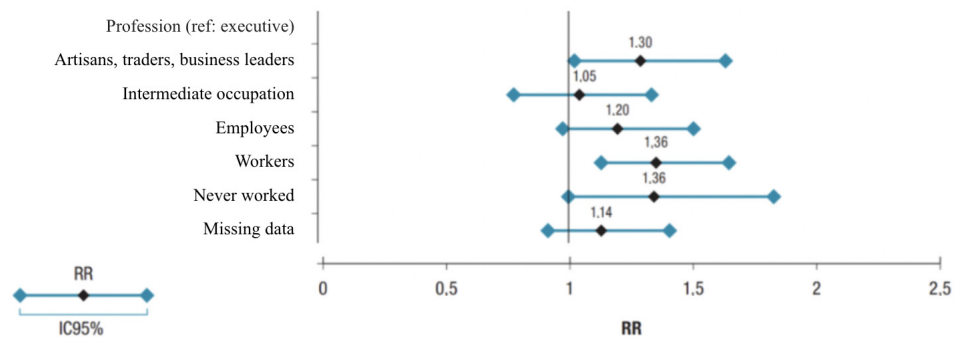

Figure 7. Relative risks according to socio-professional category with executives as reference (source: Piffaretti et al. 2016).

Figure 8 shows the distribution of socio-professional status. Based on the two figures, we estimate $\frac{q_x^{A'}}{q_x^{C'}} = \frac{1}{q_x^{C'}/q_x^{A'}}$ where A′ are executives with diabetes and C′ is decomposed into various categories:

$$\frac{q_x^{A'}}{q_x^{C'}} = \frac{1}{1*9.6\% + 1.36*9.0\% + 1.05*11.7\% + 1.20*18.4\% + 1.36*26.7\% + 1.36*6.3\% + 1.14*18.3\%} = 82\%$$

For $q_x^{B'}/q_x^{D'}$, we use mortality tables by sex and socio-professional status (Blanpain 2016). In order to be able to compare the average relative risks, we used the age and sex weighting of the article used for diabetes (Piffaretti et al. 2016).

We obtain $\frac{q_x^{B'}}{q_x^{D'}} = 71.6\%$.

This leads to the multiplier $\rho' = \frac{q_x^{A'}/q_x^{C'}}{q_x^{B'}/q_x^{D'}} = 82\%/71.6\% = 115\%$.

We searched for multipliers from other countries and found articles from which it is straightforward to obtain a multiplier for Scotland (Walter et al. 2011), the United States (Dray-Spira et al. 2010), and South Korea (Kim et al. 2016). We observe that the French multiplier is of the same order of magnitude as these multipliers, which are presented in Table 6.

Baseline characteristics of the Entred 2001 and Entred 2007 cohorts and mortality rate of people with type 2 diabetes who responded to the patient self-questionnaire (N=7 218), France

		Frequency (%)	Mortality rate for 1000 patients
Occupation and socio-professional category	Artisans, traders, business leaders	9.0	41.0
	Executive	9.6	29.2
	Intermediate occupation	11.7	24.4
	Employees	18.4	22.8
	Workers	26.7	31.7
	Never worked	6.3	30.4
	Missing data	18.3	36.5

Figure 8. Distribution of socio-professional status.

Table 6. Multiplier for diabetes in various countries.

South Korea	108%
United States	111%
Scotland	123%
France	115%

4.2.2. Short-Term Disability by Age and Age of Diagnosis

To model disability in type 1 diabetes, we relied on American database statistics from 2016 and 2017 due to the diabetes type variable only being present for these two years. During this period, 60,000 Americans completed questionnaires, capturing the impact of type 1 diabetes on the incidence and maintenance of disability. We modelled the incidence and duration as follows:

- A logistic regression yields the following average incidence rate:

$$i_x = \frac{1}{1 + e^{-(\alpha + \beta x + \gamma \, 1_{diabetes \; type \; 1} + \delta \, \tau)}}$$

where $1_{diabetes \; type \; 1}$ is 1 or 0 depending on whether the person was diagnosed with breast cancer or not, and τ is the time since diagnosis (set to 0 for those without type 1 diabetes). We then have $i_x{}^D = \frac{1}{1 + e^{-(\alpha + \beta x)}}$ and $i_x{}^C = \frac{1}{1 + e^{-(\alpha + \gamma + \beta x + \delta \, \tau)}}$. Regression found $\beta = 0.003$ (almost no impact of age, overall), $\gamma = 0.14$ (effect of type 1 diabetes), and $\delta = -0.04$. α was adjusted to match the average frequency of work stoppage in France (Kusnick-Joinville et al. 2006).

- A gamma regression yields the following average duration:

$$d_x = e^{-(\alpha + \beta x + \gamma \, 1_{diabetes \; type \; 1} + \delta \, \tau)}$$

We then have $d_x{}^D = e^{-(\alpha + \beta x)}$ and $d_x{}^C = e^{-(\alpha + \gamma + \beta x + \delta \, \tau)}$. Here, $\beta = 0.024$ (longer work stoppage with age) and $\gamma = 0.32$ (effect of type 1 diabetes). α was adjusted to match the average frequency of work stoppage in France, observed in (Kusnick-Joinville et al. 2006).

4.2.3. Long-Term Disability

In Table 2, we do not have data for type 1 diabetes, only for type 1 or 2 diabetes. We still considered the latter. We think it is prudent, as our short-term disability analysis in the US database demonstrates lower risks for type 1 diabetes than type 2 (data not shown).

- $p_x{}^C = 1 - (1 - 0.033)^{1/3} = 1.11\%$ for t between 0 and 3 years.
- $p_x{}^C = 1 - \left(\frac{1-0.076}{1-0.033}\right)^{1/7} = 0.65\%$ for t beyond 3 years.

4.2.4. Results

Figure 9 shows additional mortality premiums expressed in percentage depending on applicant age and how long ago the applicant was diagnosed with type 1 diabetes.
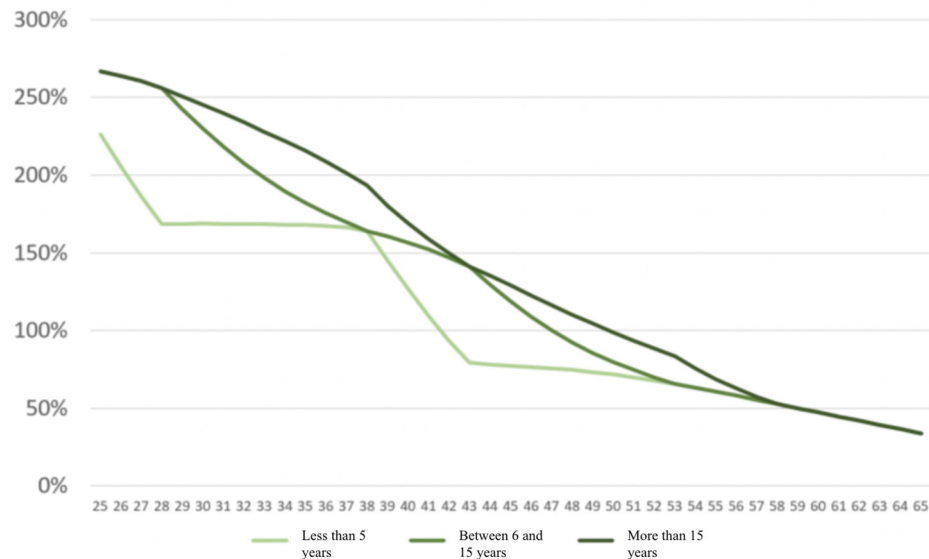

Figure 9. Additional pure premium p for mortality risk as a function of applicant age and whether diagnosis is recent or not (three cases).

