ACTUARIAL COLLOQUIUM PARIS 2020

Prospective modelling of Disability risk: Proposal for a two-dimensional model and Machine Learning algorithms combined approach

Fatoumata NDOYE[1]

SUMMARY

The main objective of this article is to propose a model for the compensation rate after a temporary disability (or temporary incapacity to work) in France based on the foundations of the LEE-CARTER model and national data.

We have structured our work in three main stages.

The first stage is a qualitative analysis of the disability risk in general (temporary incapacity to work in particular) where we present the coverage system established by the Social Security System and insurance organizations, and then a summary of its evolution over the last few years.

The second stage consists of a presentation of the data to assess the daily compensation rate by prescriber's medical specialty, followed by a proposal to model its evolution by combining two-dimensional modelling (similar to that of LEE-CARTER) and statistical learning.

In the third stage concerning the management of the temporary disability risk, we propose an approach to improve its monitoring in an ORSA framework for an insurance organization, but especially develop customized prevention solutions.

Keywords: *Temporary incapacity to work, health insurance, evolution, open source data, prospective modelling, LEE-CARTER, Artificial Intelligence, statistical learning, Boosting, interpretation, Shap Value, Tree interpreter, ORSA, prevention.*

CONTENT

[1] Actuary IA, expert ERM CERA – Consultant Manager – FIXAGE Consulting Firm

PREAMBLE

Over the past decade, we have identified a significant paradigm shift in disability claims: an increase in claims related to psycho-social risks and MusculoSkeletal Disorders (MSD). Economic, social and regulatory changes have also changed the relationship of people to work and their protection against incapacity and invalidity risks is becoming a major element for the French government and insurance companies.

Given this context, the expertise adapts or develops themselves to understand, measure and even reduce risks. The theoretical and practical basis for risk analysis and management have been changing for several years, with more recent use frameworks. Evaluation techniques evolve and are constructed in parallel with the application subjects under consideration, resulting in a multiplicity of monitoring and control tools. In spite of this evolution of evaluation techniques and in view of the volatility of disability risk, we note in some insurers a risk aversion, especially when it is hidden: risk-taking, based on data from the past, cannot always be delineated if the context or the population under risk changes significantly.

The assessment of the risks borne by insurers has certainly reached an important level of requirement and control, sometimes putting a brake on economic development and customized solutions, but the most fearful risks are those we cannot know and measure. It is necessary to reconcile qualitative and quantitative approaches in order to have a wider and global view of insurance risks in general, disability risk in particular by showing:

- The analysis of the accepted

- The evaluation of the unknown and/or the not accepted

Among the quantitative approaches, we have Artificial Intelligence (AI) algorithms that represent many challenges, technological and ethical, among which we note data bias and transparency.

In this article we focus on describing the disability risk, more particularly the temporary incapacity risk.

After a brief summary of the risk coverage by Social Security System and insurance institutions and its evolution, we give the some underlying issues in terms of estimation.

We propose a modelling of the compensate rate after a disability claim based on the LEE CARTER model and national data.

Finally, we will propose solutions to facilitate good disability risk management and the implementation of customized solutions.

1. ANALYSIS OF DISABLITY RISK

1.1. The specificities of disability risk

1.1.1. The Incapacity / Invalidity distinction

The first specific feature of work stoppage is the legal distinction between incapacity and invalidity arising from Social Security law.

Incapacity is a temporary (partial or total) inability to engage in an professional occupation. Invalidity is a permanent (partial or total) reduction in certain abilities.

Invalidity is considered as an extension of disability, with an identical cause in both cases for a claimant, when it is not immediate.

These situations lead to a drop in income to which Social Security and supplementary insurance intend to compensate. The guarantees offered relate to all or part of the shortfall. Benefits paid as long as the incapacity is temporary are Daily Allowances (DA). Once the state of invalidity is consolidated, the claimant receives as benefit an invalidity pension.

Important Notes:

- A work stoppage is a medical prescription issued to an employee whose condition of health requires him to interrupt his professional activity.
- Total and Temporary Incapacity to Work or "TIW" is the unit of measurement used in criminal law to quantify injuries suffered by a claimant. Any medical is entitled to prescribe TIW, and it can also be done in the absence of physical injury if the claimant has suffered significant psychological trauma. A medical certificate for a TIW following an assault is in no way equivalent to a work stoppage and therefore cannot legally justify an absence from his employer if a work stoppage has not been prescribed by the doctor. Since the concepts of TIW and work stoppage are completely distinct, they often do not coincide.
  Example: It is possible that the doctor examining a victim of minor violence may assess the TIW for 2 days, but consider that the psychological impact justifies a five-day work stoppage.

In our study, we discuss the subject of work stoppages prescribed by health professionals and subsequent to a TIW.

1.1.2. The causes of disability

We distinguish five causes of work stoppage: personal injury or illness, maternity, work accident or occupational disease.

We briefly describe some of the causes below.

An accident on the way to or from work or road accident is an accident sustained by an employee during the journey between his place of residence and his place of work, outside the usual working hours. The accident may also occur between the workplace and the business restaurant. The road must naturally be direct and without interruptions, except detours justified by the necessities of daily life.

The article L 411-1 of the Social Security Code considers as a work accident, whatever the cause, an accident which occurred by or in connection with work to any employee or worker, in any capacity or place, for one or more employers or entrepreneurs.

According to the Social Security Code, any disease designated in a table of occupational diseases is presumed to be of occupational origin and contracted under the conditions mentioned in this table. These data cover many diseases such as cancers, intoxications, respiratory or skin diseases, etc. The tables also determine working conditions (use of chemicals, repetitions of movements, etc.) in which the disease must appear.

Nevertheless, a disease may be considered to be occupational even if it does not appear in the table or does not fulfil all the conditions required.

The following sections describe the benefits paid by the SS and the insurance organizations.

1.1.3. The Social Security benefits

SS benefits vary in form and amount depending on the cause of the disability.

The beneficiaries of benefits paid by health insurance are the insured persons. In order to qualify as a social insured person, the person must be insured under the general scheme. The conditions for entitlement to the daily subsistence allowance are assessed at the time of the work stoppage. Cash benefits, provided in the form of a DA, constitute a replacement income for claimant.

The summary information on SS benefits paid as a result of temporary incapacity is given in the annexes, section **Erreur ! Source du renvoi introuvable.**.

The summary information on the employer's benefits is given in the annexes, section **Erreur ! Source du renvoi introuvable.**.

1.1.4. The additional services of insurance companies

Supplementary organizations which supplement the SS indemnities shall pay benefits in accordance with the arrangements of the contract or the regulations governing the relationship between the insured and the insurer. We distinguish individual and collective contracts that can be implemented:

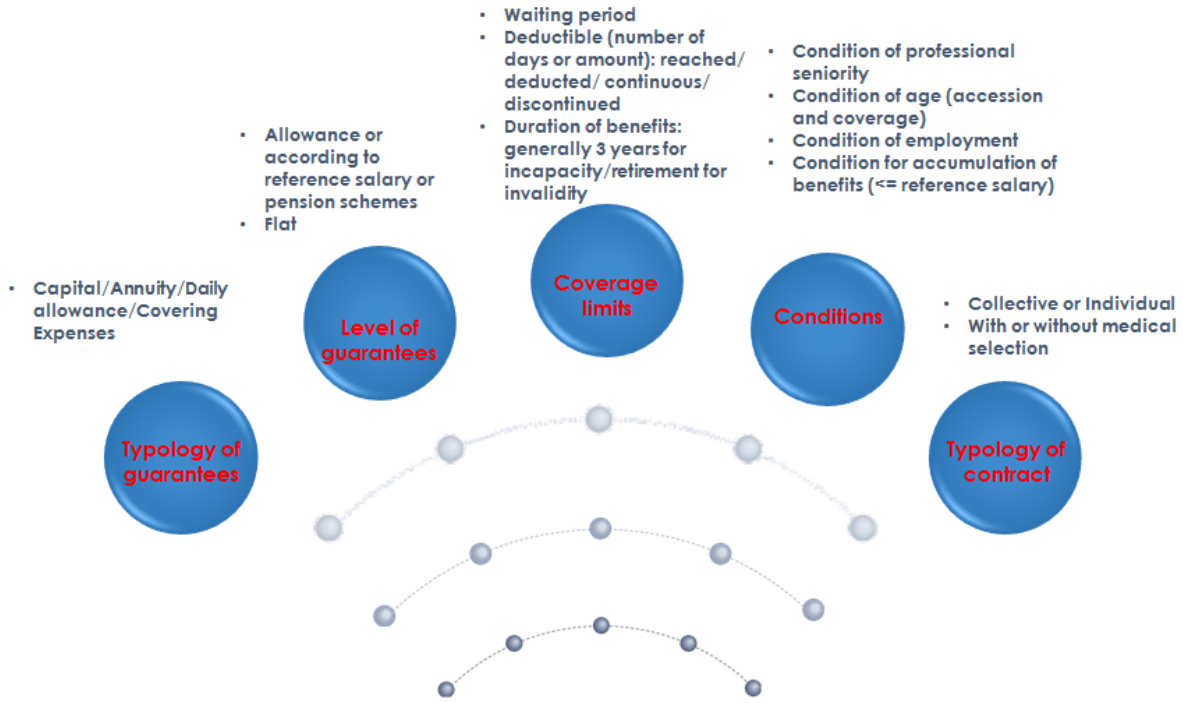
- By collective agreement (collective or company agreement),
- By referendum on a draft agreement proposed by the head of the enterprise,
- By unilateral decision of the employer.

In these three cases, the group covered must be defined objectively:

- In the case of a collective contract with compulsory membership, all or almost all employees are involved (exempt: beneficiaries of Universal Medical Coverage, employees on fixed-term contracts, seasonal workers and employees with multiple employers).
- In the case of an optional collective agreement, the employees concerned are free to join or not.

The coverage offered by insurance companies is not standardized, so we observe different types of contracts or agreements that depend on the population at risk.

The services are characterized in the following graph:



1.2. The drivers of disability risk at national level

The National Agency for the Improvement of Working Conditions (NAIWC) updates annually (since 2012) a gender analysis of work accidents, commuting accidents and occupational diseases based on data provided by the National Sickness Insurance Fund (NSIF).

In the private sector, an extensive study was carried out by the firm Ayming in 2017 (10th barometer of absenteeism and commitment) with 46540 companies employing 1.836.802 employees. It shows that the average time spent in work stoppage is 17.2 days/year in 2017 (16.8 days/year in 2016). The absenteeism rate is 4.72%/year.

1.2.1. The age

Seniors stop more often and longer than young people. They are often affected by heavier pathologies, resulting in longer downtimes.

In the private sector

The absenteeism rate for those over 56 is 7.11% in 2017.

On the other hand, the age of the employee must be compared with his seniority in his company: from five years of seniority, the rate of absenteeism decreases by 30%.

On the other hand, the decline in the retirement age has an indirect effect on the level of risk since exposure to risk increases (in incidence and duration).

1.2.2. The gender

Women stop more frequently than men, mainly because of the arduousness of the tasks associated with jobs they occupy more often. They also manage more family and household responsibilities than men. They are more single parents than men.

Work and road accidents have been increasing since 2001 among women, even though their number is half that of men and there has been an overall decrease in these accidents.

The development of occupational diseases is twice as rapid for women as for men.

It should be noted that the modification of the MSD recognition tables (the leading cause of occupational diseases) has led to a decrease in claims related to these diseases.

The increase in claims for women is greater in mixed and predominantly female sectors (temporary work, social work, health, cleaning, banking, insurance, administration, etc.).

Exposure to occupational risks is thus different according to gender since a large part of the sectors have non-mixed jobs or different conditions for performing tasks according to gender.

In the private sector

The absenteeism rate for women was 5.30% in 2017, compared to 3.54% for men.

1.2.3. The profession

Work accidents are more common in:

- Service activities (temporary work, social work, health, cleaning) for women,
- Construction and public works industries for men.

Occupational diseases (especially MSD) are more important among manual and non-manual workers.

In the private sector

Health (private sector - excluding public hospitals), industry and commerce are the sectors with the highest level of work stoppage in 2017.

This is explained by the increasing number of MSD, particularly in the health sector. Employees in the latter sector are forced to manipulate patients, are often in standing stations and often move within their establishment. The labor force does not renew as often as in other sectors. They are older than average, which has an impact on both the incidence and duration of the work stoppage.

In the industrial sector, it is the arduousness (with the carrying of heavy loads) that makes the level of work stoppages high. Increasing automation/robotization of tasks helps to limit negative effects.

In the commercial sector, the frequency of work stoppages is increased by the devaluation of jobs and competition from online businesses.

In the building sector, psychosocial risks have led to an increase in the level of risk. According to occupational physicians, this is related to pressure to meet deadlines and procedures. Competition and the proliferation of statutes and contracts have also increased the level of pressure. Dermatological and cardiac diseases, low back pain are pathologies related to building trades.

