Mémoire présenté le : 22/05/2015
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuaires

Par : Diane Desmonts

Sujet : Caractère facultatif sans sélection médicale d'un contrat collectif en prévoyance complémentaire : impact sur le taux d'adhésion et l'incidence de l'incapacité

Confidentialité : ☐ NON ■ **OUI** (Durée : ☐ 1 an ■ **2 ans**)

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

Membre présents du jury de l'Institut des Actuaires

Membres présents du jury de la filière

Secrétariat

Bibliothèque :

Entreprise :

Nom : SMACL Santé

Signature :

Directeur de mémoire en entreprise :

Nom : ***Laure Châtel***

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

Signature du candidat

Résumé

Le décret encadrant la participation des collectivités territoriales au financement de la protection sociale complémentaire de leurs agents a été publié en novembre 2011, plus de quatre ans après son homologue pour la fonction publique d'Etat.

La particularité de ce décret est qu'il impose, en prévoyance collective, une adhésion facultative sans sélection médicale pendant six mois pour les agents en activité à la date d'effet de la convention de participation. Une conséquence de cette contrainte pourrait être une mauvaise mutualisation des risques, entraînant un déséquilibre des résultats.

Nous avons fait l'hypothèse qu'il pouvait exister deux causes, non exclusives, au défaut de mutualisation des risques dans ce cadre contextuel :

- des taux d'adhésion trop faibles ;

- une prépondérance des adhésions d'agents représentant un risque plus élevé que le risque moyen pour lequel le tarif a été calculé.

Dans la mesure où l'étude a été réalisée moins de trois ans après la publication du décret, et où la prévoyance (hors frais de soins) est un risque long, nous avons surtout développé l'analyse de l'adhésion, et avons dégagé les premières tendances concernant la sinistralité.

Nous nous sommes intéressés dans un premier temps aux liens qui pouvaient exister entre certaines caractéristiques de la collectivité et le taux d'adhésion. De fortes disparités existent entre collectivités, les plus forts taux d'adhésion étant reliés à certains types de collectivités (e.g. communes) et dans des zones où les proportions d'HLM parmi les habitations principales sont les plus faibles. La participation de la collectivité à la cotisation semble avoir un effet positif sur le taux d'adhésion, mais son influence n'est pas la même selon la taille de la collectivité.

Au niveau individuel, la seule différence entre les deux sexes que nous avons mise en évidence est que les femmes adhérant avant 40 ans seraient en congé maladie indemnisé plus tôt que les hommes du même âge. Concernant l'âge, les taux d'adhésion sont plus faibles pour les agents les plus jeunes et les plus âgés, et l'incidence de l'incapacité indemnisée, comme son maintien, tendraient à augmenter avec l'âge. Suite à ces premières analyses, il ne semble pas exister de lien entre les agents qui adhèrent le plus et ceux qui représentent un plus fort risque.

Ces résultats seront cependant à confirmer avec des données et surtout un recul supplémentaires. Ils peuvent néanmoins être utilisés dès maintenant au minimum pour l'orientation du développement de l'activité.

Mots clefs

Collectivités territoriales, décret du 08 novembre 2011, prévoyance collective facultative, incapacité, participation, taux d'adhésion, régression logistique, modèle de Cox, surdispersion

Abstract

The decree that sets out the framework of local authority communities' financial contributions to the supplementary social protection of their employees was published in November 2011. It came four years after equivalent legislation concerning civil service employees.

The specificity of this decree in terms of group protection is the optional nature of adhesion, without medical based selection requirements, for newly hired employees during the first six months of their contract and as from the effective date of the participation agreement. A possible complication related to these constraints is an inappropriate risk pooling which could lead to unbalanced analytical results.

Within this theoretical framework we hypothesized that two non-exclusive factors can generate an inappropriate risk pooling.

- Low levels of adhesion.
- A high proportion of employees who present a greater risk compared to the average risk used to calculate rates.

Because the decree was only published three years ago, and because group protection (disregarding health care expenses) is a long-term risk, the current study focused on adhesion processes, and revealed emerging trends regarding claims rates.

We first examined the relationships between the characteristics of local authority communities and adhesion rates. Strong differences amongst local authority communities were found; higher adhesion rates were found in certain community categories (e.g. town councils) and were also observed in the areas where local authorities were not a major housing provider. Apparently, participation levels also positively influence adhesion rates, yet this influence varies with the size of the local authority community.

Considering individual levels of adhesion, the only difference found between sexes in our analyses is that women under 40 who join the optional protection system show a tendency to leave on a compensated disability sooner than men from the same age group. Regarding age groups, adhesion rates were lower for the youngest and the oldest employees whilst the beginning of compensated disability leave along with its maintenance tended to increase with age. The outcome of these first analyses is that there is no clear relationship between employee adhesion and risk levels.

Our results require further investigation, notably through implementation of longer-term data. Nonetheless, they are immediately useful – especially to guide business planning and development.

Keywords

Local authority communities, 2011 8th November's decree, optional group protection, disability, contribution, logistic regression, Cox proportional hazards model, overdispersion

Remerciements

Mes premiers remerciements s'adressent à mon Président, Robert Chiche, et mon Directeur Général, Nicolas Piotrowski, pour m'avoir permis d'intégrer SMACL Santé et de suivre cette formation en plus de mes fonctions au sein de l'entreprise.

Un immense merci également à Laure Châtel, sans qui ce mémoire n'aurait pu aboutir.

Je tiens également à remercier toutes les personnes ayant contribué de près ou de loin à la réalisation de ce travail, notamment notre Responsable grands comptes, Guylène de Moras, notre Responsable de la gestion, Christine Bouniot, ainsi que toute son équipe, et Bruno Massonnet.

Un merci particulier à Arnaud Cohen, pour ses divers conseils et aides depuis ma découverte du monde de l'actuariat en 2009.

Enfin, merci à ma famille et mes amis de m'avoir supportée, dans les deux sens du terme, tout au long de ces études qui ne s'arrêtent jamais…

Table des matières

Introduction 10

Partie I : La prévoyance dans les collectivités territoriales 13

A/ Statut des agents territoriaux et droits à congés maladie 13

1. Agents CNRACL 13

2. Agents IRCANTEC 17

B/ Environnement règlementaire de la prévoyance complémentaire des territoriaux 21

1. Décret n°2011-1474 du 8 novembre 2011 relatif à la participation des collectivités territoriales et de leurs établissements publics au financement de la protection sociale complémentaire de leurs agents 21

2. Loi n°89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques (dite 'Loi Evin') 24

Partie II : Présentation et préparation des données 27

A/ Données concernant les collectivités du portefeuille et leurs agents 27

1. Données disponibles 27

2. Montant de participation 28

3. Type de collectivité 30

4. Taille de la collectivité 31

5. Données complémentaires 34

6. Démographie des collectivités 35

B/ Données concernant les adhérents 36

1. Population assurée 37

2. Sinistralité 42

Partie III : Etude de l'adhésion 45

A/ Taux d'adhésion 45

1. Analyse descriptive 45

2. Présentation de la méthode utilisée 49

3. Application au taux d'adhésion 56

4. Synthèse concernant le taux d'adhésion 71

B/ Caractéristiques individuelles liées à l'adhésion 72

1. Choix d'adhérer ou non 72

2. Choix de la période d'adhésion 76

C/ synthèse générale concernant l'adhésion 80

Partie IV : Etude de la sinistralité 82

A/ Méthode utilisée 82
B/ Analyse de l'incidence de l'incapacité 84
1. Sélection du modèle 85
2. Analyse des résultats 87
3. Synthèse sur l'incidence de l'incapacité indemnisée 90
C/ Analyse du début du maintien en incapacité indemnisée 92
1. Sélection du modèle 92
2. Analyse des résultats 93
3. Synthèse concernant le début du maintien en incapacité 94
D/ Synthèse concernant la sinistralité 95
Conclusion 96
A/ Synthèse générale 96
B/ Perspectives 98
Bibliographie 99
Sources de données publiques 100
Mémoires d'actuariat consultés 101
Annexe 1 : Liens internet du décret du 08 novembre 2011 et des éléments associés 102
Annexe 2 : codes d'implémentation du test de Stukel sous SAS et du test des résidus de Schoenfeld sous R 103

Introduction

Avec près de 5,5 millions d'agents, la fonction publique représente à elle seule 20% de l'emploi en France (DGAFP, 2013). Comme les salariés du secteur privé, ces agents peuvent avoir à faire face à une perte de revenus lorsqu'ils sont en arrêt de travail pour raisons de santé, et être amenés à souscrire un contrat d'assurance complémentaire pour compenser, au moins partiellement, cette perte. Or les droits à congés pour raisons de santé ne sont pas les mêmes que ceux des salariés du secteur privé, et dans la majorité des cas la sécurité sociale n'intervient pas dans le versement des indemnités en cas de congé maladie. De plus, les contrats collectifs dans les établissements de la fonction publique ne peuvent pas faire l'objet d'une adhésion obligatoire pour leurs agents, contrairement au secteur privé, et il n'existe pas non plus de convention collective nationale pour les agents de la fonction publique. La prévoyance[1] complémentaire des agents de la fonction publique diffère donc de la prévoyance dans le secteur privé en plusieurs points et nécessite d'être traitée spécifiquement.