Figure 10 shows the additional pure premium for disability (both short-term and long-term) as a function of applicant age and time since the diagnosis of type 1 diabetes.
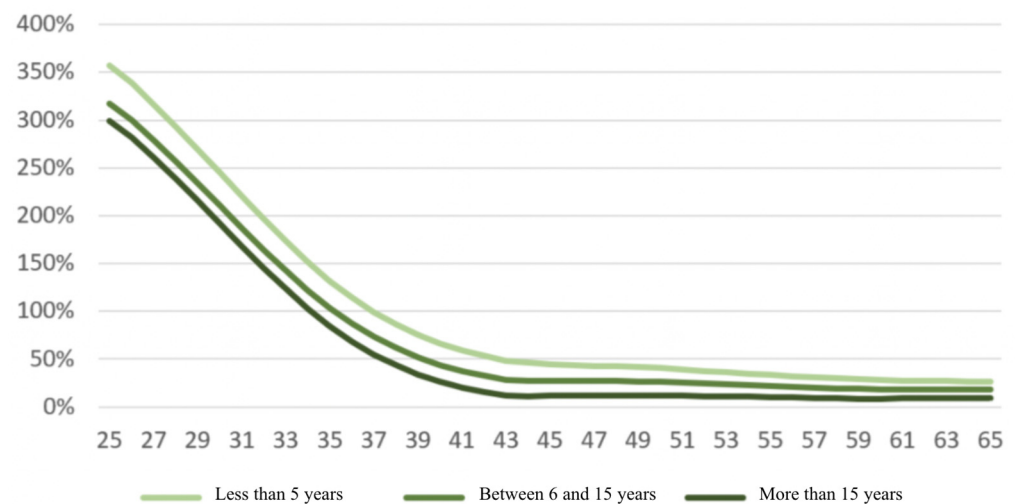

Figure 10. Additional pure premium p for disability risk as a function of applicant age, if diagnostic is recent or not (three cases).

As we can observe, the processes of obtaining premiums for type 1 diabetes and breast cancer were relatively similar. This suggests that the same process can be conducted for all severe diseases, in the worst case grouping them, such as grouping type 1 and 2 diabetes or taking prudent approximations, such as taking long-term disability risks associated with cancers instead of breast cancer. The process we conducted is the one described in

Section 3. It is an example of the process and other means to estimate risks based on open data that can lead to the best estimations, such as adjusting disability risks based on the risk of mortality.

5. Conclusions

This study has demonstrated that, based on open data, it is possible to significantly refine the description of the risks of mortality and disability for serious pathologies that currently lead to numerous insurance refusals. A more precise assessment of these risks makes it possible to include more people in the insurance with excess risks, which allows the premium to be kept to a reasonable level.

The methodology presented makes it possible to extend the work to many other diseases, even if each disease requires significant bibliographic and information cross-checking efforts. The examples of diseases taken suggest it is sometimes possible to estimate risks related to great details for some pathologies, such as the exact grade of and oestrogen receptor function of a breast cancer, and that it is sometimes not necessarily feasible except by taking prudent approximations, such as using all-cancer long-term disability risk instead of breast cancer. This means that it is possible to cover a large range of diseases without taking strongly prudent approximations if one accepts that the granularity of disease definition is sometimes coarse, depending on the found open data.

A difficult question was how to transpose risks from the general population to high socio-economic profiles. We defined this transposition via the definition of a multiplier that interestingly permits us to transpose studies that link social inequalities and mortality risks into loan insurance premiums—a different field. We were able to suggest that the multiplier would tend to be slightly smaller than one when mass screenings exist for the studied disease, and slightly greater than one otherwise (and if the disease is a frequent one). This was perfectly verified in the examples we considered.

Beyond the technical aspects of being able to use a diverse set of open data to generate loan insurance premiums comes the question of how to pool diseases to facilitate access to loans. The results we observed for breast cancer and type 1 diabetes suggest that non-metastatic breast cancers could be covered as a whole, as well as type 1 diabetes in the absence of nephropathy.