The mechanization of tasks has transformed dynamic efforts into static ones.

The level of risk for service businesses is decreasing.

1.3. The main causes of work stoppages

1.3.1. The main causes of work and road accidents and occupational diseases

The main causes of work accidents in 2015 and 2016 are:

- Manual handling, 53% of cases;
- Falls on the same level, 13% of cases;
- Falls of height, 12% of cases;
- Hand tools, 9% of cases.

The main causes of road accidents in 2015 and 2016 are:

- Loss of control of a conveyance, 60% of cases;
- Fall or misstep, 25% of the cases.

The main causes of occupational diseases in 2015 and 2016 are:

- MSD, 87% of cases. In the 42 535 MSD identified, the most common are peri-articular, chronic lumbar spine caused by heavy load handling ;
- Asbestos-related diseases, 7% of cases.

1.3.2. The psychosocial risks

Psychosocial risks have increased over the past 20 years, causing an increase in the level of risk.

Pressure to respect time and procedures, increased precariousness and work problems (harassment) increase work stoppages.

Psychosocial constraints are also involved in MSD. These factors are based on how work is perceived by employees as:

- Dissatisfaction with monotonous work,
- The tension caused by the deadlines to be respected,
- The lack of professional recognition,
- The degraded social relations,
- Lack of support from line manager and colleagues,
- The insecurity of employment.

In the private sector

We can see domino effects: employees who are in under-staffed services (as a result of a colleague's absence from work) work overtime, cannot take time off, and then go on sick leave.

This can also lead to an increase in work accidents.

Organizational changes and long absences are a source of stress and anxiety.

Even if companies are aware of the problems, improvements in working conditions can take time.

1.4. Summary

Statistical analyses of national and insurance data highlight several risk factors related to the profiles of the claimant, as well as to the economic, social and regulatory context.

- The characteristics of the claimant: the age reached at the occurrence of the claim, the SPC (socio-professional category), the gender, the geographical situation, the marital situation.
- The characteristics of the claim: the cause of the claim (sickness, accident, work accident).
- The characteristics of the guarantees of the supplementary insurance contract: the deductible, the waiting period, the type of compensation (lump sum/indemnity).
- The socioeconomic-regulatory context: the situation of the affected person in his enterprise, the exposure to risk, and the existence of supplementary pension plans.

It is difficult to assess the quantitative impact of these factors which would require quality data and adequate statistical methods.

2. PROSPECTIVE MODELLING OF TEMPORARY DISABILITY RISK

We wish to study the evolution of the temporary disability risk. To do this, we have structured our approach as follows:

- Selection and analysis of study data, including through descriptive statistics. The data used are those relating to health insurance expenditure. We deduce a study variable, the compensation rate of TIW, and the explanatory variables. This annual compensation rate is equal to the ratio between the number of days covered by health insurance after a TIW established in a given year to the French population at risk at the beginning of that year ;
- Classify the explanatory variables of the studied variable using a Boosting Machine Learning algorithm. The use of this algorithm also allows the compensation rate to be predicted based on the values of the explanatory variables ;
- Predict the specialty of the TIW prescriber to replace the un-filled values of this variable. We will also use the same Boosting Machine Learning algorithm ;
- Model the compensation rate of TIW, evaluated for a given age group, using an approach similar to LEE-CARTER model: the rate changes according to the specialty of the TIW prescriber and a temporal component.

2.1. Treatment of disability risk under Solvency II

The Solvency II reform has introduced, in an economic balance sheet, an estimation of technical provisions and a valuation of cash flows with Best Estimate assumptions, whereas they were previously estimated with conservative assumptions defined by the insurance code. Insurers are thus encouraged to use their best possible knowledge of their portfolios to model liability cash flows in stress scenarios such as financial markets ones.

Temporary and permanent disability risks are assessed using the following probability laws[2]:

- Temporary disability incidence laws;
- Persistency risk in temporary disability laws;
- Temporary to permanent disability transition laws or permanent disability incidence laws;
- Persistency risk in permanent disability laws.

These laws of probability will also be used to simulate the policyholders' claims within a stress environment as described in the SCR (Solvency Capital Requirement) calculations. In fact, the disability risk is completely apart from other risks requiring the implementation of a SCR calculation: this sub-module is the "Disability Risk" within the underwriting risk module. The "Disability" SCR corresponds to the increase in Best Estimate relative to:

- Increase of 35% in disability rates for the first year and a 25% increase in following years;
- Decrease of 20% in recovery rates;

Solvency II also introduces ORSA (Own Risk and Solvency Assessment). This is a forward-looking risk and solvency assessment process that allows the risk dimension to be integrated into the management of the company. It is carried out over 3 or 5 years and may correspond to the company's business plan.

The prospective assessment of the disability risk emerges a problem of adaptation in terms of the assumptions' appropriateness in provisioning, valuation and pricing. It can be delicate to keep the

[2] The compensation rate referred to in the following sections is the equivalent product between the incidence rate and the maintenance duration (resulting from the multiplication of maintenance rates).

estimation methods which is established from the last x years if they contain periods of stress or if we think that the evolution or the level of claims would be different in the medium term.

We are faced with the **risk of misestimating the future evolution of risk**, implying a **non-consistance of the ORSA scenarios, an incorrect quantification of the ORSA capital** for the underwriting risks and a **non-compliance with provisioning and pricing assumptions**.

In addition, the workers compensation cover may be subject to numerous legal, regulatory, contractual or socio-economic changes: we are also in a constantly evolving insurance market that lives in its social context.

We face the **risk of poor understanding of socio-economic and regulatory developments.**

In the following, we would like to propose a model regarding the evolution of the compensation rate related to a disability claim, based on social security data. The objective is to be able to apprehend the crude claims data (no filters linked to the ownership and characteristics of a benefit or income protection insurance contract).

The objective of this modelling work is to reduce the risk of misestimating the future development of temporary disability risk.

2.2. Presentation of study data

We used for our study the Open Damir data: it regroup the expenses of cross-plan health insurance and come from the national cross-plan System of Health Insurance (SNIIRAM) covering all health insurance reimbursements for all plans (including hospital services invoiced directly to health insurance).

The expenditure is detailed according to six lines of analysis (period, benefit, care institution, beneficiary of care, performing health professional, prescribing health professional) and seven indicators of amount (total expenditure, basis of reimbursement, amount reimbursed, extra billing) and volume (count, quantity, coefficient).

Geographical areas are grouped together to preserve the anonymity of health professionals and claimants.

In addition, we used data related to the French population (from INSEE) as a basis for the population at risk. They give the number of habitant by French region, gender and age group.

2.2.1. The variables studied

We have studied the following Open Damir variables:

- AGE_BEN_SNDS: Beneficiary age band at time of care;
- BEN_SEX_COD: Beneficiary Gender;
- BEN_RES_REG: Beneficiary's Region of Residence;
- PRS_NAT: Nature of Service. The modalities studied are the following :
    - 6013 : « IJ WAITING PERIOD (CRPCEN) »
    - 6014 : « SUPPLEMENT IJ > CEIL (CRPCEN) »
    - 6110 : « IJ NORMAL + 6MOIS »
    - 6111 : « IJ NORMAL -3 MOIS »
    - 6112 : « IJ NORMAL +3MOIS »
    - 6113 : « IJ REDUCED -3MOIS »

  - 6114 : « IJ REDUCED +3MOIS »
  - 6115 : « IJ INCREASED -3MOIS »
  - 6116 : « IJ INCREASED +3MOIS »
  - 6117 : « IJ PARTIAL, SALARY LOSS -3MOIS »
  - 6118 : « IJ PARTIAL, SALARY LOSS +3MOIS »
  - 6119 : « IJ INCREASED + 6 MOIS »
  - 6120 : « TEMPORARY INCAPACITY BENEFIT WA/OD »
  - 6134 : « IJ IMPA DISEASE - Insured Medical Practitioners and Auxiliaries - LESS THAN 3 MONTHS »
  - 6135 : « IJ IMPA DISEASE MORE THAN 3 MONTHS »
- ASU_NAT: Nature of Insurance;
- EXO_MTF: Reason for exemption of the Moderator Ticket;
- CPT_ENV_TYP: Benefits envelope Type;
- PRS_ACT_QTE: Quantity of benefits (number of days of incapacity paid);
- PSP_SPE_SNDS: Medical Speciality PS Prescriber;
- ETP_CAT_SNDS : Description of Category Etb Prescriber;
- FLX_ANN_MOI: Year and Month of Treatment of the benefit;
- SOI_ANN: Year of Care;
- SOI_MOI: Month of Care.

The modalities of the different variables are given in annexes (section **Erreur ! Source du renvoi introuvable.**).

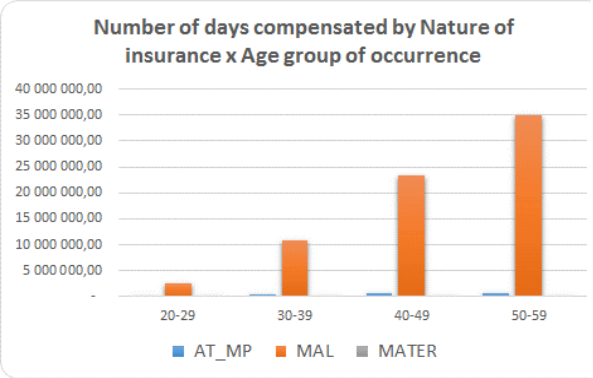
Note: When using crude indicators such as "PRS_ACT_QTE", it is necessary to filter the repayment type (the variable "PRS_REM_TYP") by selecting modalities 0 and 1.

The number of DA related to a claim is obtained by adding the variable "PRS_ACT_QTE" by processing date; this allows to obtain the number of days compensated following a claim occurred in the year.

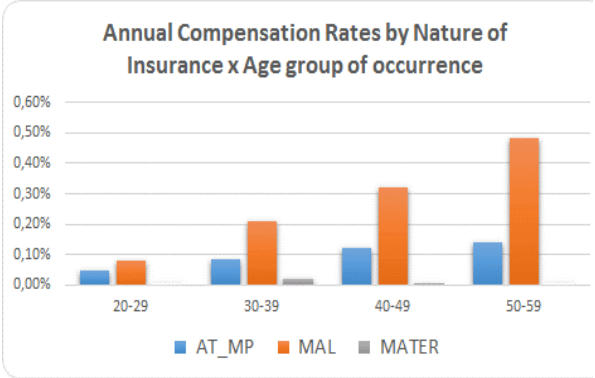
2.2.2. Open Damir data quality and descriptive statistics

Some data are not present in the Open Damir database, such as the SPC, marital status or income group. In addition, some variables are present but are not correctly entered, in particular the variable "ATT_NAT" which must indicate the nature of the work accident.

The specialty of the TIW prescriber is not known for about **15.3%** of the lines (**12.32%** between 2015 and 2018). The region is not filled in for about **16.34%** of the lines before 2015. It is fully filled in from 2015. The following graphs give some statistics on the distribution of claims according to the most significant variables.

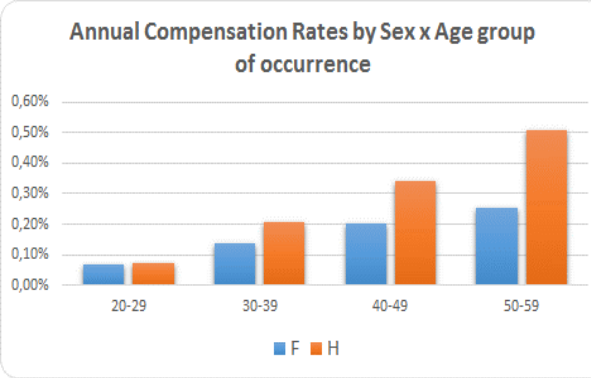


Distribution of claims by insurance type and age group at occurrence (historical 2010-2018)

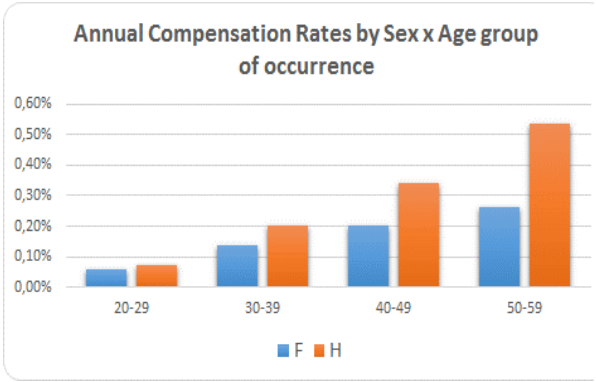


Annual compensation frequency by insurance type and age group at occurrence (historical 2010-2018)

The number of TIW claims increases with age as seen in section **Erreur ! Source du renvoi introuvable.**. This increase in claims is more explained by privacy diseases ("MAL").