La fonction publique comprend trois versants : la Fonction Publique de l'Etat (FPE) pour 45% des emplois, la Fonction Publique Hospitalière (FPH) pour 21% des emplois et la Fonction Publique Territoriale (FPT) pour 34% des emplois (DGAFP, 2013). Si la Fonction Publique de l'Etat comprend majoritairement des agents de catégorie A (fonctions d'études générales, de conception et de direction) et seulement 20% d'agents de catégorie C (tâches d'exécution), la Fonction Publique Territoriale n'a, elle, que 9% d'agents de catégorie A contre 75% de catégorie C. Elle est donc constituée en majorité de personnes ayant des salaires[2] relativement bas, la moitié des agents ayant en effet un revenu net mensuel inférieur à 1636€ (revenu moyen 1823€). De plus, la Fonction Publique Territoriale est celle des trois fonctions publiques présentant la moyenne d'âge la plus élevée : 44 ans d'âge moyen, contre 42 ans dans la Fonction Publique de l'Etat et 41 ans dans la Fonction Publique Hospitalière, la moyenne d'âge dans le secteur privé étant, elle, de 40 ans. Enfin, même s'il a été montré qu'il est difficile de comparer directement les mesures du niveau d'absentéisme pour raisons de santé de chacun des trois versants de la fonction publique et celles réalisées dans le secteur privé (iFRAP, 2014), la tendance montre un niveau plus élevé dans la fonction publique, particulièrement dans la Fonction Publique Territoriale.

Les agents de la Fonction Publique Territoriale constituent donc une population particulièrement sensible en matière de prévoyance : ils peuvent se retrouver en-dessous du seuil de pauvreté s'ils restent longtemps en arrêt de travail, et moins de 50% d'entre eux sont couverts par un contrat de prévoyance complémentaire. Cette situation assez critique est également liée au fait que les collectivités territoriales et les établissements publics qui leurs sont associés[3], n'ayant pas l'obligation de participer au financement de la prévoyance complémentaire de leurs agents, et gênés par de trop faibles moyens financiers, participent très peu comparé aux entreprises privées. En revanche, s'ils désirent y participer, la procédure est maintenant précisément règlementée suite à la publication du décret du 8

[1] Tout au long de ce document, l'utilisation du terme 'prévoyance' ne concernera pas les frais de soins, et sera
[2] Le terme utilisé ci-après ne sera plus 'salaire' mais 'traitement', terme usuel pour la fonction publique.
[3] Le terme 'collectivités' sera utilisé pour mentionner les collectivités territoriales et les établissements publics qui leurs sont associés.

novembre 2011 (décret n° 2011-1474). Ce décret définit également les règles de mise en concurrence des opérateurs d'assurance pour l'attribution des contrats collectifs, ainsi que les conditions d'adhésion des agents à ces contrats. Il précise notamment que tout agent de la collectivité qui n'est pas en arrêt de travail à la date d'effet du contrat, a 6 mois pour adhérer sans questionnaire médical, l'adhésion restant facultative, tandis que l'opérateur ne peut refuser aucune adhésion au cours de ces 6 premiers mois.

Historiquement, il existait déjà des contrats collectifs à adhésion facultative avec participation éventuelle de l'employeur pour les agents des collectivités territoriales. Mais l'opérateur avait la possibilité de soumettre l'acceptation de l'adhésion à un questionnaire médical. Ce questionnaire était la plupart du temps supprimé au-delà d'un taux d'adhésion jugé suffisant pour la mutualisation des risques dans une collectivité donnée, ou au niveau d'un segment du portefeuille de l'opérateur d'assurance (la majorité des agents couverts en prévoyance complémentaire étaient répartis entre deux mutuelles avant la sortie du décret, une seule en regroupant plus de la moitié). Les conditions du décret entraînent donc pour les acteurs de l'assurance, qu'ils soient historiques ou plus récents sur ce marché, une situation nouvelle qui combine un contrat collectif à adhésion facultative à l'absence de possibilité de sélection du risque. La mise en concurrence les oblige néanmoins à proposer des tarifs les plus compétitifs possibles, donc avec une marge de prudence très limitée.

Les préoccupations ne sont cependant pas les mêmes pour un opérateur historique que pour un opérateur plus récent n'ayant pas encore de portefeuille significatif. En effet, même si beaucoup d'agents des collectivités ne sont pas encore couverts en prévoyance complémentaire (seulement 33% étaient couverts avant la mise en place du décret), la priorité pour un opérateur historique est de conserver son portefeuille. C'est notamment le cas lorsque les collectivités participant déjà, ou souhaitant participer au financement de la prévoyance complémentaire de leurs agents, doivent se mettre en conformité avec le décret. L'opérateur plus récent, qui lui est en phase de conquête de portefeuille, doit le faire rapidement pour deux raisons principales : avoir une part significative du marché pour être reconnu, ce qui facilitera son développement, et atteindre un nombre d'assurés permettant une bonne mutualisation du risque.

Pour les deux types d'opérateurs, mais surtout pour celui en phase de conquête de portefeuille, il est essentiel de déterminer quel peut être l'effet du caractère facultatif sans sélection de ces contrats collectifs, à la fois sur l'adhésion aux contrats et sur leur sinistralité. Savoir quelles sont les caractéristiques des agents qui adhèrent, dans quelles proportions ils le font, à quel moment par rapport à la date d'effet du contrat, permet dans un premier temps d'orienter les campagnes commerciales pour les rendre plus efficaces et d'optimiser l'investissement dans l'effort de développement. Dans un second temps, combinées à l'analyse de la sinistralité, ces informations contribueront à déterminer s'il existe de l'anti-sélection sur ces contrats (*i.e.* si les agents qui choisissent d'y adhérer représentent un risque plus élevé que ceux qui choisissent de ne pas adhérer), ou encore si certains types ou tailles de collectivités peuvent représenter un risque plus élevé. L'application ultime étant l'ajustement des tarifs en fonction de critères pertinents. Cependant, la prévoyance étant un risque long, l'étude de la sinistralité ne peut pas être menée en totalité moins de trois ans après la mise en place du décret et ces derniers points ne pourront pas être traités dans cette étude. Il sera néanmoins possible de commencer à dégager des tendances concernant l'incidence de l'incapacité et le début du maintien en incapacité.

Cette étude concerne le portefeuille en prévoyance collective complémentaire de SMACL Santé, une mutuelle du livre II du code de la mutualité, créée en 2007 pour couvrir spécifiquement les agents des collectivités territoriales en protection sociale complémentaire (santé et prévoyance). La SMACL Santé a connu un fort développement depuis la mise en place du décret du 8 novembre 2011, en particulier en prévoyance puisque son nombre d'adhérents à des contrats collectifs est passé d'environ 2 000 au 31 octobre 2011 à plus de 36 200 au 31 août 2014. Il a ainsi été multiplié par 18 en moins de trois ans. SMACL Santé est donc un opérateur récent sur le marché de la prévoyance complémentaire des agents territoriaux, mais le volume atteint par son portefeuille collectif depuis la mise en place du décret permet d'ores et déjà d'apporter des réponses aux questions suivantes :

- Quels sont les taux d'adhésion aux contrats collectifs mis en place dans le cadre du décret du 8 novembre 2011, et comment varient-ils en fonction de différentes caractéristiques de la collectivité comme sa taille ou son type (*e.g*. commune) ?

- Quels sont les agents qui choisissent d'adhérer à un tel contrat ? Quand les agents choisissent-ils d'adhérer : à la date d'effet du contrat ou plus tard ?

- Le montant de la participation de la collectivité a-t-il un effet sur le taux d'adhésion ?

- Quelles sont les caractéristiques individuelles des adhérents présentant une incidence plus précoce de l'incapacité ? Sont-elles les mêmes que celles des agents semblant rester plus longtemps en incapacité ? Ces caractéristiques individuelles correspondent-elles à celles des agents qui adhèrent le plus aux contrats collectifs de prévoyance complémentaire ?

L'étude présentée ci-dessous comprend quatre parties principales :

- La première partie donnera les éléments clefs du cadre contextuel et conceptuel afin de poser clairement les bases nécessaires à la compréhension : elle décrira les droits à congés maladie des agents de la fonction publique et l'environnement règlementaire spécifique à la prévoyance complémentaire des agents territoriaux ;

- La seconde partie présentera les données ainsi que leur préparation pour les analyses ;

- La troisième partie s'intéressera aux éléments influençant le taux d'adhésion et la structure de la population adhérente en comparaison de la population d'origine ;

- La quatrième partie traitera de l'incidence de l'incapacité en prévoyance complémentaire et des premiers mois de son maintien.