For even greater inclusion of patients, a framework would be needed to mutualize severe risks with low risks, e.g., by covering all cases of breast cancer and diabetes or even by eliminating medical selection. The latter option would dilute the excess risk across all insured individuals, with a potential but limited increase in basic premiums as we find in part 1, but runs the risk of high risks being concentrated in the first insurer or insurers to try this.

Notes

[1] For insured risks this gender breakdown is based on expert opinion, as the data at our disposal does not distinguish between the sexes.

[2] Average, maximum and minimum of best estimate actuarial tables among a group of French insurer for the mortality and temporary disability incidence of borrowers. Female risks were obtained by dividing by 1.5 (expert judgement) and male risks where deducted by considering that the tables contained 55% males (expert judgement).

[3] Since the conditional transition probabilities depend on the time spent in the state, semi-Markovian models are better suited, but given the context of this study, this simplification is acceptable.

[4] Surveillance Research Program, National Cancer Institute SEER* Stat software (seer.cancer.gov/seerstat) version 8.3.9.

[5] In blue and red for men and women respectively. The solid lines represent those without kidney disease. The numbers associated with the curves (15, 30 and 45) are the ages at diagnosis.

References

Ameli. 2018a. «Fiches sur les Pathologies» Available on the Website Ameli.fr. Available online: https://assurance-maladie.ameli.fr/etudes-et-donnees/entree-par-theme/pathologies/cartographie-assurance-maladie/donnees/fiches-pathologies/fiches-pathologies (accessed on 5 November 2021).

Ameli. 2018b. *Femmes prises en charge pour cancer du sein actif*. Fiche Pathologie. Available online: https://assurance-maladie.ameli.fr/etudes-et-donnees/cartographie-fiche-cancer-sein-femme-actif (accessed on 5 November 2021).

Ameli. 2018c. *Femmes prises en charge pour cancer du sein sous surveillance*. Fiche Pathologie. Available online: https://assurance-maladie.ameli.fr/etudes-et-donnees/cartographie-fiche-cancer-sein-femme-sous-surveillance (accessed on 5 November 2021).

Ameli. 2018d. *Diabète*. Fiche Pathologie. Available online: https://assurance-maladie.ameli.fr/etudes-et-donnees/cartographie-fiche-diabete (accessed on 5 November 2021).

Bagui, Hasna. 2013. *Refonte des lois de maintien en incapacité temporaire de travail*. Mémoire de Master. ISFA. Available online: http://www.ressources-actuarielles.net/C12574E200674F5B/0/CA2755D9E71E059AC1257C78006164B9 (accessed on 20 February 2022).

Blanpain, Nathalie. 2016. *Les inégalités sociales face à la mort. Tables de mortalité par catégorie sociale et par diplôme*. INSEE, Résultats. Available online: https://www.insee.fr/fr/statistiques/1893092 (accessed on 20 February 2022).

Blanpain, Nathalie. 2018. *Tables de mortalité par niveau de vie*. INSEE, Résultats. Available online: https://www.insee.fr/fr/statistiques/3311422 (accessed on 20 February 2022).

Cuerq, Anne, Michel Païta, and Philippe Ricordeau. 2008. *Points de repère Ameli n°16: Les causes médicales de l'invalidité en 2006*. Available online: https://assurance-maladie.ameli.fr/etudes-et-donnees/2008-causes-medicales-invalidite-2006 (accessed on 20 February 2022).

Dabakuyo, Tienhan Sandrine, Franck Bonnetain, Patrick Roignot, Marie-Laure Poillot, Gilles Chaplain, Thierry Altwegg, Guy Hedelin, and Patrick Arveux. 2008. Population-based study of breast cancer survival in Cote d'Or (France): Prognostic factors and relative survival. *Annals of Oncology* 19: 276–83. Available online: https://www.annalsofoncology.org/article/S0923-7534(19)41345-8/fulltext (accessed on 20 February 2022).