Annual compensation frequency by gender and age group at occurrence (historical 2010-2018)



Annual compensation frequency by gender and age group at occurrence (historical 2015-2018)

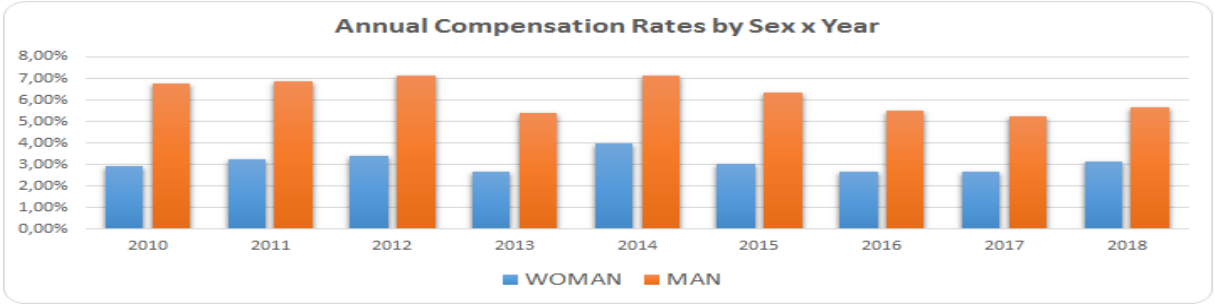
The dynamics of DA are linked to several structural factors:

- Wage developments, which are the basis for calculating DA;
- The development of employment, which acts mechanically on the volume of DA with a delayed effect;
- The share of older people in the labor force. The latter have more serious health problems, which require longer TIW than younger employees;
- Precarious employment, as employees with fixed-term contracts or on probation are significantly less absent than others.

The bumps in the evolution of the DA can be explained by the striking pace of the volume of sickness and "WA-OD" benefits in the general scheme:

- The slight decline from mid-2011 to the end of 2013 is linked to two factors: the deterioration of the labor market and the favorable epidemiological context in 2011 and 2012;
- Despite a less pronounced flu epidemic than in 2013, the volumes of DA resumed in 2014. Therapeutic time, the easing of the conditions for granting rights to long-term DA and the creation of "disease" DA for farmers;
- We also see in 2015, as in 2014, a volume effect significantly higher than in previous years. This growth is mainly driven by the seasonal flu epidemic, which is more pronounced than in 2014.

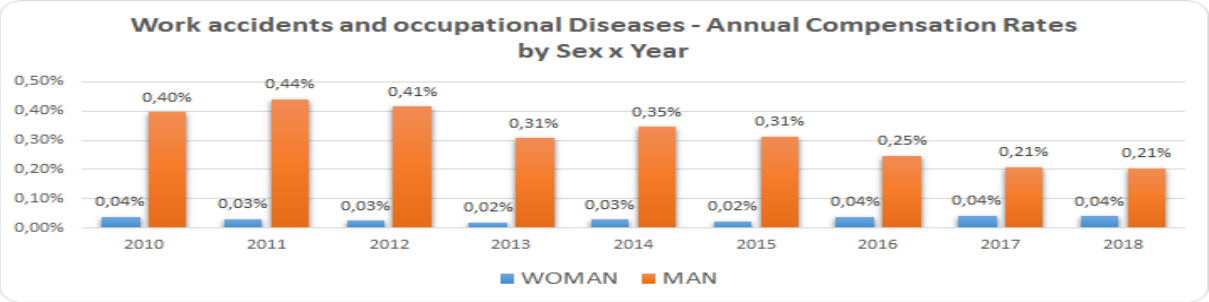
Claims are higher for men than women for all age groups. This difference increases with age, with a gap of almost 50% between the ages of 50 and 60.

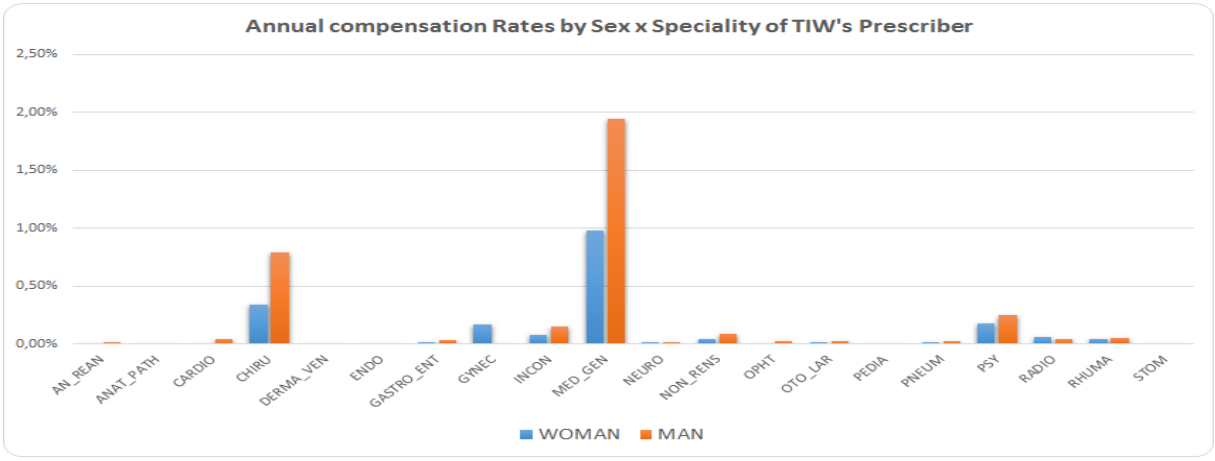


Annual compensation rates by gender and year

The overall decline in claims between 2014 and 2017 was particularly marked among men.

By focusing on work accidents and occupational diseases, the drop in claims among men is more marked but hides an increase among women as we can see on the following graph:



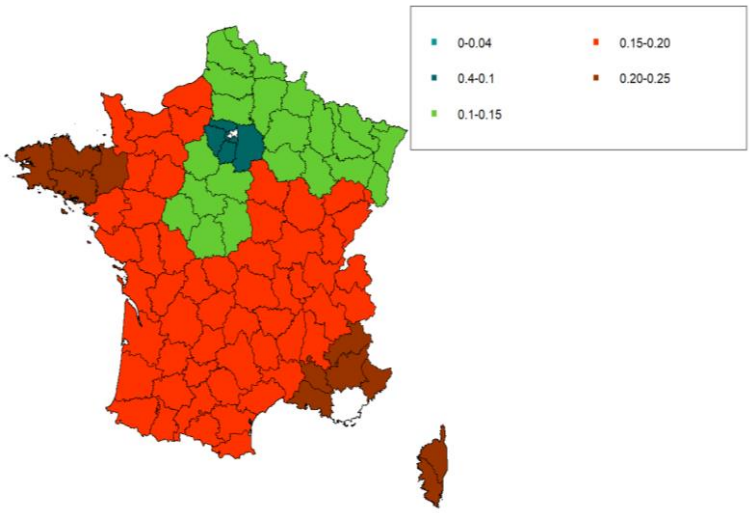


Annual compensation rates by gender and specialty of TIW's prescriber (2015-2018 History)

Claims for TIW are mainly prescribed by specialists in the following disciplines: General Medicine, Surgery, Psychiatry, Radiology and Rheumatology. Unfortunately, almost half of prescribers' specialties are unknown in the database, but the proportion of claims affected is less than 20%.

The regional vision

The following graph (this is an area) gives the 2018 compensation rate by major French region:



The North-East has a higher level of sinistrality because of its industry and a less favorable level of health (the highest rates of cancer in France or cardiovascular diseases).

The Southwest has had a high level of unemployment for several years, dissatisfaction at work is high: the TIW becomes a recourse to lack of interest in his work.

The compensation rate is the lowest in Île-de-France: This may be due to the high pressure of work.

2.3. Presentation of the models used

2.3.1. The Machine Learning algorithms

Artificial intelligence and assurance: between understanding, interpretability and use

Statistical learning makes it possible to construct a functional relationship from data by a computer; this relationship is the model. When supervised, it can predict an event from its descriptive data. It has been used for several years in life and non-life insurance to assess risk, manage it and/or improve profitability.

Statistical learning has evolved in recent years with the arrival of new methodologies such as neural networks and Boosting, but also with advances in computing. One of the challenges of the modeler is to obtain a model that is interpretable, understandable and takes into account the structure and limitations of the data (including bias). It is important to understand how the data are obtained, to study and reprocess them, and not just to study correlations.

A growing complexity with AI is that of interpretability. As simple as it was able to justify that the gender was correlated with the TIW claim, when only this criterion and age were taken into account. But when we are on voluminous data, both in rows and columns (multiplicity of discriminating variables), generalization or “stereotyping” becomes more delicate and undermines the principle of customization, since we can no longer talk about an average individual.

The risk groups obtained by the algorithms may be meaningless, but the results obtained will still have to be explained to the project sponsors.

Understanding the model is the second issue: the deep learning algorithms, which abound in areas of application such as image recognition and text analysis, are difficult to understand, since they are considered as black boxes.

In statistical evaluations, these models can predict the level of probability (to price or provisioning), but they are not yet used to give reasons for understanding the origin or estimating the amount of a claim. Also, just because your model gives you multiple risk groups doesn’t mean you have to segment your study population. Two of the limitations of segmentation are the loss of the « pooling effect » and the increase in estimation error due to sampling fluctuations.

In fact, a model introducing a lot of "overfit" or over-adjusted segmentation, does not generalize enough. We should therefore be able to find a balance between pooling and segmentation. The ideal would be to use AI for insurance by:

- Avoiding some pitfalls such as over-learning and lack of interpretability.
- Identifying cause-effect relationships to better target prevention, and indirectly reduce claims.

The Boosting

We used a Boosting supervised learning algorithm, the eXtreme Gradient Boosting (XGBOOST), based on tree aggregation. It is a whole technique which consists in aggregating classifiers (models) developed sequentially on a learning sample whose individuals’ weights are corrected as they go. Classifiers are weighted according to their performance. The XGBOOST uses the loss gradient function to calculate individual weights when constructing each new model. It’s a bit like gradient descent for neural networks.

It is first and foremost a pragmatic approach which makes it possible to manage problems of regression and classification (like ours, since we will try to classify unknown values in the classes of known values). As with the GBM (Gradient Boosting Machine), the algorithm works sequentially, allowing it to improve

by capitalization compared to previous runs. So it starts by building a first model that it will overestimate. From this first evaluation, each line of data will then be weighted according to the performance of the prediction. XGBOOST therefore behaves remarkably in machine learning competitions, but not only because of its sequential self-improvement principle, as it includes a large number of hyper-parameters that can be modified and adjusted for improvement. In fact it has some specificities making the learning fast and able to handle large volumes of data like that of Open Damir.

- It allows to learn a random forest;
- Each node is a threshold on a variable, this threshold is chosen from all possible values or quantiles, these quantiles being fixed for a tree;
- Quantiles used to determine thresholds can be adjusted for each node;
- A compression of the values of the variables by columns is done to reduce the memory footprint is performed;
- Missing values are not treated as other values. Each node of a tree has a default value given to missing values;
- The treatments are parallel.

Algorithms developed for model interpretation

In this section, we will give the principles of two approaches used to interpret the outputs of machine learning algorithms, one specific to decision tree models and an agnostic model to models, which allow predictions' explanations. It should be noted that these approaches, even if they can provide elements of interpretability, can also be complex.

The **first approach** is the interpretation tree (Tree Interpreter) which is a set of nodes forming a passage for each observation. This approach allows the variables that are involved in the decision to be developed while giving weight to them, thereby providing a rationale for a decision. Unfortunately, when there are a large number of explanatory variables, we will have many nodes and the explanation will be long. Furthermore, we do not have an explanation of why such a threshold was chosen for a given node.

The **second approach** is the Shap method, which allows post explanations of the variable's contribution for all models. The Shap algorithm applies minor modifications to the model's input and looks at the impact of these changes on predictions.

This implies that this algorithm favours strong signals that have a strong impact on the decision, whereas it is weak signals that improve the performance of certain algorithms. Two observations may have the same shap value represented without having the same decision.

2.3.2. Prospective modelling by LEE-CARTER

The objective of the prospective modelling is to take into account future changes in the risk studied. In life insurance, the risk that has been the subject of more prospective modelling studies is mortality. This may be due to:

- Mortality is the main risk covered in the welfare, borrower and pension segments;
- National data are available and exploitable despite some limitations (basic risk or risk of decorrelation between national and study population mortality);
- The trend towards lower mortality is seen in several countries, particularly in France.

The models used to construct the mortality tables initially adjust to past trends and then extrapolate to the future. The construction stages of the prospective tables are as follows:

1. Construction of crude rates by age x and calendar year t, marked $q_{x,t}$. The history to construct the crude rates must be long (greater than 3-5 years) to approximate the past trends.
2. Change from the crude mortality quotient $q_{x,t}$ to the instantaneous mortality rate which is the modelled variable. A hypothesis of constancy of the instantaneous rate in each square of the Lexis diagram is made and leads to obtain the following estimator :

$$\mu_{x,t} = -\ln(1 - q_{x,t})$$

3. Closing rate curves. Since crude rates beyond a certain age are inoperative, closure methods consisting in extrapolating instantaneous mortality rates can be used (examples of methods: DENUIT & GODERNIAUX, DENUIT & QUASHIE, COALE & KISKER).
4. Use of parametric models such as LEE-CARTER and LOG-POISSON.