Partie I : La prévoyance dans les collectivités territoriales

A/ Statut des agents territoriaux et droits à congés maladie

Selon leur statut (titulaire ou non titulaire) et leur quotité travaillée (plus ou moins de 28h hebdomadaires), les agents des collectivités territoriales sont rattachés à la CNRACL (Caisse Nationale de Retraite des agents des Collectivités Locales) ou à l'IRCANTEC (Institut de Retraite Complémentaire des Agents Non Titulaires de l'Etat et des Collectivités publiques). Cette affiliation détermine l'intervention ou non de la sécurité sociale, leurs droits à congés maladie et leurs droits en cas d'invalidité. En 2011, parmi les 1,83 millions de personnes travaillant dans la Fonction Publique Territoriale, environ 1,3 millions étaient des agents titulaires affiliés à la CNRACL, et 0,5 million des agents affiliés à l'IRCANTEC (dont environ 0,4 million de non-titulaires et 0,1 million de titulaires) (DGAFP, 2013).

NB$_1$: Le traitement des agents des collectivités territoriales se décompose en traitement de base et régime indemnitaire (RI ou 'primes'), ce dernier pouvant représenter jusqu'à plus de 30% du revenu annuel d'un agent. Les droits à congés maladie définis par le statut ne concernent que le traitement de base, les collectivités ayant toute liberté quant au maintien ou non, et au niveau de maintien, du RI (hors supplément familial de traitement et indemnité de résidence qui sont maintenus intégralement), en l'absence de transposition aux collectivités territoriales du décret (n° 2010-997 du 26 août 2010) prévoyant le maintien des primes et indemnités pour les agents publics de l'Etat (Collectivites-locales.gouv, 2010).

NB$_2$: Les droits à congés maladie sont calculés en année glissante, c'est-à-dire que pour chaque jour d'arrêt, la rémunération à laquelle a droit l'agent dépend du nombre de jours d'arrêts maladie dans les 364 jours précédents.

1. Agents CNRACL

Les agents titulaires et stagiaires (période probatoire avant la titularisation, durant en général un an) travaillant à temps complet ou temps non complet plus de 28h par semaine sont affiliés à la CNRACL. Lorsqu'ils sont en arrêt maladie, la sécurité sociale n'intervient pas dans la rémunération. C'est la collectivité qui maintient le traitement selon les modalités décrites ci-après. Sont exclus de ces droits à congés les agents en congé parental ou en disponibilité pour raisons personnelles.

Il existe cinq types de congés donnant droit à cinq régimes de rémunération :

- **Maladie Ordinaire** (MO) : dans ce régime entrent tous les arrêts maladie qui ne peuvent donner droit à aucun autre régime. Le plein traitement est maintenu pendant 3 mois, puis la collectivité ne verse plus que 50% de son traitement à l'agent pendant 9 mois.
- **Longue Maladie** (LM) : ce type de congé est accordé si la maladie présente un caractère invalidant et de gravité confirmée et nécessite un traitement et des soins prolongés. Le plein traitement est maintenu pendant 1 an, puis réduit à demi-

traitement pendant 2 ans. Pour avoir droit à un autre congé de longue maladie, l'agent doit avoir repris l'exercice de ses fonctions pendant un an.

- **Longue Durée** (LD) : ce régime n'est accordé que pour cinq grandes catégories de maladies. Il s'agit des cas de maladie mentale, tuberculose, affection cancéreuse, poliomyélite ou déficit immunitaire grave et acquis. Le plein traitement est maintenu pendant 3 ans, puis réduit à demi-traitement pendant 2 ans. L'agent n'a droit qu'à un congé de ce type par catégorie de maladie au cours de sa carrière. Si la maladie a été contractée en service, le plein traitement est maintenu 5 ans et le demi-traitement 3 ans.

  NB : Sauf dans le cas où l'agent n'a plus droit au plein traitement en LM, le congé de LD ne peut être accordé qu'à l'issue de la période à plein traitement de LM, cette dernière constituant néanmoins une période de LD à plein traitement pour la maladie en cause.
- **Congé pour accident de service ou maladie ayant une cause exceptionnelle** (AT/MP) : ce régime est accordé si la maladie résulte d'un accident de service, ou a été contractée ou aggravée soit en service, soit en accomplissant un acte de dévouement dans un intérêt public, soit en exposant ses jours pour sauver la vie d'une ou plusieurs personnes. Le traitement est alors maintenu en intégralité jusqu'au retour au travail de l'agent ou jusqu'à sa mise en retraite.
- **Congé pour infirmité de guerre** : a droit à ce type de congé tout agent atteint d'infirmités contractées ou aggravées au cours d'une guerre ou d'une expédition déclarée campagne de guerre. Le traitement est maintenu en totalité, pour une durée de deux ans maximum au cours de la carrière.

A l'expiration des droits à congés maladie, l'agent qui n'est pas apte à reprendre en totalité ses fonctions peut :

- Bénéficier d'un **Temps Partiel Thérapeutique** (TPT) s'il en a la capacité et la nécessité (pour favoriser l'amélioration de son état de santé ou s'il doit faire l'objet d'une rééducation ou d'une réadaptation professionnelle). Ce temps partiel ne peut en aucun cas être inférieur au mi-temps, et est accordé pour un an maximum. L'ouverture des droits au TPT peut se faire : après 6 mois consécutifs de congé en MO pour une même affection, après un congé de LM, après un congé pour maladie de LD, après un congé pour AT/MP. Pendant cette période, l'agent perçoit la totalité de son traitement et bénéficie de ses droits à l'avancement et à la retraite ;
- Etre **reclassé** pour inaptitude physique s'il est reconnu apte à exercer une autre activité ;
- Etre placé en **Disponibilité d'Office** (DO), pour un maximum d'un an renouvelable deux fois, avec une possible dérogation pour un troisième renouvellement si le comité médical estime que l'agent reprendra ses fonctions avant la fin de la quatrième année. La durée maximale est donc de quatre ans. Pendant cette période, l'agent cesse de bénéficier de ses droits à l'avancement et à la retraite, et ne perçoit plus de traitement. Cependant, il peut percevoir une indemnité de coordination par la collectivité, pendant une durée maximum de deux ans, uniquement après une Maladie Ordinaire, qui correspond à 50% du traitement de base et de l'indemnité de résidence + 100% du supplément familial (66.66% à la place de 50% pour les agents ayant 3 enfants à charge ou plus). Après ces deux années, ou suite à une Longue

Maladie ou une maladie de Longue Durée, l'agent peut demander à percevoir, s'il est temporairement inapte et atteint d'une invalidité réduisant aux 2/3 sa capacité de travail, une Allocation d'Invalidité Temporaire (AIT), après avis de la Caisse Primaire d'Assurance Maladie (CPAM) et de la Commission de Réforme (CR). Le montant de l'AIT suit les mêmes règles que le montant de la pension d'invalidité de la Sécurité sociale (30% pour le 1er groupe, 50% pour les 2è et 3è groupes).

- Etre admis à la **retraite pour invalidité** s'il n'a pas atteint l'âge limite d'activité (entre 65 et 67 ans selon l'année de naissance), qu'il ne peut pas être reclassé dans un emploi correspondant à ses aptitudes physiques et qu'il est reconnu définitivement inapte par la Commission de Réforme et la CNRACL. Il ne s'agit pas d'une mise en invalidité comme pour les salariés du secteur privé mais bien d'une mise en retraite pour invalidité : l'agent perçoit sa pension de retraite dès sa mise en invalidité. De plus, la mise en retraite d'office à l'initiative de la collectivité ne peut être prononcée qu'à l'expiration des droits à congé maladie.

Dans le cadre de la **couverture complémentaire en prévoyance, l'assureur peut avoir à intervenir sur la Maladie Ordinaire, la Longue Maladie et la maladie de Longue Durée lorsque l'agent est à demi-traitement**. Il est **parfois également demandé que l'assureur intervienne sur les congés sans traitement pour raison de santé, soit la Disponibilité d'Office et le temps partiel pour raison de santé qui fait suite au Temps Partiel Thérapeutique** quand l'agent ne peut pas reprendre son activité à temps plein (ou à la quotité en vigueur avant son arrêt maladie).

La figure 1 ci-après schématise les droits à congés maladie des agents CNRACL avec les périodes et niveaux d'intervention de la collectivité, ainsi que l'intervention potentielle d'une couverture complémentaire. La hauteur d'un rectangle correspond à 100% du traitement. Le niveau d'intervention de chacune des parties correspond au maximum possible.