Dray-Spira, Rosemary, Tiffany Gary-Webb, and Frederick Brancati. 2010. Educational disparities in mortality among adults with diabetes in the US. *Diabetes Care* 33: 1200–5. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2875423/ (accessed on 20 February 2022).

Gentil-Brevet, Julie, Marc Colonna, Arlette Danzon, Pascale Grosclaude, Gilles Chaplain, Michel Velten, Franck Bonnetain, and Patrick Arveux. 2008. The influence of socio-economic and surveillance characteristics on breast cancer survival: A French population-based study. *British Journal of Cancer* 98: 217–24. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2359707/ (accessed on 20 February 2022).

Jørgensen, Marit Eika, Thomas Peter Almdal, and Bendix Carstensen. 2013. *Time Trends in Mortality Rates in Type 1 Diabetes from 2002 to 2011*. Available online: https://pubmed.ncbi.nlm.nih.gov/23949580/ (accessed on 20 February 2022).

Kim, Nam Hoon, Tae Joon Kim, Nan Hee Kim, Kyung Mook Choi, Sei Hyun Baik, Dong Seop Choi, Yousung Park, and Sin Gon Kim. 2016. Relative and combined effects of socioeconomic status and diabetes on mortality: A nationwide cohort study. *Medicin* 95. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5265873/ (accessed on 20 February 2022).

Kusnick-Joinville, Odile, Céline Lamy, Yvon Merlière, and Dominique Polton. 2006. *Points de repère Ameli n°5: Déterminants de l'évolution des indemnités journalières maladie*. Available online: https://assurance-maladie.ameli.fr/sites/default/files/2006-11_determinants-evolution-indemnites-journalieres-maladie_points-de-repere-5_assurance-maladie.pdf (accessed on 20 February 2022).

Maruani, Margaret, and Monique Meron. 2012. *Un siècle de travail des femmes en France*. Paris: La Découverte.

NHIS (National Health Survey). 2016. Centers for Disease Control and Prevention. Available online: https://www.cdc.gov/nchs/nhis/1997-2018.htm (accessed on 20 February 2022).

Piffaretti, Clara, Anne Fagot-Campagna, Grégoire Rey, Juliana Antero-Jacquemin, Aurélien Latouche, and Laurence Mandereau-Bruno. 2016. Déterminants de la mortalité des personnes diabétiques de type 2. Cohortes Entred, France, 2002–2013. *Bulletin Epidémiologique Hebdomadaire-BEH* 681–90. Available online: https://www.santepubliquefrance.fr/maladies-et-traumatismes/diabete/documents/article/determinants-de-la-mortalite-des-personnes-diabetiques-de-type-2.-cohortes-entred-france-2002-2013 (accessed on 20 February 2022).

Ridsdale, B. 2012. Annuity underwriting in the United Kingdom. Note for the International Actuarial Associ-ation Mortality Working Group. Available online: https://www.actuaries.org/CTTEES_TFM/Documents/Zagreb_item19_underwriting_annuities_UK.pdf (accessed on 20 February 2022).

SEER. 2020. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence—SEER Research Data, 9 Registries, Nov 2020 Sub (1975–2018)—Linked to County Attributes—Time Dependent (1990–2018) Income/Rurality, 1969–2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Released April 2021, Based on the November 2020 Submission. Available online: https://seer.cancer.gov/seerstat/ (accessed on 5 November 2021).

Tomas, Julien, and Frédéric Planchet. 2014. Prospective mortality table and portfolio experience. In *Computational Actuarial Science with R*. Edited by A. Charpentier. The R Series; Chapman and Hall: chp. 9. Available online: http://www.ressources-actuarielles.net/gtmortalite (accessed on 20 February 2022).

Walter, Jeremy, Shona Livingstone, Helen Colhoun, Robert Lindsay, John McKnight, Andrew Morris, John Petrie, Sam Philip, Naveed Sattar, Sarah Wild, and et al. 2011. Effect of socioeconomic status on mortality among people with type 2 diabetes: A study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetes Care* 34: 1127–32. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3114515/ (accessed on 20 February 2022).