Specificities of the LEE-CARTER model

The LEE-CARTER model extrapolates the instantaneous mortality rate as follows:

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$$

With:

- α_x which is interpreted as the average of $\ln(\mu_{x,t})$;
- β_x which reflects the sensitivity of instantaneous mortality at age x in relation to overall evolution kt ;
- $\varepsilon_{x,t}$ which are random variables i.i.d. (independent and identically distributed) according to a $N(0, \sigma^2)$ law and representing a random phenomenon.

To make the model identifiable, the following constraints are added:

$$\sum_x \beta_x = 1 \text{ and } \sum_t k_t = 0$$

The parameters are obtained by minimizing the least squares criterion (non-linear).

$$\left(\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t\right) = \arg min \sum_{x,t} (\ln(\mu_{x,t}) - \alpha_x + \beta_x k_t)^2$$

LEE-CARTER model limitations

The LEE-CARTER model assumes stability over time of sensitivity parameter β_x : quotients of relative changes in mortality rates at different dates do not depend on date t.

The model involves homoscedasticity in mortality rates, which is not true in practice, since the variance of crude rates increases with age with declining numbers.

It has a selection of optimal parameters that has no probabilistic justification.

Furthermore, there is no assumption about age correlation in the LEE-CARTER model: each age has a separate parameter. However, close or distant ages may have the same changes over time.

Typology of data used in the LEE-CARTER model

The LEE-CARTER model can be used on data internal to a structure, which has the advantage of being able to measure the endogenous evolution of the mortality risk, but with a limit linked to the data especially at the high ages.

It can be used on external data, such as those of the INSEE, which has the advantage of reducing, or even cancelling, the limit related to the volume but with the disadvantage of increasing the basic risk (different evolution between external and internal data).

It is customary to position internal data rates in relation to external data rates.

2.4. Presentation of our prospective modelling approach

2.4.1. Study of the compensation rate's risk factors: Using Machine Learning Algorithms

We wanted to use statistical learning algorithms, in particular XGBOOST, on our study data, given the available volume, to help us classify explanatory variables and measure their influences on the level of TIW claims.

The data history is 2010-2018. The explanatory variables are the gender, the age group at the time of the claim, the nature of insurance, the nature of the benefit, the reason for waiving the user fee, the type of envelope and the medical specialty of the TIW's prescriber. The region is added when we limit ourselves to the historical 2015-2018. The variable to be explained is the annual compensation rate after a TIW.

The following tables give the classification of the explanatory variables by importance.

VARIABLE	WORDING	IMPORTANCE
ETP_CAT_SNDS	Category Etb Prescriber	22,39%
LIB_PSP_SPE_SNDS	Specialty Medical PS Prescriber	21,82%
AGE_BEN_SNDS	Age group Beneficiary at time of care	19,52%
EXO_MTF	Reason for Waiver of the User Ticket	10,08%
BEN_SEX_COD	Sex of the Beneficiary	8,24%
CPT_ENV_TYP	Type of Envelope	7,20%
PRS_NAT	Nature of Service	5,68%
LIB_ASU_NAT	Nature of Insurance	3,13%
PSP_ACT_CAT	Category of the Prescriber	1,71%
PSP_ACT_SNDS	Nature of Activity PS Prescriber	0,22%

Classification of explanatory variables (without region). History: 2010-2018. Claimants between 20 and 59 years old.

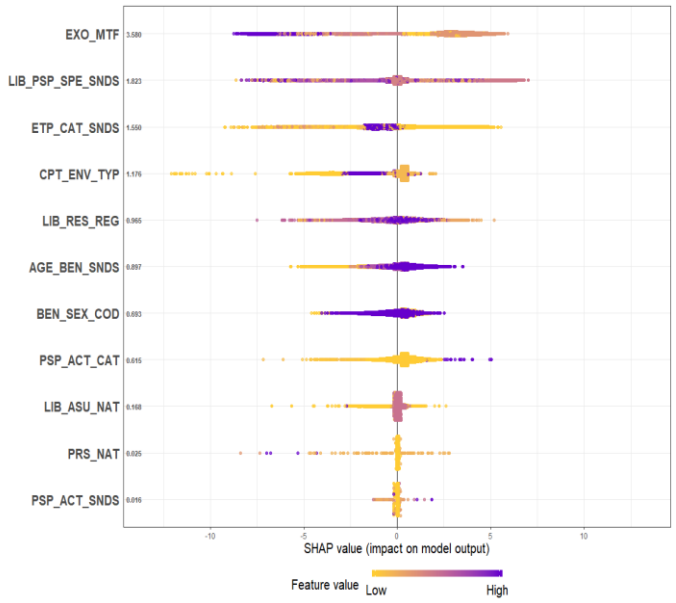
VARIABLE	WORDING	IMPORTANCE
LIB_RES_REG	Region of Residence of the Beneficiary	21,03%
CPT_ENV_TYP	Type of Envelope	17,55%
LIB_PSP_SPE_SNDS	Specialty Medical PS Prescriber	15,91%
ETP_CAT_SNDS	Category Etb Prescriber	15,21%
AGE_BEN_SNDS	Age group Beneficiary at time of care	11,85%
EXO_MTF	Reason for Waiver of the User Ticket	8,72%
BEN_SEX_COD	Sex of the Beneficiary	5,85%
PSP_ACT_CAT	Category of the Prescriber	1,70%
LIB_ASU_NAT	Nature of Insurance	1,46%
PSP_ACT_SNDS	Nature of Activity PS Prescriber	0,43%
PRS_NAT	Nature of Service	0,29%

Classification of explanatory variables. History: 2015-2018. Claimants between 20 and 59 years old.

We used the SHAP Value approach to help us interpret the outputs of the XGBOOST algorithm. The graph below classifies the explanatory variables according to the magnitude of the effect on the compensation rate for the 2015-2018 history.

Explanatory variables are ranked in order of importance. The horizontal bar shows whether the effect of the value of the explanatory variable is associated with a high or low compensation rate's prediction.

The violet color is associated with a high rate value, while the yellow color is the reverse.



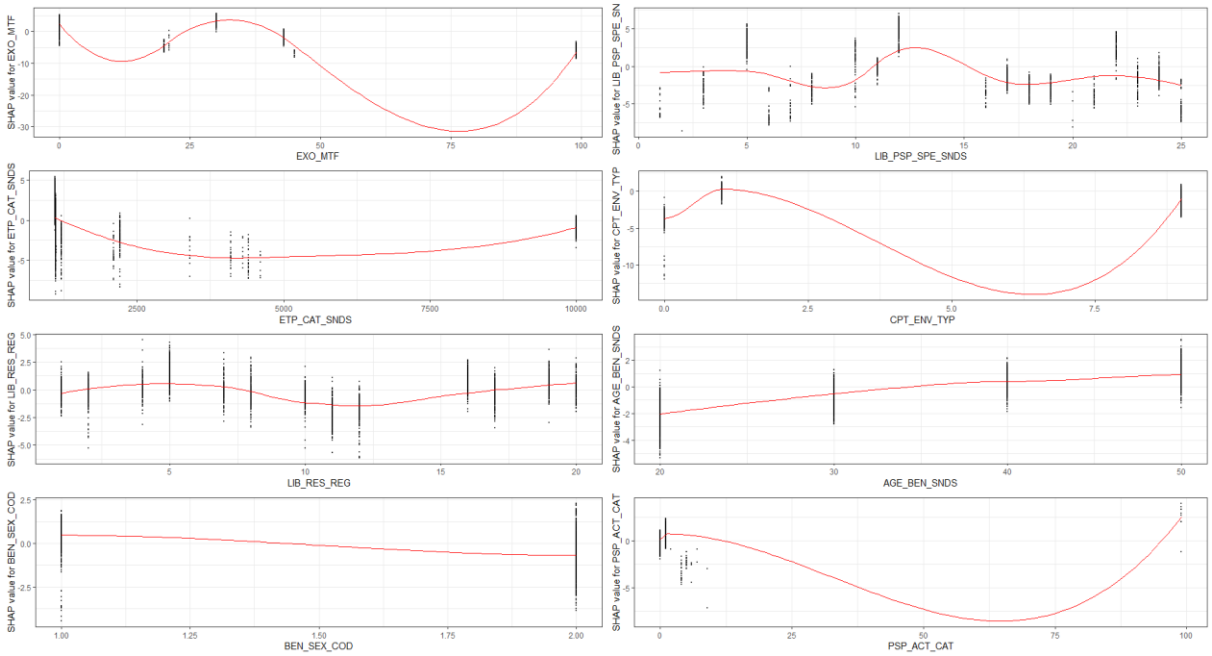
A low value of the variable “Reason for Waiver of User Fees” is associated with a high value of the compensation rate; the correlation is negative. In fact, the compensation rates are lower for fertility treatments, pathological leave (after the 6th month of pregnancy), preventive measures and specialized education (modalities from 51).

The compensation rate is high regardless of the value of the prescriber’s specialty, but we reach a maximum for value 12 (General Medicine).

A high age of the claimant is associated with a high level of compensation rate.

The compensation rate is high for all genders, but it is higher for men.

The following graph shows the compensation rate’s evolution according to the value taken by the 8 main explanatory variables.

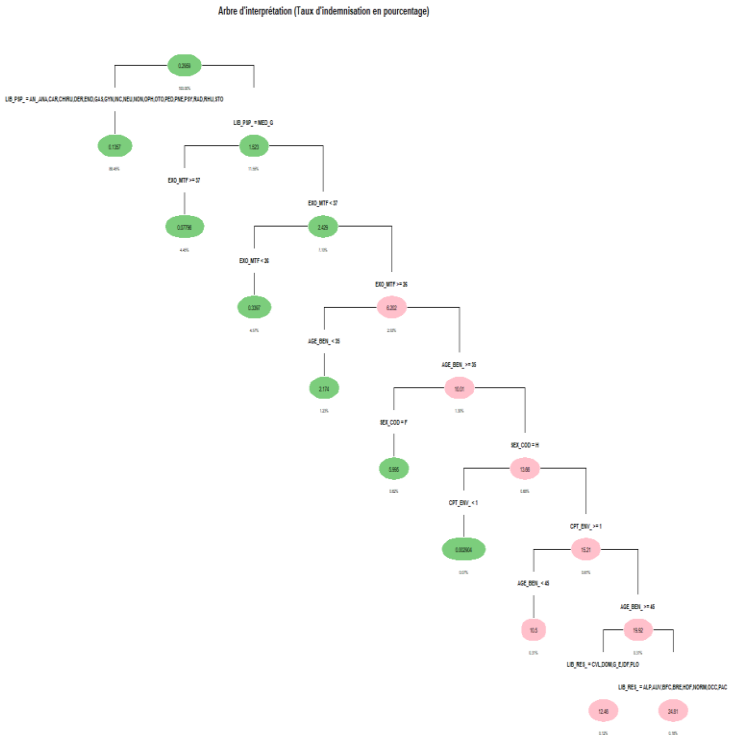


The analysis of the rate’s evolution according to the variable "ETP_CAT_SNDS", providing information on the prescriber’s center, shows that the levels of compensation are higher in regional hospital centers, hospital centers, psychiatric and other hospital centers and local hospitals.

The variable "CPT_ENV_TYP" makes it possible to isolate the refunds attached to the legal benefit envelopes. The highest levels of compensation are those for city care, hospitalization and medico-

social and intermediate consumption (medicines) and statutory health insurance benefits outside the ONDAM envelope.

To complete our interpretation process, we also used the Tree Interpreter, which allows us to obtain the following graph:



The study of the compensation rate's explanatory variables following a TIW gives us a lot of information, including the importance of gender and age, but also on the possible origins of the claim thanks to the speciality and belonging center of the prescriber. Variables such as the reason for exonerating the user fee and the service envelope also provide some explanation of the compensation level. The nature of the benefit (type of DA) does not provide any further explanation of the compensation level.

In the remainder of our study, we propose to focus on the three lines of analysis that are the specialty of the TIW's prescriber, the gender and the age group of the claimant.

2.4.2. Prediction of the specialty of incapacity's prescriber

Given that one of the important variables used in our modelling, the speciality of the TIW's prescriber, is not provided for a significant part of our database, we used the XGBOOST to predict its value.

As part of our prediction study, our approach is very classic: after having previously divided our data into three files (learning, testing and application), we build predictive models that we evaluate by the criterion of error rate test to compare performance.

The learning and test files do not contain the lines where the specialty of the TIW's prescriber is not entered or known.

The application file contains only the claim lines where the prescriber's specialty is not entered.