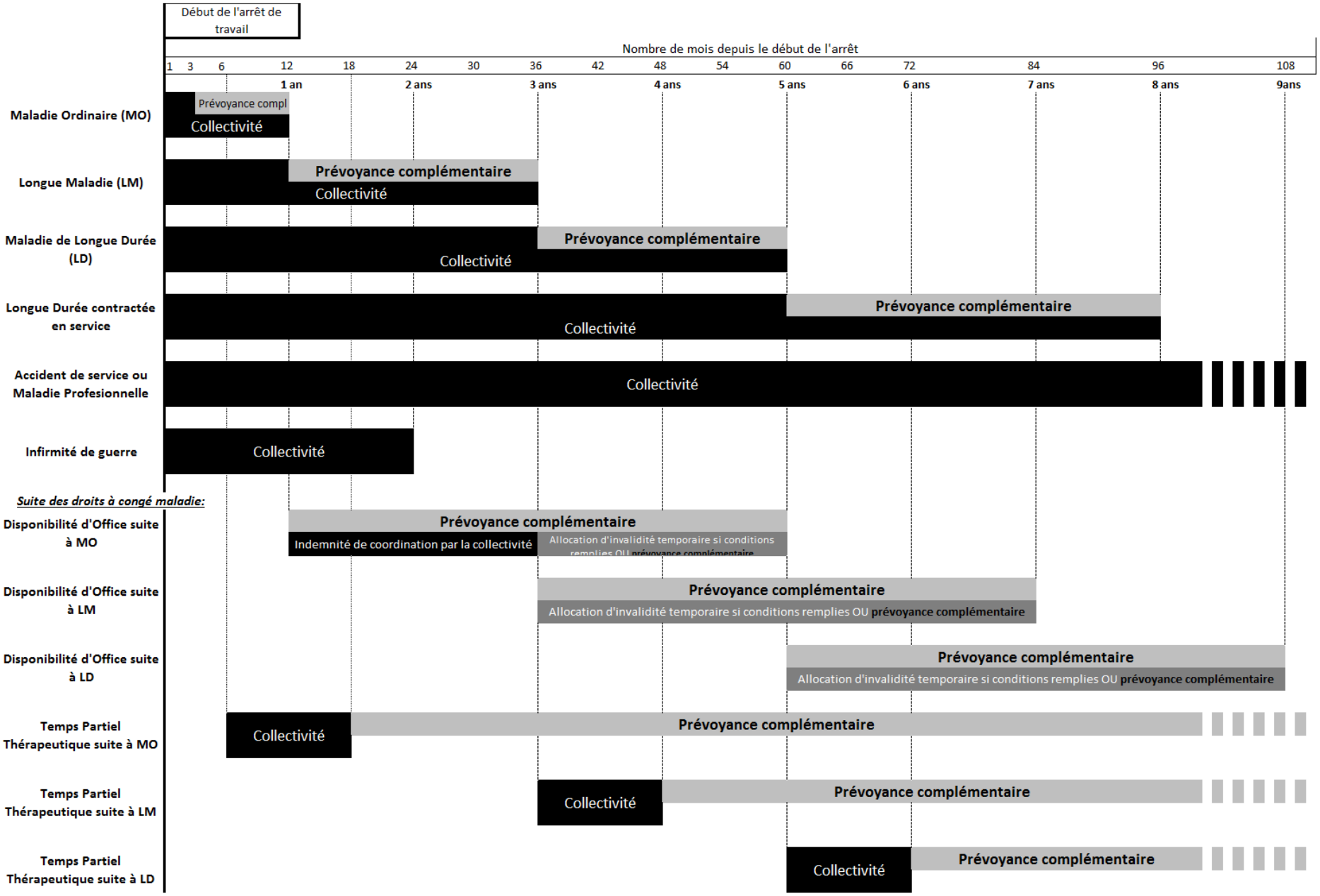


Figure 1: Droits à congés maladie des agents titulaires et stagiaires affiliés à la CNRACL et intervention potentielle de l'opérateur en prévoyance complémentaire

2. Agents IRCANTEC

Les agents titulaires et stagiaires travaillant moins de 28h par semaine, et les agents non titulaires sont affiliés à l'IRCANTEC. La sécurité sociale intervient avec la collectivité dans les prestations en cas d'arrêt maladie, en proportions variables selon que l'agent travaille moins de 200 heures ou plus de 200 heures par trimestre.

Les droits diffèrent entre les titulaires et stagiaires, et les non titulaires.

a/ Droits des agents IRCANTEC titulaires et stagiaires

- **Maladie Ordinaire** (MO) : les droits sont les mêmes que pour les agents CNRACL, soit 3 mois à plein traitement puis 9 mois à demi-traitement (les 50% sont portés à 66% si 3 enfants à charge) ;
- **Grave Maladie** (GM) : ce type de congé est l'équivalent de la Longue Maladie pour les agents CNRACL. Les agents ont leur plein traitement pendant 1 an puis un demi-traitement pendant 2 ans (les 50% sont portés à 66% si 3 enfants à charge) ;
- **Accident du Travail ou Maladie Professionnelle** (AT/MP) : L'agent bénéficie de ce type de congé jusqu'à la guérison ou la consolidation de la blessure. Il perçoit son plein traitement pendant 3 mois, puis 80% jusqu'à son retour au travail ou sa mise en retraite.

A l'expiration des droits à congés maladie, l'agent qui n'est pas apte à reprendre en totalité ses fonctions peut :

- Bénéficier d'une **reprise d'activité partielle pour motif thérapeutique**. La collectivité versera le traitement de l'agent au prorata du temps de travail effectué. La CPAM pourra, sous certaines conditions, indemniser en tout ou partie la perte de traitement de l'agent ;
- Etre **reclassé** pour inaptitude physique s'il est reconnu apte à exercer une autre activité ;
- Etre placé en **Disponibilité d'Office** (DO) aux mêmes conditions de durée que les agents CNRACL s'il est temporairement inapte ;
- Etre reconnu **invalide** ou en **incapacité permanente** s'il remplit les conditions définies par la Caisse Primaire d'Assurance Maladie (CPAM).

Dans le cadre de la **couverture complémentaire en prévoyance, l'assureur peut avoir à intervenir sur la Maladie Ordinaire et la Grave Maladie lorsque l'agent est à demi-traitement**. Il peut également lui être demandé d'intervenir sur l'**Accident de Travail/Maladie Professionnelle et la Disponibilité d'Office**.

La figure 2 ci-après schématise les droits des agents IRCANTEC titulaires et stagiaires avec les périodes et niveaux d'intervention de la collectivité et de la sécurité sociale, ainsi que l'intervention potentielle d'une couverture complémentaire (Le 1er mois présente une sous-partie correspondant aux 3 jours de carence de la sécurité sociale ; lorsque le niveau d'indemnisation diffère de 50% ou 100%, il est indiqué en chiffres sur le schéma ; la reprise d'activité partielle pour motif thérapeutique n'est pas représentée car trop complexe à schématiser.).

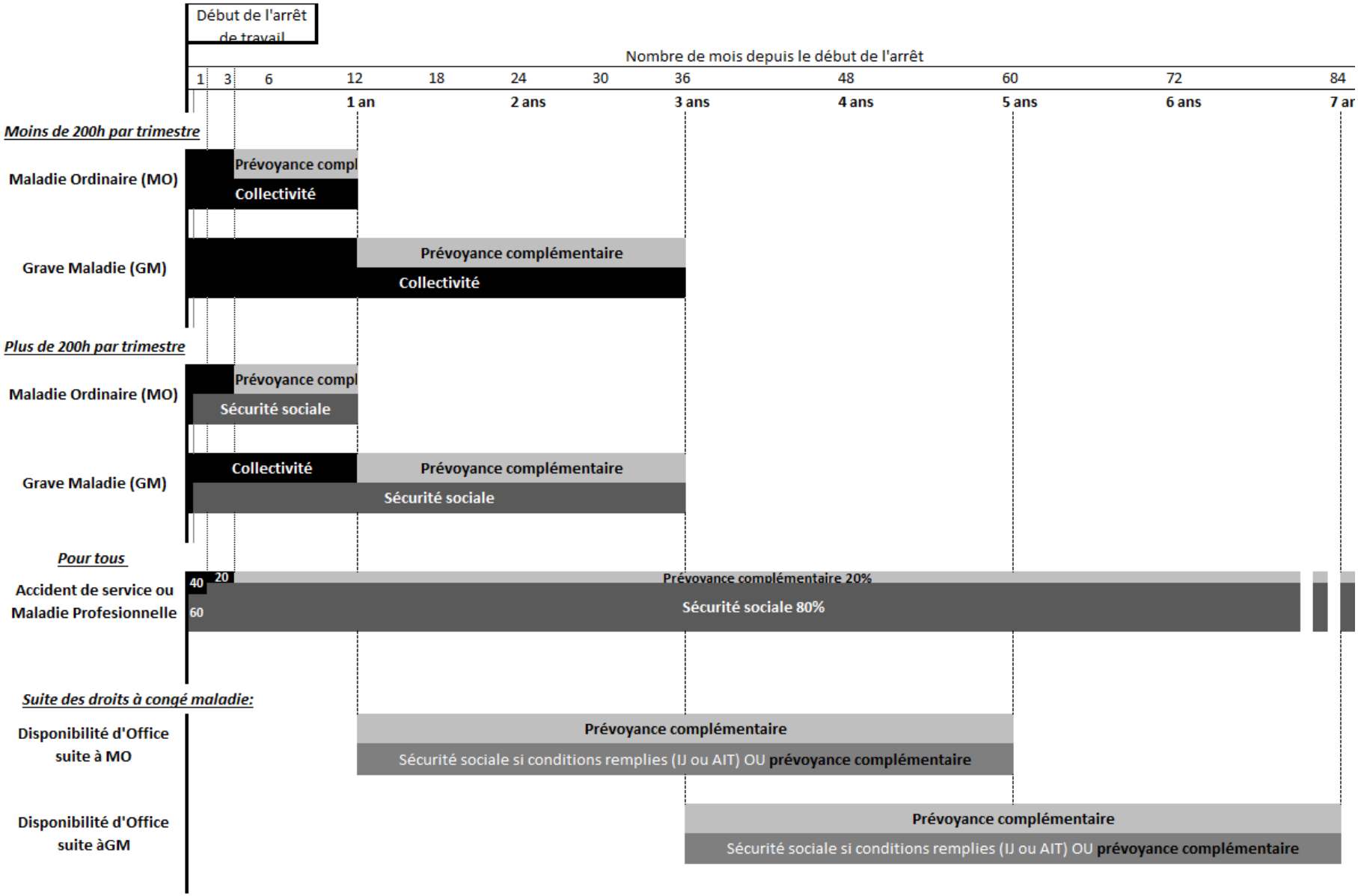


Figure 2: Droits à congés maladie des agents titulaires et stagiaires travaillant moins de 28h par semaine (affiliés à l'IRCANTEC) et intervention potentielle de l'opérateur en prévoyance complémentaire