- We establish a first model (noted XG_BOOST_MOD_1) by the XGBOOST algorithm with as a learning base 70% of the data from the 2015-2018 history and integrating the region as an explanatory variable;
- We test the XG_BOOST_MOD_1 model on the test basis made up of the rest of the 2015-2018 history data (30%) and integrating the region. We get the error rates (number of bad predictions on the total number of predictions) given in the table below.

Studied segment (learning and testing)	Claimant aged 20 to 59 – Historical 2015-2018	Claimant aged 20 to 39 – Historical 2015-2018	Claimant aged 40 to 59 – Historical 2015-2018
Error rate (testing)	3,73%	3,26%	4,13%

This model seems relevant for us to predict the specialty of the TIW's, and thus replace the missing values. On the other hand the region being the second most important variable of the XG_BOOST_MOD_1 model, and given that the quality of this variable is not good for the 2010-2014 history, we cannot use it;

- We establish a second model (noted XG_BOOST_MOD_2) with a learning base composed of 70% of the data from the 2015-2018 history, but not integrating the region as an explanatory variable ;
- We test the XG_BOOST_MOD_2 model on the 30% test basis of 2015-2018 history data. We get the following error rates :

Studied segment (learning and testing)	Claimant aged 20 to 59 – Historical 2015-2018	Claimant aged 20 to 39 – Historical 2015-2018	Claimant aged 40 to 59 – Historical 2015-2018
Error rate (testing)	3,37%	3,43%	3,93%

The error rate is slightly lower, despite the removal of the second most important variable in the machine learning algorithm. In fact, the annual compensation rate, very correlated to the region, takes over in the prediction of the specialty.

- We are using the XG_BOOST_MOD_2 model to replace the missing values of the 2010-2018 History TIW's prescriber specialty.

The following graphs show the classifications of the variables explaining the specialty of the TIW's prescriber.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	58,67%
EXO_MTF	Reason for Waiver of the User Ticket	14,13%
LIB_RES_REG	Region of Residence of the Beneficiary	13,04%
BEN_SEX_COD	Sex of the Beneficiary	5,27%
AGE_BEN_SNDS	Age group Beneficiary at time of care	4,98%
CPT_ENV_TYP	Type of Envelope	2,25%
LIB_ASU_NAT	Nature of Insurance	1,65%
PRS_NAT	Nature of Service	0,01%

Integration of the region in the explanatory variables. History: 2015-2018. Claimants between 20 and 39 years.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	69,42%
EXO_MTF	Reason for Waiver of the User Ticket	9,98%
AGE_BEN_SNDS	Age group Beneficiary at time of care	6,14%
BEN_SEX_COD	Sex of the Beneficiary	4,93%
CPT_ENV_TYP	Type of Envelope	4,83%
LIB_ASU_NAT	Nature of Insurance	4,33%
PRS_NAT	Nature of Service	0,37%

Removal of the region from explanatory variables. History: 2015-2018. Claimants between 20 and 39 years.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	59,93%
EXO_MTF	Reason for Waiver of the User Ticket	14,95%
LIB_RES_REG	Region of Residence of the Beneficiary	12,39%
BEN_SEX_COD	Sex of the Beneficiary	5,03%
CPT_ENV_TYP	Type of Envelope	3,32%
AGE_BEN_SNDS	Age group Beneficiary at time of care	3,26%
LIB_ASU_NAT	Nature of Insurance	0,99%
PRS_NAT	Nature of Service	0,12%

Integration of the region in the explanatory variables. History: 2015-2018. Claimants between 40 and 59 years.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	69,32%
EXO_MTF	Reason for Waiver of the User Ticket	14,42%
BEN_SEX_COD	Sex of the Beneficiary	4,90%
CPT_ENV_TYP	Type of Envelope	4,10%
AGE_BEN_SNDS	Age group Beneficiary at time of care	3,97%
LIB_ASU_NAT	Nature of Insurance	2,69%
PRS_NAT	Nature of Service	0,60%

Removal of the region from explanatory variables. History: 2015-2018. Claimants between 40 and 59 years.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	57,62%
EXO_MTF	Reason for Waiver of the User Ticket	12,82%
LIB_RES_REG	Region of Residence of the Beneficiary	11,46%
AGE_BEN_SNDS	Age group Beneficiary at time of care	8,85%
BEN_SEX_COD	Sex of the Beneficiary	4,69%
CPT_ENV_TYP	Type of Envelope	3,31%
LIB_ASU_NAT	Nature of Insurance	1,11%
PRS_NAT	Nature of Service	0,15%

Integration of the region in the explanatory variables. History: 2015-2018. Claimants between 20 and 59 years.

VARIABLE	WORDING	IMPORTANCE
TX_IJ	Annual compensation rate	67,55%
EXO_MTF	Reason for Waiver of the User Ticket	12,51%
AGE_BEN_SNDS	Age group Beneficiary at time of care	8,67%
CPT_ENV_TYP	Type of Envelope	4,08%
BEN_SEX_COD	Sex of the Beneficiary	4,07%
LIB_ASU_NAT	Nature of Insurance	2,74%
PRS_NAT	Nature of Service	0,38%

Removal of the region from explanatory variables. History: 2015-2018. Claimants between 20 and 59 years.

Using the predictions obtained from the XGBOOST model, we have replaced the unknown values of the variable giving the specialty of the TIW's prescriber. The following table provides the distribution of predictions for unknown values:

AN_REAN	ANAT_PATH	CARDIO	CHIR_DENT	CHIRU	DERMA_VEN	ENDO	GASTRO_ENT	GERIA	GYNEC	MED_GEN	MED_INT
0,13%	0,00%	0,18%	0,00%	50,83%	0,06%	0,11%	0,66%	0,01%	9,97%	4,53%	0,30%

MED_PHY	NEPHR	NEURO	OPHT	OTO_LAR	PEDIA	PNEUM	PSY	RADIO	RHUMA	STOM
1,45%	0,02%	0,59%	0,30%	1,18%	0,03%	0,12%	23,68%	0,56%	5,14%	0,16%

Nearly half of the predictions relate to SURGERY and a quarter to PSYCHIATRY. Nearly 5% are attributed to RHEUMATOLOGY and GENERAL MEDICINE.

2.4.3. Projection of compensation rates by the prospective LEE-CARTER approach

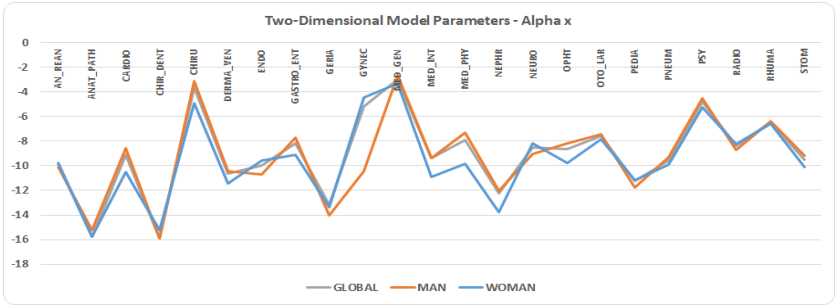
In this part, we use the LEE-CARTER model to broadcast the compensation rate for TIW. The following table makes a comparison between the classic approach and ours.

	Classical approach	Proposed approach
Modelled variable	Instantaneous mortality rate at age x and period t: $\mu_{x,t}$	Annual compensation rate for temporary incapacity to work for the speciality of the prescriber x and period t: $\gamma_{x,t}$
Parametric model	$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$	$\ln(\gamma_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$
Interpretation of the parameters	α_x is the average of $\ln(\mu_{x,t})$	α_x is the average of $\ln(\gamma_{x,t})$
	β_x reflects the sensitivity of instantaneous mortality at age x relative to the overall evolution k_t	β_x reflects the sensitivity of the compensation rate to the prescriber's specialty x relative to the general evolution k_t
	$\varepsilon_{x,t}$ are random variables i.i.d. according to a $N(0, \sigma^2)$ law and representing a random phenomenon	
Constraints on model identification	$\sum_x \beta_x = 1$ and $\sum_t k_t = 0$	
Obtaining the parameters	$(\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t) = \arg min \sum_{x,t} (\ln(\mu_{x,t}) - \alpha_x + \beta_x k_t)^2$	

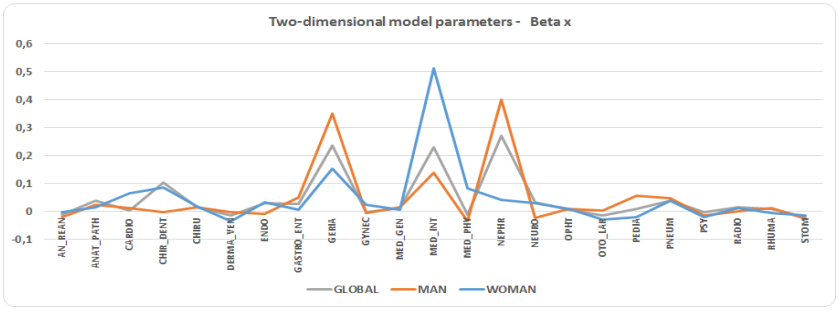
The compensation rate (relating to a given age group and gender) for the speciality of the TIW's prescriber *x* and year *t* is therefore split on the logarithmic scale, to an error term, the sum of a component specific to the speciality x and a product between a time parameter describing the general evolution of the claims experience and a parameter specific to the specialty describing the evolution of the rate in x in relation to those relating to other specialties.

$$\ln(\gamma_{x,t})^{Age,Sexe} = \alpha_x^{Age,Sexe} + \beta_x^{Age,Sexe} k_t^{Age,Sexe} + \varepsilon_{x,t}^{Age,Sexe}$$

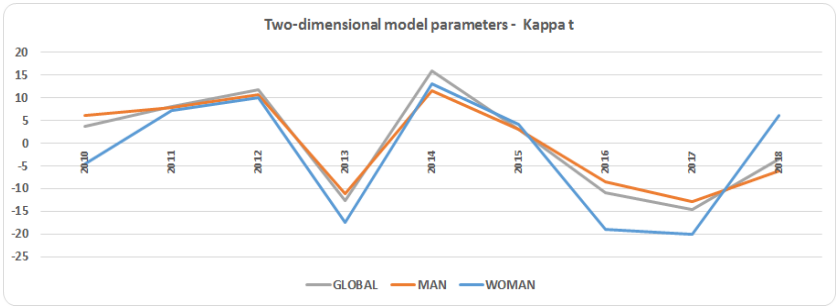
The application of the two-dimensional model on our data allows to obtain the parameters indicated in the following graphs.



Analysis of the average behaviour of compensation rates over time: Rates relating to specialties (logarithmic scale) SURGERY, GYNECOLOGY (for women), GENERAL MEDICINE, PSYCHIATRY and RHEUMATOLOGY are the highest.

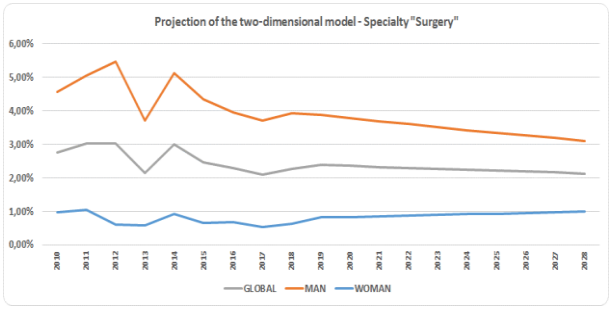


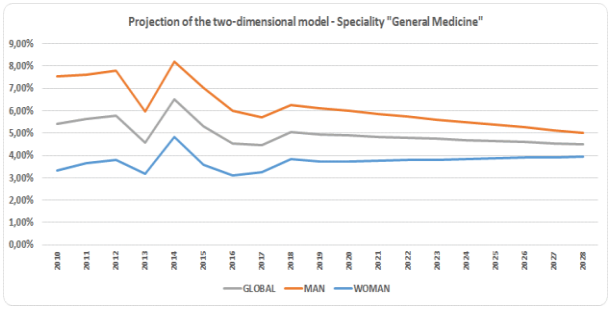
Analysis of the rate's evolution compared to other specialties: DENTAL SURGERY, GERIATRY, INTERNAL MEDICINE and NEPHROLOGY have a higher evolution ratio than other specialties. The limit of these results is that they are specialties with a very low rate, therefore a more erratic evolution than the others.

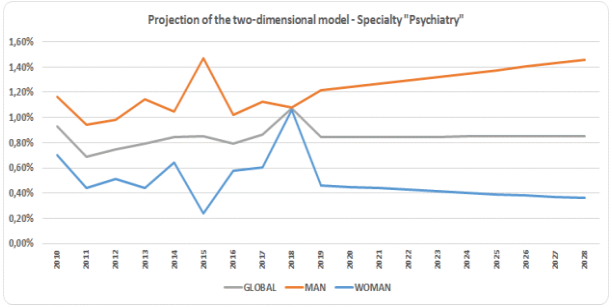


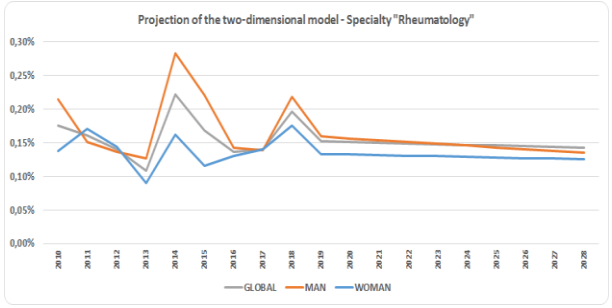
Analysis of the temporal evolution of TIW risk: The highest level of loss is achieved in 2014. The years 2013, 2016 and 2017 are those with the lowest levels.

The four graphs below show the evolution of annual compensation rates for the main specialties (General Medicine, Surgery, Psychiatry and Rheumatology).





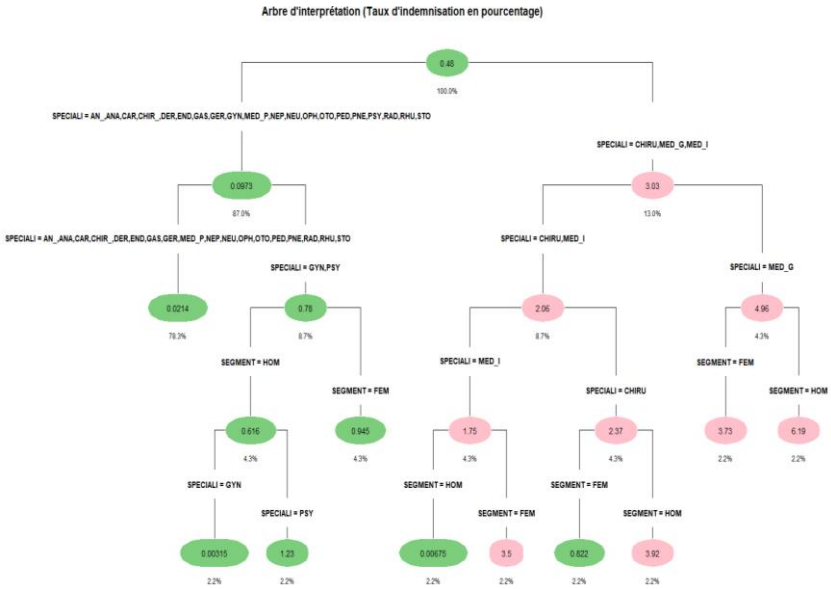




In view of the low compensation rates obtained for other specialties, resulting in a large volatilité and thus a higher β_x, we propose to use a decision algorithm to group specialties and gender. The objective

of this grouping is to decrease the estimation error and the number of estimated parameters in our two-dimensional model.

The decision algorithm (Tree Interpreter) will have as variable explained the annual compensation rate between 2010 and 2028, and as explanatory variables the specialty of the TIW's prescriber and the gender of the claimant.



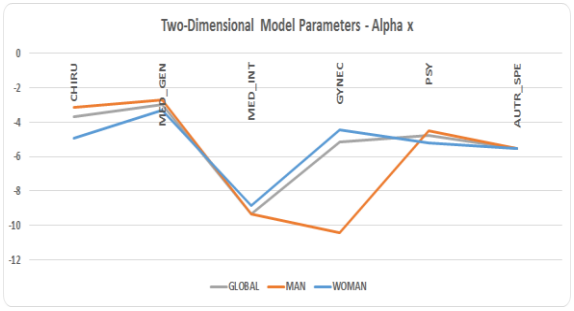
The analysis of the decision tree's results shows that:

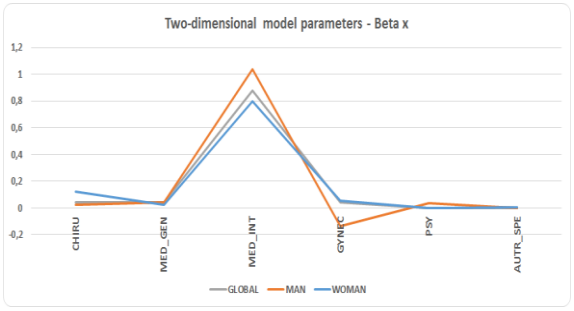
- The modelling of the annual compensation rate (current + future) needs to be distinguished by gender for the specialties GENERAL MEDICINE, INTERNAL MEDICINE, SURGERY, GYNECOLOGY and PSYCHIATRY.
- Rate modelling for all other specialties can be done regardless of gender and specialty.

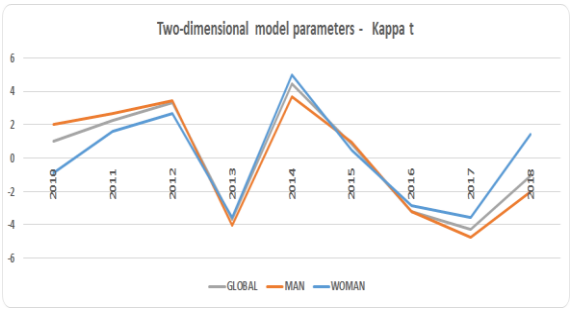
The new simplified two-dimensional model reads:

$$\ln(\gamma_{x,t})^{Age,Sexe} = 1_{\{x=MED_{GEN},MED_{INT},CHI,GYN,PSY\}} \times \left(\alpha_x^{Age,Sexe} + \beta_x^{Age,Sexe} k_t^{Age,Sexe}\right) + 1_{\{x\neq MED_{GEN},MED_{INT},CHI,GYN,PSY\}} \times \left(\alpha_x^{Age} + \beta_x^{Age} k_t^{Age}\right) + \varepsilon_{x,t}^{Age}$$

The application of the new two-dimensional model allows to obtain the parameters indicated in the following graphs.







2.4.4. Validation of the prospective approach

In this section, we propose to compare the compensation rates obtained by our two-dimensional model, limiting our estimate history to 2010-2016, to those observed in 2017 and 2018.

The following table gives the sum of the squares of differences between predictions and observed, weighted by exposures.

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
MAN	CHIRU	56	2	58
	MED_GEN	1	443	444
	MED_INT	0	0	0
	GYNEC	0	0	0
	PSY	38	33	71
	OTH_SPECIALTY	5	3	8
	ALL SPECIALTIES	100	481	581

Male population [20-39] years - Sum of squares differences between observed and two-dimensional model predictions

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
WOMAN	CHIRU	53	25	78
	MED_GEN	164	18	182
	MED_INT	1	1	1
	GYNEC	73	269	342
	PSY	8	251	259
	OTH_SPECIALTY	2	8	10
	ALL SPECIALTIES	302	571	872

Female population [20-39] years - Sum of squares differences between observed and two-dimensional model predictions

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
MAN	CHIRU	925	1	926
	MED_GEN	890	3 920	4 811
	MED_INT	8	0	8
	GYNEC	0	0	0
	PSY	256	727	983
	OTH_SPECIALTY	11	100	111
	ALL SPECIALTIES	2 090	4 750	6 839

Male population [40-59] years - Sum of squares differences between observed and two-dimensional model predictions

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
WOMAN	CHIRU	150	1	150
	MED_GEN	119	562	680
	MED_INT	3	2	5
	GYNEC	3	367	370
	PSY	123	63	186
	OTH_SPECIALTY	49	34	83
	ALL SPECIALTIES	446	1 029	1 475

Female population [40-59] years - Sum of squares differences between observed and two-dimensional model predictions

By way of comparison, we calculated the average compensation rates obtained in 2012 and 2016 and considered this as an estimate for 2017 and 2018. The table gives the sum of the squares of differences:

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
MAN	CHIRU	512	266	778
	MED_GEN	1 326	421	1 748
	MED_INT	0	0	0
	GYNEC	0	0	0
	PSY	0	2	2
	OTH_SPECIALTY	1	10	11
	ALL SPECIALTIES	1 839	699	2 538

Male population [20-39] years - Sum of squares differences between mean predictions and observed

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
WOMAN	CHIRU	18	4	22
	MED_GEN	178	14	191
	MED_INT	0	0	1
	GYNEC	219	511	730
	PSY	12	270	282
	OTH_SPECIALTY	1	10	11
	ALL SPECIALTIES	428	808	1 236

Female population [20-39] years - Sum of squares differences between mean predictions and observed

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
MAN	CHIRU	455	204	659
	MED_GEN	3 948	133	4 081
	MED_INT	19	3	22
	GYNEC	0	0	0
	PSY	291	633	924
	OTH_SPECIALTY	11	90	101
	ALL SPECIALTIES	4 724	1 063	5 787

Male population [40-59] years - Sum of squares differences between mean predictions and observed

		PERIOD		
	SPECIALTY	2017	2018	2017-2018
WOMAN	CHIRU	153	6	159
	MED_GEN	116	2 340	2 456
	MED_INT	1	5	6
	GYNEC	9	415	424
	PSY	45	12	57
	OTH_SPECIALTY	11	94	105
	ALL SPECIALTIES	334	2 872	3 206

Female population [40-59] years - Sum of squares differences between mean predictions and observed

The differences observed among 20-39 year olds by taking a simple average of the previous observations are higher than those obtained with the predictions of the two-dimensional model, especially for men.

For male claimants aged [40-59] years, the differences between the predictions of the two-dimensional model and the number of DA observed are higher than those obtained by taking an average of the observation between 2012 and 2016.

We thus obtain a projection model closer to short-term observation.

3. TREATMENT OF INCAPACITY RISK AND MONITORING

3.1. What is the prevention approach?

Prevention approaches and tools in place in companies and jurisdictions can take time to be effective. This may be due to:

- Lack of consideration for differences in risk exposure by gender, occupation, age, place of work and residence;
- A problem of integrating new hires.

It is necessary to renew the elements of evaluation and prevention for disability risk in order to improve the working conditions of all employees.

The adaptation of work systems and policies to prevent disability claims can be done by:

- Systematizing the production of disability claims indicators by gender, age, causes and occupation (if available);
- Systematizing the production of indicators related to the quality of working life;
- Adapting work systems to take into account differences in morphology, size, muscular strength, articulation of working times and obscuring differences in age;
- Developing the HSWCC or SEC approaches;
- Implementing GBV prevention mechanisms in occupational health action plans;
- Adapting agreements on psychosocial risks if necessary.

3.2. What strategy should be adopted in the own risk assessment?

One of the main problems of insurers is the reconciliation of the following five quantities:

The quantity modelled	The quantity declared	The quantity compensated	The quantity realized	The quantity future
• It is biased mainly due to modelling errors	• It corresponds to what is declared by the insured. Filters related to contract exclusions, misrepresentations, take-over conditions result in a correlation with the quantity compensated	• It corresponds to what is paid by insurers and therefore generally lower than what is reported by policyholders	• This is the true sinistrality which is not always equal to the quantity declared because of a lack of reporting time, a lack of opportunity, a lack of knowledge of the guarantees offered by the contracts, etc.	• It is the future realization of the sinistrality, dependent on all the possible evolutions of the world.

For a risk such as incapacity, a no (or bad) consideration of its evolution may lead to a more or less significant differentiation between the modelled/compensated quantities and the future quantity.

By choosing to integrate a prospective dimension for incapacity risk, we introduce a stochastic simulation logic around the trend (often adopted in economic capital calculations) which allows for the probabilistic description of the modelled variable.

We recommend establishing increased monitoring of disability, in particular incapacity, and its evolution over time. The risk monitoring approach we recommend is as follows:

- Compare the risk evolution of the insured population with that of the national population covered by the SS. This comparison can be done by homogeneous risk group (gender x age x region, etc.);
- Adjust the assumptions (incidence and maintenance laws) at a frequency of at least two years if the level of modelled claims is different from the real level;

- Identify relevant ORSA scenarios and quantify the associated economic capital based on the evolution of risk.

The two-dimensional projection model proposed in this article can be used on the insurance organization's own data in order to obtain its own probabilistic distribution and quantify the ORSA capital related to the disability risk.

A good assessment of the overall claims level, more particularly, that of incapacity risk and its explanatory factors allows:

- To adapt the characteristics and clauses when creating new products or reviewing existing products, with the objective of setting up customized covers, for example;
- To measure and limit issues such as anti-selection and moral hazard that are important, especially for a protection contract where the health of insured persons is the central issue;
- Put in place better risk management strategies.

3.3. Appropriation of new evaluation techniques

The data used to evaluate insurance risks are becoming more numerous and heterogeneous, with an internal and/or external origin.

The appropriation of new risk assessment techniques such as AI requires:

- Streamline the collection, processing and qualification of data, while being "GRDP-compliant". Documentation, incorporating a flow audit trail, is essential.
- Compare and evolve current actuarial models to AI models. Studies combining data of different origins and structures improve risk assessment/management by facilitating more detailed works (pricing of customized contracts, measurement of cyber or emerging risks, etc.). But we must be vigilant about the criteria of robustness, interpretability and comprehensibility of models.
- Automate simple or low value-added tasks by involving the people concerned.