b/ Droits des agents non titulaires

- **Maladie Ordinaire** (MO) : les droits sont calculés sur une période de 12 mois consécutifs ou 300 jours de travail effectif en cas de service discontinu. La durée du congé et le versement de tout ou partie du traitement varient en fonction de l'ancienneté de l'agent en qualité de non titulaire (fig. 3), qui doit être de 4 mois minimum pour ouvrir des droits. En cas d'ancienneté insuffisante, l'agent est soit placé en congé sans traitement pour un an au maximum, soit licencié si l'incapacité est permanente ;
- **Grave Maladie** (GM) : les causes d'attribution de ce congé sont les mêmes que pour les IRCANTEC titulaires. Pour prétendre à ces droits, l'agent doit avoir une ancienneté minimale de 3 ans continus. Il conserve son plein traitement pendant 1 an puis perçoit un demi-traitement pendant 2 ans. En cas d'ancienneté insuffisante, l'agent est soit placé en congé sans traitement pour un an au maximum, soit licencié si l'incapacité est permanente.
- **Accident du Travail ou Maladie Professionnelle** (AT/MP) : L'agent bénéficie de ce type de congé jusqu'à la guérison ou la consolidation de la blessure. Il perçoit son plein traitement en fonction de son ancienneté puis 80% de ce dernier.

A l'expiration des droits à congés maladie, l'agent qui n'est pas apte à reprendre en totalité son activité peut :

- Bénéficier d'une **reprise d'activité partielle pour motif thérapeutique**. La collectivité versera le traitement de l'agent au prorata du temps de travail effectué. La CPAM pourra, sous certaines conditions, indemniser en tout ou partie la perte de traitement de l'agent ;
- Etre **reclassé** s'il est reconnu apte à exercer une autre activité ou, en cas d'impossibilité, licencié ;
- Etre placé en **congé sans traitement** pour une durée maximale d'1 an s'il est inapte temporairement.

Dans le cadre de la **couverture complémentaire en prévoyance, l'assureur peut avoir à intervenir sur la Maladie Ordinaire et la Grave Maladie lorsque l'agent est à demi-traitement**. Il peut également lui être demandé d'intervenir sur l'**Accident de Travail/Maladie Professionnelle et le congé sans traitement pour raisons de santé**.

La figure 3 ci-après schématise les droits des agents IRCANTEC non titulaires avec les périodes et niveaux d'intervention de la collectivité et de la sécurité sociale, ainsi que l'intervention potentielle d'une couverture complémentaire.

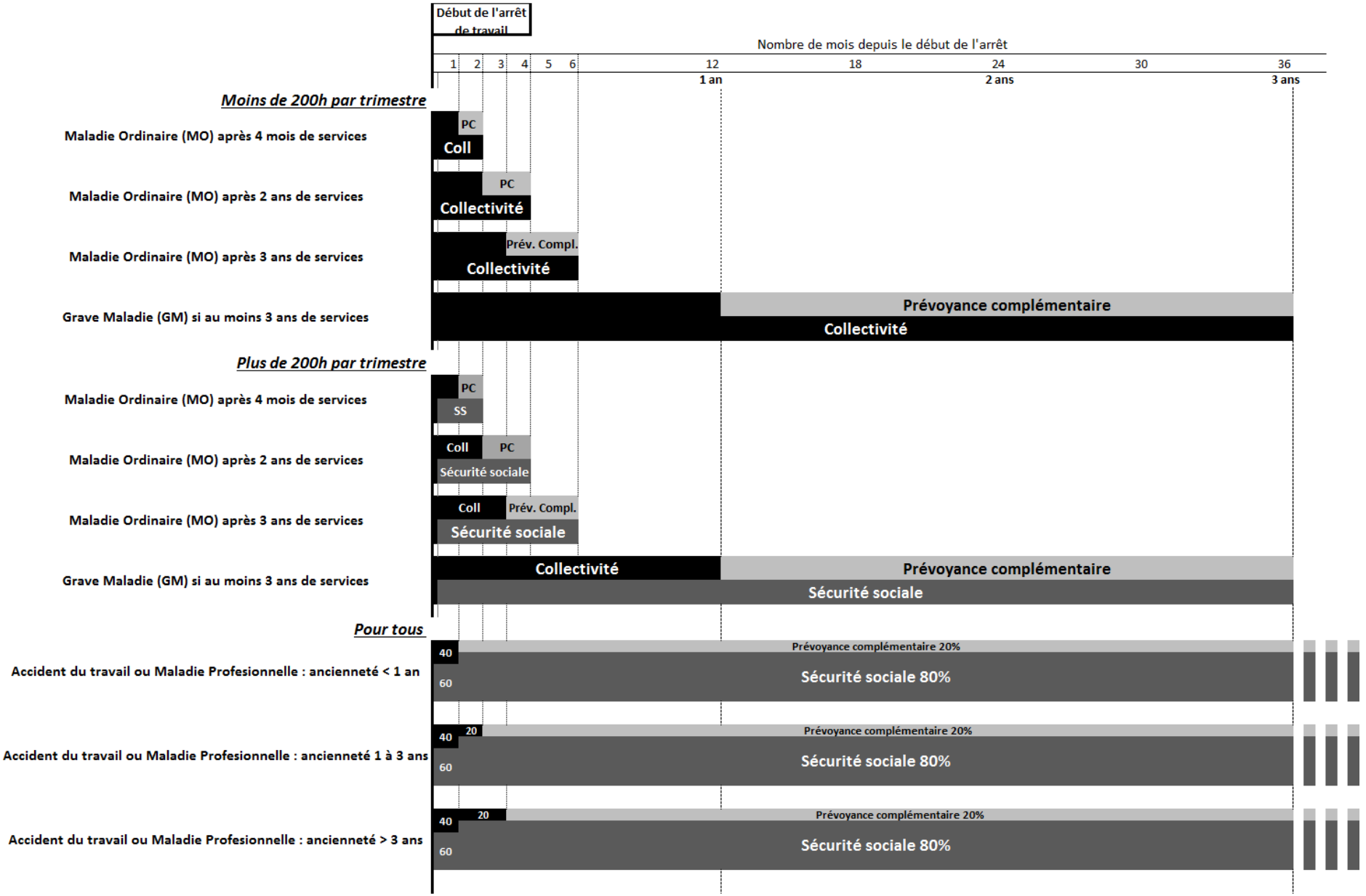


Figure 3: Droits à congés maladie des agents non titulaires affiliés à l'IRCANTEC et intervention potentielle de l'opérateur en prévoyance complémentaire

B/ Environnement règlementaire de la prévoyance complémentaire des territoriaux

Historiquement, l'article R. 523-2 de l'ancien code de la mutualité et l'arrêté interministériel du 19 septembre 1962 (« arrêté Chazelle ») autorisaient la participation financière de l'employeur public au financement des mutuelles de ses agents, que cette aide soit directe, ou indirecte par la mise à disposition de personnels et de locaux. Suite à une plainte contre la Mutualité Fonction Publique (MFP) et ses mutuelles membres, la Commission Européenne a estimé que le soutien de l'Etat en faveur de ces dernières n'était plus compatible avec l'évolution du marché commun, et proposait à la France, en juillet 2005, de revoir l'ensemble du dispositif juridique. Parallèlement à cela, le Conseil d'Etat remettait en cause l'arrêté Chazelle en septembre 2005. Différents travaux ont été conduits et ont abouti à la loi de modernisation de la Fonction Publique le 2 février 2007 (n°2007-148), contenant un article (*art.* 39) permettant aux employeurs des trois fonctions publiques de participer au financement de la protection sociale complémentaire de leurs agents, à condition qu'elle soit attribuée à des contrats mettant en œuvre des critères de solidarité. Les modalités d'application de cette loi devaient être fixées par des décrets en Conseil d'Etat : un pour chacun des versants de la Fonction Publique. Le premier décret publié a été celui de la Fonction Publique d'Etat, le 19 septembre 2007 (n°2007-1373). Le décret pour la Fonction Publique Hospitalière n'est toujours pas publié, et celui pour la Fonction Publique Territoriale l'a été bien plus tard que pour la Fonction d'Etat puisqu'il date du 8 novembre 2011 (décret n° 2011-1474).