Insurance organizations face strategic and organizational choices:

- Select the level of segmentation in the pricing, provisioning, evaluating risk and performance indicators (BE, SCR, etc.).
- Acquire skills to develop and adapt evaluation models, taking into account changes brought by regulatory reforms. IT (Information Technology) is a key issue as storage capacities, interfaces and processing software must follow the evolution of data.
- Adapt working methods by adopting accelerated implementation and development processes.

3.4. Fight against cognitive bias

Cognitive biases are forms of thinking that deviate from logical or rational thinking and tend to be systematically used in a variety of situations, including in insurance when designing coverage products or when modelling risks. Some of the most common cognitive biases in assurance are:

- Model illusion: It is the tendency to see models in random events, such as the idea that an employee with children will be more likely to fall into TIW than other employees.

- Confirmation Bias: This is the tendency to listen only to information that supports our preconceived ideas.
- Conservatism bias: People prefer past evidence to new evidence or information that has emerged.
- Recency bias: This is the tendency to give more weight to the latest information than to the older data.
- Information bias: This is the tendency to look for information when it does not affect the action. More information is not always better, because with less information, people can often make more accurate predictions. In our study, making predictions with many variables or data did not necessarily yield better results compared to those made on a limited number of data and/or variables.
- Pro-innovation bias: it is the overestimation of the usefulness of an innovation and/or underestimating its limits. Machine Learning algorithms can bring innovation, but they have limitations, which can be more important than those of more traditional methods.
- Salience bias: This is the tendency to focus on the most easily recognizable characteristics of a person or concept. Interpretation trees are often complex due to the number of nodes. It is important not to focus on certain traits or knots at the expense of others.
- Stereotype bias: it is the expectation that a group or person will have certain characteristics without having any real information about the latter.

Insurers (management and operational) should be encouraged to reduce these biases by taking a step back from current observations that may be limited by the data's unavailability or quality and the predictions of the models used.

4. LIMITS

Our approach of prospective modelling based on national data and combining two-dimensional models and machine learning algorithms has certain limitations that we wish to bring to the attention of the reader.

The data used do not allow the disability risk to be split between incidence and maintenance. On the other hand, the results obtained, such as replacing unknown data with predictions from machine learning algorithms, may be different from those we could obtain with the real data. In addition, some variables of interest are not present in the Open Damir database, such as occupation, income or marital status, which does not allow us to refine our modelling.

The risk exposure considered for a given year is the French population at the beginning. There were no adjustments made according to the months in order to take into account the inflows and outflows (migration flows, deaths, births, incapacity/invalidity claims in progress, etc.). There have been no adjustments made to return to the population covered by French Health Insurance.

The Open Damir data allow us to understand the French evolution for the risk under consideration. The results of our modelling can be used by an insurance organization as part of a claims' prospective modelling associated with its portfolio. Unfortunately, there will always be the presence of basis risk, as may be the case with studies on prospective modelling of mortality risk.

5. CONCLUSION

This article propose a model for estimating the compensation rate following a temporary incapacity to work. Given the limitations of frequently used actuarial models, we present in this work a method combining a two-dimensional structure similar to the LEE CARTER model whose parameterization is performed using Machine Learning algorithms. This approach achieves a prospective compensation rate for each age group, gender and medical specialty.

Our approach allows:

- To facilitate the monitoring of the incapacity risk over time,
- Easier analysis and explanation of the causes of risk evolution,
- Anticipate future developments especially with the emergence of new coverage solutions and preventive measures,
- Improve the risk management process,
- Facilitate the strategic choices of senior management with faster access to relevant information.

That said, despite all the efforts to anticipate and identify risks, our approach has limits and some changes will be unpredictable, since they can only be observed retroactively: the best management method is risk governance. The latter must be reactive, efficient and equipped with human resources "connected" with developments in the economic, social and regulatory environment.

6. ANNEXES

6.1. Glossary

ABBREVIATION	TERM USES	DEFINITION
TIW	TOTAL AND TEMPORARY INCAPACITY to WORK	
IJ	INDEMNITES JOURNALIERES	
MSD	MUSCULOSKELETAL DISORDERS	Musculoskeletal disorders include disorders affecting structures on the periphery of joints such as muscles, tendons, nerves, ligaments, bursa, joint capsules, vessels…
SS	SOCIAL SECURITY	
ORSA	OWN RISK SOLVENCY ASSESSMENT	
AI	ARTIFICIAL INTELLIGENCE	
GRDP	EUROPEAN GENERAL REGULATION ON DATA PROTECTION	
BE	BEST ESTIMATE	
SCR	SOLVENCY CAPITAL REQUIREMENT	
SNIIRAM	NATIONAL INTER-SYSTEM OF HEALTH INSURANCE	
DAMIR	HEALTH INSURANCE EXPENSES CROSS PLAN	
INSEE	NATIONAL INSTITUTE FOR STATISTICS AND ECONOMIC STUDIES	
HSWCC	HEALTH, SAFETY AND WORKING CONDITIONS COMMITTEE	
SEC	SOCIAL AND ECONOMIC COMMITTEE	

6.2. References / Bibliography

- Extract from the Official Journal (October 2017). Labour & Safety n° 787.
- DREES (2016). Health spending in 2015. https://drees.solidarites-sante.gouv.fr/IMG/pdf/fiche34-2.pdf.
- Ayming firm in partnership with AG2R LA MONDIALE (2018). 10th Barometer of Absenteeism and Engagement. *http://presse.ayming.com/communique/176960/10eme-Barometre-de-l-Absenteisme-de-l-Engagement-Ayming*.

- AMELI (2019). Understanding musculoskeletal disorders. *https://www.ameli.fr/assure/sante/themes/tms/comprendre-troubles-musculosquelettiques*.
- Igotz AUBIN, Anne ROLLAND, Thierry POINCELIN (2010). Guidelines for the construction of disability probability laws. *Institut des Actuaires.*
- Frédéric PLANCHET & Vincent LELIEUR (2010). Use of Lee-Carter and Log-Poisson methods for adjusting mortality tables for small samples.
- Franck CHEVALIER (2016). Methodological concepts & frameworks. *Formation ERM Institut des Actuaires*
- Vivien GUETTE (2010). Determination of a mortality rate for a disaster with a 200-year return period. *Bulletin Français d'actuariat.*
- D. W. APLEY (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *Department of Industrial Engineering & Management Sciences, Northwestern University*.
- T. CHEN & C. GUESTRIN (2016). XGBoost: A Scalable Tree Boosting System. *Universty of Washington*.
- S. LUNDBERG & L. SU-IN (2017). A unified approach to interpreting model predictions. *CoRR, abs/1705.07874.*

6.3. Focus on the specificities of Social Security benefits

6.3.1. Daily Allowance following personal illness or injury

	Incapacity to work of less than 6 months	Incapacity to work of 6 months or more
Conditions of entitlement	To be compensated for the first 6 months of the interruption, the person affected must justify the following conditions on the day of the interruption: - have worked at least 150 hours during the 3 calendar months or the 90 days preceding the stop, - or have contributed, during the 6 calendar months preceding the judgment, on the basis of remuneration at least equal to 1,015 times the amount of the hourly minimum wage fixed at the beginning of that period.	To continue to be compensated if the stop extends beyond 6 months, the person affected must meet the following conditions: - on the date of interruption of work, proof of membership of the Social Security system for at least 12 months, - have worked at least 600 hours during the 12 calendar months or the 365 days preceding the stop, - have contributed, during the 12 calendar months or the 365 days preceding the stop, on the basis of remuneration at least equal to 2030 times the hourly minimum wage fixed at the beginning of that period.
Sickness Daily Allowance	50 % of the basic daily salary within the limit of the Monthly Social Security Ceiling (MSSC), calculated on the salary's average of the 3 months preceding the absence from work. If the beneficiary is paid monthly, the basic daily salary is equal to the total of the last three gross salary (used as a basis for calculating contributions due from sickness, maternity, invalidity and death insurance) received before the absence from work, divided by 91.25.	
Supplement for dependent children (>3)	66,66 % of the basic daily salary, from the 31st day of absence from work.	
Basic compensation range	Max (as of January 1, 2019): - €45.01 per day if less than 4 dependent children - €60.02 per day if more than 3 dependent children and from the 31st day of stay until the 360th day of compensation (or up to 3 years of compensation for a long-term condition)	
Waiting period	DA are paid after a waiting period of 3 days. The waiting period applies to each TIW, except in the following cases: - resumption of activity between two TIW not exceeding 48 hours, - successive work stoppages due to Long-Term Illness (LTI).	
Payment	DA are due for each calendar day of work interruption (including Saturday and Sunday). CPAM pays the DA every 14 days.	
Accumulation of income	- The DA can be combined with: the invalidity pension (if the invalidity rate is compatible with a resumption of employment), the old age pension (if the person affected is retired and is also employed), paid leave allowances and salary (if maintained, in whole or in part, by the employer during the sinister). - DA cannot be combined with: unemployment benefits, daily maternity benefits and daily allowances for work accidents and occupational diseases (WA-OD).	
Supplementary allowances paid by the employer	Conditions: In order to receive additional benefits, the affected person must meet all of the following conditions: - have at least one year's seniority in the enterprise (calculated from the first day of absence), - have forwarded the medical certificate to the employer within 48 hours, - benefit from the DA paid by the Social Security Office, - be cared for in France or in one of the Member States of the European Economic Area (EEA), - not be a home worker or seasonal, intermittent or temporary employee. Waiting period : A 7-day waiting period is provided for each TIW. Thus, except for more favourable collective agreements or collective agreements, the payment of supplementary benefits begins on the 8th day of the sick leave. Duration of payment: The payment's duration of allowances paid by the employer varies according to seniority, as follows : - 1-5 years: 60 days (30 days at 90% and 30 days at 66.66%) - 6-10 years: 80 days (40 days at 90% and 40 days at 66.66%) - 11-15 years: 100 days (50 days at 90% and 50 days at 66.66%) - 16-20 years: 120 days (60 days at 90% and 60 days at 66.66%) - 21-25 years: 140 days (70 days at 90% and 70 days at 66.66%) - 26-30 years: 160 days (80 days at 90% and 80 days at 66.66%) - 31 years and over: 180 days (90 days at 90% and 90 days at 66.66%) If the person concerned has already received one or more periods of sickness benefit from the employer in the previous 12 months, the payment period shall be deducted from the number of days already paid. Amount: During the first 30 to 90 days of rest according to professional seniority, the claimant receives benefits, which, in addition to DA, make it possible to receive 90% of the gross remuneration which he would have received if he had worked. During the next 30 to 90 days, the percentage is reduced to 2/3 (or 66.66%) of his remuneration. To calculate the amount paid by the employer, it is necessary to deduct the DA paid by Social Security and, if this is the case, the benefits resulting from the employer's payments under a supplementary provident scheme. The collective agreement may provide for more favourable provisions (such as full salary maintenance)	
Revaluation	If the absence from work lasts longer than three months, the daily sickness allowance may be increased in the event of a general increase in salary.	

TABLE 1 : INCAPACITY TO WORK FOLLOWING PERSONAL ILLNESS OR INJURY

6.3.2. Daily allowances following an occupational illness or work accident

	Incapacity due to a work accident	Incapacity due to an occupational disease
Conditions of entitlement	To be compensated, the claimant have to be a salaried employee and have suffered the interruption following a work accident.	To be compensated, the claimant must be a salaried employee and have suffered the suspension as a result of a condition related to his professional activity. - As long as the occupational nature of the disease is not recognised, the person affected is entitled to basic DA, - As soon as the disease is recognised as being of occupational origin by the CPAM, it pays increased DA.
Daily Allowance	If the victim is a monthly employee, his CPAM determines a reference daily salary. This daily salary is calculated by dividing by 30.42 the gross salary received in the month preceding the start of the sick leave. However, if the result is higher than €337.97, this amount is taken into account. The DA are equal to a percentage of this reference daily salary which varies according to the duration of the illness within the limit of an amount called net daily earnings (the daily salary minus 21%). Their amount is capped. - From 1st to 28th day off: the percentage of the reference daily salary is 60%. The net daily earnings are 1/30.42 of the previous month's salary, minus a flat rate of 21%. The ceiling for DA is €202.78. - From 29th day off: the percentage of the reference daily salary is 80%. The net daily salary is 1/30.42 of the previous month's salary, minus a flat rate of 21%. The ceiling for DA is €270.38.	
Supplement for dependent children (>3)	Not applicable	
Basic compensation range	Max (as of January 1, 2019): - 202.78€ from the 1st to the 28th day of the stay, - 270.38€ from the 29th day off.	
Waiting period	In the event of a work accident, the day on which the accident occurs shall be paid in full by the employer. The DA are paid by the CPAM from the first day following the stoppage of work, without waiting period. The same applies in the case of relapse or worsening of the employee's health.	In the case of an occupational disease, there is no waiting period. In this case, the DA are paid by the CPAM from the first day following the TIW. The same applies in the case of relapse or worsening of the employee's health.
Payment	DA are paid throughout the period of incapacity to work, until the injury (or death) is fully healed or consolidated. If the injury is consolidated, a Permanent Incapacity Rate (PIR) is set by CPAM and entitles the claimant to a lifetime annuity or lump sum capital.	
Accumulation of income	- The DA may be combined with: the invalidity pension (if the invalidity rate is compatible with a resumption of employment), the old age pension (if the person affected is retired and is also employed), paid vacation and salary (if the employer maintains all or part during the TIW). - The DA cannot be combined with: unemployment benefits, daily maternity benefits and daily sickness or accident benefits from private life.	
Revaluation	When the absence from work is more than 3 months, the DA can be revalued in the event of a general increase in salary : - by ministerial decree: a revaluation coefficient (for example: 1.013) is applied to the basic daily earnings, - or if the affected person is covered by a collective agreement: he may request by mail to his CPAM the application of the revaluation provided for in the amendment to the agreement by attaching to his application a certificate from his employer. These two revaluations cannot be combined, only the most favourable one is applied.	

TABLE 2 : INCAPACITY FOLLOWING A WORK ACCIDENT (WA) OR OCCUPATIONAL DISEASE (OD)

6.4. Focus on the specificities of the employer's benefits

6.4.1. Daily allowance following personal illness or injury

	Incapacity due to a personal disease
Conditions of entitlement	In order to receive additional benefits, the affected person must meet all of the following conditions: - have at least one year's seniority in the enterprise (calculated from the first day of absence), - have forwarded the medical certificate to the employer within 48 hours, - benefit from the daily subsistence allowance (DA) paid by the Social Security Office, - be cared for in France or in one of the Member States of the European Economic Area (EEA), - not be a home worker or seasonal, intermittent or temporary employee.
Daily Allowance	During the first 30 to 90 days' stay according to seniority, the affected person receives benefits, which, in addition to the SS DA, make it possible to receive 90% of the gross remuneration he would have received if he had worked. During the next 30 to 90 days, the percentage is reduced to 2/3 (or 66.66%) of his remuneration. To calculate the amount paid by the employer, it is necessary to deduct the DA paid by Social Security and, if this is the case, the benefits resulting from the employer's payments under a supplementary welfare scheme. The collective agreement may provide for more favourable provisions (such as full salary maintenance). If DA paid by Social Security are reduced (for example, due to hospitalization or a penalty by CPAM for non-compliance with its regulations), the amount paid by the employer remains calculated taking into account the full amount of the DA.
Supplement for dependent children (>3)	Not applicable
Basic compensation range	Not applicable
Waiting period	A 7-day waiting period is provided for each TIW. Thus, except for more favourable collective agreements, the payment of supplementary benefits begins on the 8th day of the TIW.
Payment	Duration of payment: The duration of DA payment by the employer varies according to seniority, as follows: - 1-5 years: 60 days (30 days at 90% and 30 days at 66.66%) - 6-10 years: 80 days (40 days at 90% and 40 days at 66.66%) - 11-15 years: 100 days (50 days at 90% and 50 days at 66.66%) - 16-20 years: 120 days (60 days at 90% and 60 days at 66.66%) - 21-25 years: 140 days (70 days at 90% and 70 days at 66.66%) - 26-30 years: 160 days (80 days at 90% and 80 days at 66.66%) - 31 years and over: 180 days (90 days at 90% and 90 days at 66.66%) If the affected person has already received one or more periods of sickness benefit from the employer in the previous 12 months, the payment period shall be deducted from the number of days already paid.
Accumulation of income	Not applicable
Revaluation	Not applicable

TABLE 3 : EMPLOYER'S BENEFITS IN THE EVENT OF PERSONAL ILLNESS OR INJURY

6.4.2. Daily allowances following an occupational illness or work accident

	Incapacity due to an occupational disease or a work accident
Conditions of entitlement	In order to receive additional benefits, the affected person must meet all of the following conditions: - have at least one year's seniority in the enterprise (calculated from the first day of absence), - have forwarded the medical certificate to the employer within 48 hours, - benefit from the daily subsistence allowance (DA) paid by the Social Security Office, - be cared for in France or in one of the Member States of the European Economic Area (EEA), - not be a home worker or seasonal, intermittent or temporary employee.
Daily Allowance	The amount of compensation paid by the employer is calculated as follows: - during the first 30 days of TIW, the amount corresponds to 90 % of the gross remuneration which the person would have received if he had worked, - during the following 30 days, the amount corresponds to 2/3 (66.66%) of the same remuneration. Contractual provisions may provide for a higher amount. If his absence results in an increase in the working hours of other employees, this increase shall not be taken into account in calculating gross remuneration. When the DA paid by the social security are reduced (for example, due to hospitalization or a sanction of the CPAM for non-compliance with the procedure), the additional allowances paid by the employer remain calculated on the basis of the full amount of the DA.
Supplement for dependent children (>3)	Not applicable
Basic compensation range	Not applicable
Waiting period	There is no waiting period. The payment of additional benefits begins on the 1st day of absence.
Payment	Duration of payment: Supplementary allowances shall be paid until the end of the absence from work, up to a maximum period of 12 months. The duration of the payment of benefits by the employer varies according to the seniority of the affected person in the undertaking, as follows: - 1-5 years: 60 days (30 days at 90% and 30 days at 66.66%) - 6-10 years: 80 days (40 days at 90% and 40 days at 66.66%) - 11-15 years: 100 days (50 days at 90% and 50 days at 66.66%) - 16-20 years: 120 days (60 days at 90% and 60 days at 66.66%) - 21-25 years: 140 days (70 days at 90% and 70 days at 66.66%) - 26-30 years: 160 days (80 days at 90% and 80 days at 66.66%) - 31 years and over: 180 days (90 days at 90% and 90 days at 66.66%) The 12-month period of payment of benefits is valued, for each day of TIW, from date to date. All calendar days are taken into account. If the affected person has already received one or more periods of sickness benefit from the employer in the previous 12 months, the payment period shall be deducted from the number of days already paid.
Accumulation of income	Not applicable
Revaluation	Not applicable

TABLE 4 : EMPLOYER'S BENEFITS IN THE EVENT OF TIW FOLLOWING A WORK ACCIDENT OR AN OCCUPATIONAL DISEASE

6.5. Zoom on the variables studied in the Open Damir database

The following tables list the modalities of the variables studied from the Open Damir database.

Age group Beneficiary at time of care

AGE_BEN_SNDS	Description Age Group Beneficiary at the time of care
0	0-19 ANS
20	20 - 29 ANS
30	30 - 39 ANS
40	40 - 49 ANS
50	50 - 59 ANS
60	60 - 69 ANS
70	70 - 79 ANS
80	80 ANS ET +
99	UNKNOWN AGE

Sex of the Beneficiary

BEN_SEX_COD	Description Beneficiary's gender
0	UNKNOWN OR LEGAL PERSON
1	MALE
2	FEMININ
9	UNKNOWN VALUE

Nature of Insurance

ASU_NAT	LIB_ASU_NAT
10	MAL
22	CNMSS
30	MATER
40	AT_MP
50	DC
70	PREST_SUP
80	INVAL
90	PREV_MAL
99	INC

Nature of the work accident

ATT_NAT	LIB_ATT_NAT
0	SS_OBJ
1	ACC_TRAV
2	MAL_PRO
3	ACC_TRAJ
9	INCON

Reason for Waiver of the User Ticket

EXO_MTF	Description Reason for Waiver of the User Ticket
0	NOT APPLICABLE
10	INSURANCE
20	REGIMES
21	MILITARY INVALIDITY PENSION (ART. 115)
30	NATURE OF THE SERVICE
41	BENEFICIARY WITH A LISTED DISEASE - UNRELATED CONDITION
42	BENEFICIARY WITH A LISTED DISEASE - RELATED DISEASE
43	OFF-LIST PATHOLOGY - UNRELATED CONDITION
44	OFF-LIST PATHOLOGY - RELATED CONDITION
45	OFF-LIST POLYPATHOLOGY - UNRELATED CONDITION
46	OFF-LIST POLYPATHOLOGY - RELATED CONDITION
47	ARTICLE L324 NOT EXONERATING (EXCEPT TRANSPORT)
51	COEFFICIENT AND ACT IN RELATION
54	DIAGNOSIS AND TREATMENT OF STERILITY
55	6th MONTH OF PREGNANCY
56	SPECIAL EDUCATION WITHOUT REPORT
57	SPECIAL EDUCATION IN RELATION
62	OTHER CASES OF EXONERATION
63	HOSPIT PLUS 30 DAYS
81	PREVENTION
91	ACTS CCAM/NGAP EXONERANTS
92	CARE IN CONNECTION WITH AN EXONERANT CCAM ACT
99	UNKNOWN VALUE

Type of envelope

CPT_ENV_TYP	Description Type of envelope
0	SERVICES NOT TAKEN OVER BY RG
1	CARE OF THE CITY
2	HOSPITALIZATION AND MEDICO-SOCIAL AND INTERMEDIATE CONSUMPTION (MEDICAMENTS AND LPP)
3	LEGAL BENEFITS HEALTH INSURANCE OVER ONDAM ENVELOP
4	PREVENTION
5	FNAS
6	IVG REMOVES
7	SUPPLEMENTARY CMU
8	ALSACE MOSELLE
9	UNKNOWN VALUE
98	NOT APPLICABLE

Medical Specialty PS of prescriber

PSP_SPE_SNDS	LIB_PSP_SPE_SNDS	Description Medical Specialty PS Prescriber
0	NON_RENS	NOT FILLED IN
1	MED_GEN	GENERAL MEDICINE
2	AN_REAN	ANESTHESIOLOGY - REANIMATION
3	CARDIO	CARDIOVASCULAR PATHOLOGY
4	CHIRU	SURGERY
5	DERMA_VEN	DERMATOLOGY AND VENEROLOGY
6	RADIO	RADIOLOGY
7	GYNEC	OBSTETRIC GYNECOLOGY
8	GASTRO_ENT	GASTROENTEROLOGY AND HEPATOLOGY
9	MED_INT	INTERNAL MEDICINE
11	OTO_LAR	RHINO-LARYNGOLOGY OTO
12	PEDIA	PEDIATRICS
13	PNEUM	PNEUMOLOGY
14	RHUMA	RHEUMATOLOGY
15	OPHT	OPHTHALMOLOGY
18	STOM	STOMATOLOGY
31	MED_PHY	PHYSICAL MEDICINE AND READAPTATION
32	NEURO	NEUROLOGY
33	PSY	PSYCHIATRY
34	GERIA	GERIATRY
35	NEPHR	NEPHROLOGY
36	CHIR_DENT	DENTAL SURGERY (SPECIALIST O.D.F.)
37	ANAT_PATH	ANATOMY-CYTOLOGY-PATHOLOGY
38	MED_GEN	DIRECTOR LABORATORY DOCTOR
42	ENDO	ENDOCRINOLOGY AND METABOLISMS
80	MED_GEN	PUBLIC HEALTH AND SOCIAL MEDICINE
99	INCON	UNKNOWN VALUE

The executor is the practitioner, auxiliary or supplier who practices the act, provides the pharmacy or other products. In the case of a practitioner who performs his or her own acts, he or she is considered to be the executor. The prescriber is equivalent to any health professional who prescribes an act or product.

Example 1: A doctor is an executor for the consultation he or she gives to a patient.

Example 2: A general practitioner consults a patient and prescribes blood test and medication. We are talking about PS Executor for his consultation activity performed. We are talking about PS prescribing for his activity of prescription drugs and biology. The pharmacy that will deliver the drugs and the laboratory that will take the blood will in turn be considered as PS performing in their activity.

What is the difference between “PRS_ACT_NBR” and “PRS_ACT_QTE” enumeration?

The quantity is used to price and calculate the refund according to the following formula:

- Amount reimbursed=Basis of repayment x quantity of acts x repayment rate

The document count is only given for certain schemes (01 General Scheme, O6 ENIM, 08 CNMSS, 10 CRPCEN and 90 CAVIMAC) and is 0 for the other schemes.

To count acts, we use the variable "Quantity of acts or PRS_ACT_QTE" except in the following cases:

- Transportation: the quantity contains the number of invoices, while the count counts the number of routes (if 1 round-trip invoice: quantity=1; count=2);
- Mileage allowances: the quantity contains the number of kilometers invoiced;
- Daily allowances: the quantity contains the number of days compensated;
- Living expenses: the quantity contains the number of days hospitalized.

In our study, it is appropriate to use the variable PRS_ACT_QTE to count the number of days compensated.

What is the difference between the prefilled/crude indicators?

The distinction between the two types of indicators is based on the repayment type (PRS_REM_TYP).

- The prefilled indicators only concern the Compulsory Scheme (PRS_REM_TYP=0 or 1);
- Crude indicators cover all types of repayment.