Ce décret encadre la prévoyance complémentaire des agents territoriaux en définissant les règles de participation des collectivités au financement, et les critères d'éligibilité des contrats et règlements auxquels leurs agents choisissent d'adhérer. Ces critères ont pour vocation de mettre en place une aide à caractère social permettant de maintenir une solidarité entre les bénéficiaires.

L'application de ce décret doit également se combiner avec le respect de la loi Evin, qui concerne tous les contrats d'assurance en prévoyance complémentaire.

1. *Décret n°2011-1474 du 8 novembre 2011 relatif à la participation des collectivités territoriales et de leurs établissements publics au financement de la protection sociale complémentaire de leurs agents*

a/ Généralités

Ce décret (Annexe 1) définit deux modes possibles de participation des collectivités à la protection sociale souscrite par leurs agents (*i.e.* santé et prévoyance) : la **labellisation** et la **convention de participation**[4]. Il précise également que « *la souscription à une protection sociale complémentaire ou à un mécanisme de participation est facultative pour les agents et les collectivités.* ». Les collectivités n'ont donc aucune obligation de participer à la protection sociale complémentaire de leurs agents, et dans le cas où la collectivité choisit de participer,

[4] La convention de participation dans le cadre du décret, correspond à un contrat collectif à adhésion facultative ; c'est ce terme qui sera utilisé dans la suite du document.

aucun agent n'est dans l'obligation de souscrire au contrat ou adhérer au règlement retenu par la collectivité (*art.* 3). La participation vient en déduction du montant de cotisation ou de la prime due par l'agent et peut-être versée soit à l'agent, soit à l'organisme assureur (*art.* 24). Dans le but de maximiser la solidarité, la collectivité peut décider d'ajuster le montant de la participation au revenu des agents et, le cas échéant, à leur situation de famille (*art.* 23). En prévoyance, les contrats et règlements doivent prévoir au minimum la couverture du risque incapacité de travail (*art.* 30). Tous les agents qui souhaitent adhérer à une convention de participation ou à un contrat labellisé doivent au minimum choisir l'incapacité. C'est de cette garantie uniquement que traiteront les analyses dans la suite de ce travail.

Si une collectivité décide de participer au financement de la protection sociale complémentaire de ses agents, elle doit obligatoirement choisir, après avis du comité technique, entre labellisation et convention de participation (*art.* 4).

b/ Labellisation

Dans le cas de la labellisation, la collectivité indique à l'agent où trouver la liste des contrats et règlements labellisés (site du ministère chargé des collectivités territoriales), et ce dernier fait son choix de façon autonome. Pour qu'un contrat ou règlement puisse être labellisé, le dossier doit au préalable être étudié et validé par un prestataire habilité par l'Autorité de Contrôle Prudentiel et de Résolution (ACPR ; *art.* 6). Le label est délivré pour une période de 3 ans (*art.* 7), renouvelable (*art.* 9). La souscription d'un contrat ou l'adhésion à un règlement labellisé fonctionne sur le même principe que l'assurance individuelle et peut donc faire l'objet d'une sélection médicale.

c/ Convention de participation

Une convention de participation est conclue entre une collectivité et un organisme assureur. Il s'agit d'un contrat collectif à adhésion facultative dont la mise en place est très précisément règlementée.

Dans un premier temps, la collectivité doit publier un avis d'appel public à la concurrence dans deux, voire trois, types de publications (une habilitée à recevoir des annonces légales et une spécialisée dans le secteur des assurances, et dans certains cas le Journal Officiel de l'Union Européenne). Dans cet avis sont détaillés :

- « *Les modalités de présentation des offres de candidature, dont le délai de réception ne peut être inférieur à quarante-cinq jours à compter de la date de publication de l'avis d'appel à la concurrence ;*
- *Les niveaux minimaux de capacité demandés aux candidats et les renseignements à fournir à cet effet ;*
- *Les caractéristiques essentielles de la convention envisagée, notamment son objet, sa nature et les personnels intéressés ;*
- *Les critères de choix de la collectivité territoriale ou de l'établissement public.* » (*art.* 15).

Dans un second temps, la collectivité envoie aux organismes qui en ont fait la demande un cahier des charges décrivant les prestations à proposer, ainsi que des informations quantitatives et qualitatives concernant la population à assurer (*art.* 16). La collectivité a la possibilité de faire appel à un cabinet de conseil pour constituer le cahier des charges.

Dans un troisième temps, chaque candidat renvoie sa proposition à la collectivité en respectant la date limite de réponse. Cette proposition doit détailler les conditions d'adhésion, les prestations offertes, le tarif proposé pour chacune des options et une précision du degré effectif de solidarité et de la maîtrise financière du dispositif envisagé (*art.* 17). Le **tarif** proposé est exprimé en **proportion du traitement** de l'agent et doit être au **même taux pour tous les agents** (*art.* 31).

Enfin, après examen des différentes propositions, délibération et avis du comité technique, la collectivité retient l'organisme répondant au mieux aux principes de solidarité fondant le décret, et le mieux noté selon les critères ci-dessous :

- « *Le rapport entre la qualité des garanties et le tarif proposé ;*
- *Le degré effectif de solidarité entre les adhérents ou les souscripteurs, intergénérationnelle, en fonction de la rémunération […];*
- *La maîtrise financière du dispositif ;*
- *Les moyens destinés à assurer une couverture effective des plus âgés et des plus exposés aux risques ;*
- *Tout autre critère objectif respectant l'obligation de transparence et de non-discrimination, adapté à la couverture de la population intéressée.* » (*art.* 18)

L'organisme retenu signe une convention de participation avec la collectivité pour une durée de 6 ans, renouvelable 1 an au maximum, et doit appliquer les règles suivantes :

- Il s'engage à accepter, pendant les **6 premiers mois de la convention**, l'**adhésion** de **tous** les **agents en activité** à la date d'effet de la convention **sans questionnaire médical**.
  « *L'adhésion des agents au contrat ou au règlement ne peut être conditionnée par leur âge ou leur état de santé. Les agents qui ne sont pas en arrêt de travail à la date d'effet du contrat ou du règlement et qui ne sont pas inscrits à celui-ci peuvent y adhérer sous réserve que leur inscription intervienne pendant les six premiers mois qui suivent la date de prise d'effet du contrat ou du règlement. Les agents embauchés postérieurement à la date de prise d'effet du contrat ou du règlement peuvent y adhérer sous réserve que leur inscription intervienne dans les six premiers mois qui suivent la date d'embauche.* » (*art.* 31) ;
- Dans les 6 premiers mois de la convention, il **ne peut pas refuser l'adhésion des agents en arrêt de travail** mais peut **définir** les **conditions** d'adhésion de ces derniers.
  « *Les agents en arrêt de travail peuvent adhérer au contrat ou au règlement dans les conditions prévues par celui-ci.* » (*art*. 31) ;
- **Au-delà** des 6 premiers mois, il peut soumettre l'adhésion des agents à un **questionnaire médical** et même **refuser** leur adhésion.
  « *Passé ce délai de six mois suivant la date de prise d'effet du contrat ou du règlement, ou la date d'embauche, si l'adhésion au titre du contrat ou du règlement est acceptée, elle pourra être subordonnée à une tarification différente fondée sur un questionnaire médical.* » (*art.* 31) ;

- Il ne peut augmenter ses tarifs que dans les cas énumérés ci-dessous, et si le changement « *revêt un caractère significatif* » (*art.* 20) :
  « *1° Aggravation de la sinistralité*
  2° Variation du nombre d'agents […] adhérents […]
  3° Evolutions démographiques
  4° Modifications de la réglementation » ;
- Durant toute la durée de la convention, il ne peut résilier par décision unilatérale ;
- Il doit produire, aux termes de la 3è et de la 6è années, un rapport « *retraçant les opérations réalisées au vu des critères des 2° et 4° de l'article 18* », afin que la collectivité puisse vérifier si le degré effectif de solidarité prévu à l'origine a bien été respecté et si les moyens destinés à assurer une couverture effective des plus exposés aux risques ont bien été mis en œuvre. Si tel n'est pas le cas, la collectivité a le droit de résilier la convention par décision unilatérale (*art.* 19). La collectivité peut aussi décider à tout moment de résilier la convention, en justifiant sa décision, si elle constate que l'organisme ne respecte plus les dispositions du décret (*art.* 21).

Ce décret, en conditionnant l'attribution de la participation au respect de critères de solidarité, vise notamment à maximiser la mutualisation des risques. Cependant, dans le cas des conventions de participation, l'adhésion étant facultative et sans sélection médicale, si le taux d'adhésion est faible et/ou si la population adhérente n'est pas représentative de la population initiale pour laquelle le tarif avait été calculé, l'objectif de mutualisation des risques sera difficilement atteint. Il est donc essentiel, au-delà de l'intérêt commercial que cela peut apporter, de déterminer quels sont les taux d'adhésion aux conventions de participation, à quels facteurs sont liées leurs variations, ou encore quels sont les agents qui choisissent d'adhérer, pour ensuite pouvoir mettre ces résultats en relation avec la sinistralité afin d'évaluer le niveau de mutualisation des risques.

Un autre élément à prendre en compte pour cette activité d'assurance est le respect de la loi Evin, qui s'applique à toute la protection sociale complémentaire.

2. *Loi n°89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques (dite 'Loi Evin')*

La loi Evin s'applique à l'assurance « *du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage* » (*art.* 1). Elle concerne par conséquent la prévoyance complémentaire des agents des collectivités territoriales et de leurs établissements associés.

Cette loi a pour but de protéger les assurés en imposant aux assureurs le respect de certaines règles.

Ainsi, les articles 2 et 3 de la loi précisent que pour les opérations collectives (et individuelles), l'assureur doit prendre en charge « *les suites des états pathologiques survenus antérieurement à la souscription du contrat ou de la convention ou à l'adhésion à ceux-ci, sous réserve des sanctions prévues en cas de fausse déclaration* ».

Néanmoins, pour les contrats entrant dans le cadre de l'article 3, une modulation est permise dans la mesure où l'assureur peut « *refuser de prendre en charge les suites d'une maladie contractée antérieurement à l'adhésion* […] » si deux conditions sont respectées :

« *a) Que la ou les maladies antérieures dont les suites ne sont pas prises en charge soient clairement mentionnées dans le contrat individuel ou dans le certificat d'adhésion au contrat collectif ;*
b) Que l'organisme apporte la preuve que la maladie était antérieure à la souscription du contrat ou à l'adhésion de l'intéressé au contrat collectif. »

NB : Il est bien fait mention, dans ces deux conditions, à une 'maladie' et non à un 'état pathologique'. La couverture de ce qui n'est pas considéré comme une maladie mais comme un état pathologique ne peut donc pas être exclue du contrat. L'exemple régulièrement cité pour clarifier cet aspect est celui de la séropositivité, qui est un état pathologique et ne peut donc pas faire l'objet d'une exclusion, tandis que le sida, qui lui est considéré comme une maladie, peut être exclu de la couverture par le contrat (Economie.gouv, 2012). Cependant, la jurisprudence tend à ne pas faire de différence entre 'état pathologique' et 'maladie' (Paris A., 2011).

S'il semble évident que les contrats labellisés, assimilables à des opérations individuelles, entrent dans le champ d'application de l'article 3, la réponse ne l'est pas autant pour les conventions de participation. En effet, la logique, en termes de mutualisation des risques et de recherche d'équilibre des contrats d'assurance, tendrait à rattacher les opérations collectives à adhésion obligatoire à l'article 2, et les opérations collectives à adhésion facultative à l'article 3. Cependant, la loi n'est pas rédigée aussi précisément et la question peut donc se poser. Des éléments de réponse 'officiels' sont apportés par une fiche pratique publiée sur le site du ministère de l'économie, qui mentionne l'article 2 quand il s'agit de contrats collectifs à adhésion obligatoire et l'article 3 quand il s'agit de contrats collectifs à adhésion facultative (Economie.gouv, 2012).

Dans la suite de ce travail, les conventions de participation pour les agents territoriaux seront considérées comme entrant dans le champ d'application de l'article 3. Il est donc considéré que les suites d'états pathologiques antérieurs doivent être prises en charge par la couverture d'une convention de participation, mais qu'il est possible d'exclure une ou des maladies dans certaines conditions.

L'autre article majeur de la loi Evin concernant les conventions de participation en prévoyance pour les agents territoriaux est l'article 7. Ce dernier mentionne que la résiliation ou le non-renouvellement de la convention est « *sans effet sur le versement des prestations immédiates ou différées, acquises ou nées durant son exécution* ». Ainsi, l'assureur doit indemniser toute maladie ayant son origine pendant la période de couverture de la convention, et constituer les provisions nécessaires pour couvrir ces engagements.

Les autres articles concernent plutôt les aspects pratiques de la convention.

Ainsi, l'article 12 stipule que le souscripteur de la convention, soit ici la collectivité, qui agit pour le compte de l'organisme assureur à l'égard de l'adhérent (*art.* 8), doit « *remettre à l'adhérent une notice d'information détaillée qui définit notamment les garanties prévues par la convention ou le contrat et leurs modalités d'application* ».

L'article 15 quant à lui précise que l'organisme assureur doit remettre chaque année à la collectivité un rapport sur les comptes de la convention, justifiant notamment le niveau des provisions et leur caractère prudent.

En **synthèse du cadre règlementaire** des conventions de participation à la prévoyance complémentaire des agents des collectivités territoriales, nous pouvons retenir que la **combinaison** de l'application du **décret** du 8 novembre 2011 et du respect de la **loi Evin** rend la **mutualisation des risques compliquée**. En effet, une adhésion facultative sans sélection médicale pendant une durée étendue si l'agent est en activité à la date d'effet du contrat, et l'obligation de prendre en charge les suites d'états pathologiques antérieurs s'ils ne sont pas mentionnés précisément au moment de l'adhésion, peuvent faciliter l'existence d'anti-sélection sur ce type de contrats.

De plus, le décret impose d'appliquer le même taux de cotisation à tous les agents, alors que les agents affiliés à l'IRCANTEC ont moins de droits que les agents affiliés à la CNRACL et donc moins de prestations en prévoyance complémentaire, ce qui peut entraîner un biais supplémentaire pour la mutualisation des risques.

En conséquence, si comme précisé dans l'introduction il n'est pas possible, moins de trois ans après la mise en place du décret, d'évaluer précisément le risque lié aux conventions de participation en prévoyance car il s'agit d'un risque long, il est essentiel de dégager les premières tendances concernant les taux d'adhésion aux contrats et les caractéristiques individuelles majeures des agents qui choisissent d'adhérer à ces contrats.

La suite de ce travail présentera dans un premier temps les données disponibles et utilisées dans les analyses ainsi que leur préparation, puis l'analyse des taux d'adhésion et des caractéristiques individuelles des agents qui choisissent d'adhérer, et enfin les premières tendances qui se dégagent concernant l'incidence de l'incapacité et le début de son maintien.

Partie II : Présentation et préparation des données

Deux grands types de données ont été utilisés :

- des données caractérisant les collectivités et l'ensemble des agents de ces collectivités ;

- des données concernant plus spécifiquement les agents du portefeuille de la mutuelle, *i.e.* ayant adhéré à une convention de participation.

Que ce soit au niveau de la collectivité ou au niveau de l'individu, les données utilisées ont été sélectionnées de façon à ce que les analyses puissent porter sur un maximum de collectivités et être répétées dans le futur avec l'évolution du portefeuille. La sélection a également tenu compte des informations disponibles dans les appels d'offres, de façon à pouvoir utiliser dans le futur les résultats de ces analyses pour affiner la méthode de tarification.

L'extraction des données pour ces analyses a eu lieu le 31 juillet 2014.

A/ Données concernant les collectivités du portefeuille et leurs agents

1. Données disponibles

Les données fournies par une collectivité lorsqu'elle publie un appel d'offre sont la plupart du temps assez succinctes. Il est extrêmement rare de disposer d'une liste (anonyme) des agents avec les informations tête par tête.

Les informations présentes systématiquement sont le type de la collectivité (commune,…) et le nombre total d'agents, soit le potentiel assurable. Dans la quasi-totalité des cas nous disposons également de la répartition entre hommes et femmes ainsi que d'une répartition par tranche d'âge, mais parfois ces deux informations sont fournies distinctement et il n'est pas possible de répartir chaque sexe par tranche d'âge.

La masse salariale brute est souvent renseignée, mais au global et sans mettre en parallèle le nombre d'ETP (Equivalents Temps Plein) ; or la proportion d'agents ne travaillant pas à temps plein est relativement élevée dans les collectivités territoriales puisqu'environ 13% des titulaires et 37% des non titulaires travaillent à temps non complet (CSFPT, 2011). Cette donnée n'est donc pas particulièrement exploitable pour les questions qui nous intéressent.

La répartition entre agents titulaires et non titulaires, et/ou par affiliation à la CNRACL ou l'IRCANTEC, et/ou par catégorie (A, B ou C) est parfois renseignée, mais trop peu souvent pour être utilisée dans le cadre de la présente étude.

Le montant de participation par agent n'est pas souvent renseigné au moment de la publication de l'appel d'offres mais il s'agit d'une information qui est dans la plupart des cas transmise au moment de la signature et de la mise en place de la convention de participation.

Le portefeuille étudié ici, qui est donc constitué de conventions de participation mises en place dans le cadre du décret du 8 novembre 2011, avec une date d'effet comprise entre la publication du décret et le 31 juillet 2014, contient **451 collectivités** pour un potentiel d'environ **56 000 agents**, soit plus de 10% du potentiel total des appels d'offres publiés depuis la mise en place du décret.

Deux conventions, l'une ayant débuté le 1er avril 2014 et l'autre le 1er juillet 2014, sont exclues du jeu de données car ne présentent pas assez de recul au moment de l'extraction pour étudier le taux d'adhésion.

2. Montant de participation

La variable la plus contraignante, car moins disponible, que nous voulons utiliser dans l'analyse du taux d'adhésion en fonction de caractéristiques de la collectivité, est son niveau de participation à la prévoyance complémentaire. Parmi les **449** collectivités conservées, le **niveau de participation** de la collectivité pour ses agents était disponible pour 355 d'entre elles (données recueillies auprès de la personne responsable des relations avec les collectivités). Il a été possible d'obtenir rapidement l'information pour 16 collectivités supplémentaires en les contactant directement par téléphone, ce qui porte à **371** le nombre de collectivités pour lesquelles nous connaissons la participation. Cette dernière est le plus souvent définie par un montant en euros par agent et par mois, mais peut présenter d'autres formes. Ainsi, certaines collectivités la déterminent en pourcentage du montant de la cotisation (pour la garantie incapacité la plupart du temps), d'autres allouent un montant différent en fonction de la catégorie de l'agent (A, B ou C) ou de son indice majoré[5]. Pour rendre les données comparables entre collectivités, nous avons fait en sorte d'obtenir un montant en euros par agent et par mois, qui soit unique pour une collectivité donnée. Lorsque le montant variait en fonction de certaines caractéristiques de l'agent (dans 21 collectivités, soit moins de 6% des cas), nous avons conservé le montant maximum alloué. Lorsque la participation était définie en pourcentage de la cotisation (dans 17 collectivités, soit moins de 5% des cas), nous avons calculé un montant moyen par agent et par mois à

[5] Multiplié par la valeur du point, cet indice permet de déterminer le traitement de base brut d'un agent.

l'aide des fichiers de précompte sur salaire envoyés par les collectivités, et mentionnant le montant de cotisation de chaque agent adhérent à la convention de participation. Le montant de participation ainsi obtenu varie de 0€ (certaines collectivités décident de mettre en place une convention de participation sans participer au financement de la prévoyance complémentaire de leurs agents) à 65€ par agent et par mois, avec une moyenne de 9€ et un écart-type de 8€. La figure 4 présente la fréquence cumulée croissante des montants de participation (arrondis à l'euro le plus proche) par agent et par mois dans une collectivité donnée, et permet d'observer que la médiane est à 7€ (*i.e.* 50% des collectivités versent une participation inférieure à 7€ par agent et par mois).

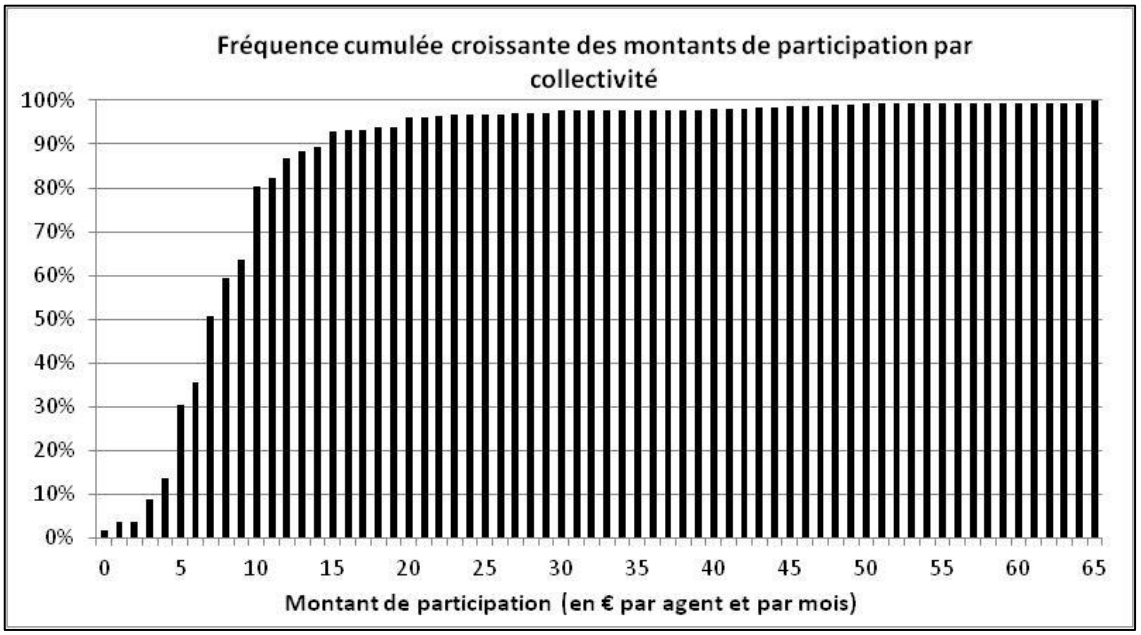


Figure 4 : Fréquence cumulée croissante, en nombre de collectivités, du montant de participation par agent et par mois

NB : Tester l'effet de la participation de la collectivité en prenant en compte un montant en euros par agent et par mois peut être biaisé par le fait que la cotisation s'exprime en pourcentage du traitement de l'agent et que la garantie minimale obligatoire n'est pas systématiquement l'incapacité seule. Ainsi, un même montant de participation peut être plus ou moins avantageux pour deux agents ayant des traitements différents et/ou entre deux collectivités n'ayant pas choisi le même niveau de garantie minimale. Pour limiter ce biais, la participation aurait pu être exprimée en pourcentage moyen de la garantie minimale obligatoire. Hors la collectivité ne fournit pas les informations nécessaires pour calculer ce pourcentage : pas de liste, même anonyme, des traitements par agent, mais une masse salariale globale brute que l'on ne peut pas rattacher à un nombre d'ETP ni même la plupart du temps à un nombre précis d'agents. Nous avons donc estimé que le biais serait plus limité en utilisant un montant en euro par agent qu'en posant plusieurs hypothèses et en estimant un traitement moyen par agent et par collectivité. Enfin, nous n'avons pas intégré le niveau minimal de garantie dans les analyses car dans la quasi-totalité des collectivités du portefeuille il s'agit de l'incapacité seule.

Deux autres variables qui nous intéressent pour l'étude du taux d'adhésion sont le type de collectivité et sa taille. Ces informations sont présentes pour l'ensemble des collectivités du portefeuille mais nous les avons retravaillées pour les rendre plus faciles à utiliser dans les analyses et l'interprétation des résultats.

3. Type de collectivité

Le portefeuille comportait à l'origine **14 types** de collectivités, répartis comme dans la table I ci-après.

Les communes sont bien entendu les plus représentées en nombre et constituent 62% des collectivités du portefeuille. Elles sont suivies par les centres communaux d'action sociale, puis les communautés de communes et syndicats (mixtes et intercommunaux).

Table I : Répartition des collectivités par type avant regroupement :

Type de collectivité	Nombre dans le portefeuille
Commune	231
Communauté de communes (Comcom)	26
Communauté d'agglomérations (Comagglo)	9
Conseil Général (CG)	7
Conseil Régional (CR)	1
Centre Communal d'Action Sociale (CCAS)	43
Centre Intercommunal d'Action Sociale (CIAS)	4
Etablissement d'Hébergement pour Personnes Agées Dépendantes (EHPAD)	16
Syndicat	25
Centre de Gestion (CDG)	2
Caisse des Ecoles	3
Parc ou Espace Naturel Régional	2
Service Départemental d'Incendie et de Secours (SDIS)	1
Centre Universitaire Départemental	1

Pour limiter le nombre de classes et augmenter le nombre de collectivités par classe, nous avons regroupé les types similaires en une seule classe. Les regroupements sont les suivants (table II) :

- communautés de communes et communautés d'agglomérations (« ccom-comagglo ») ;

- conseils généraux et régionaux (« dpt-region ») ;

- CCAS, CIAS et EHPAD (« ccas-ehpad ») ;

- une catégorie « autre » pour les collectivités trop peu représentées et ne pouvant se regrouper avec un autre type. Cette catégorie comprend les types 'CDG', 'caisse des écoles', 'parc ou espace naturel régional', 'SDIS' et 'centre universitaire départemental'.

Table II : Répartition des collectivités par type après regroupement

Type de collectivité	Nombre de collectivités dans le portefeuille
Commune	231
Ccom-comagglo	35
Dpt-région	8
Ccas-ehpad	63
Syndicat	25
Autre	9

4. Taille de la collectivité

Afin d'interpréter et surtout d'utiliser plus facilement les résultats des analyses, nous avons souhaité discrétiser le potentiel de la collectivité. Pour créer des classes de taille, nous avons tenu compte de trois éléments : le nombre total de classes, le nombre de collectivités par classe et le nombre d'agents par classe. La figure 5 montre que ce portefeuille est constitué en majorité de collectivités de petite taille mais que la majorité des agents sont dans les collectivités de grande taille. En effet, 50% des collectivités ont un potentiel inférieur à 20 agents et représentent moins de 5% des agents, tandis que 50% des agents sont répartis sur 10 collectivités, soit moins de 3% des collectivités, d'un potentiel supérieur à 1 500 agents.

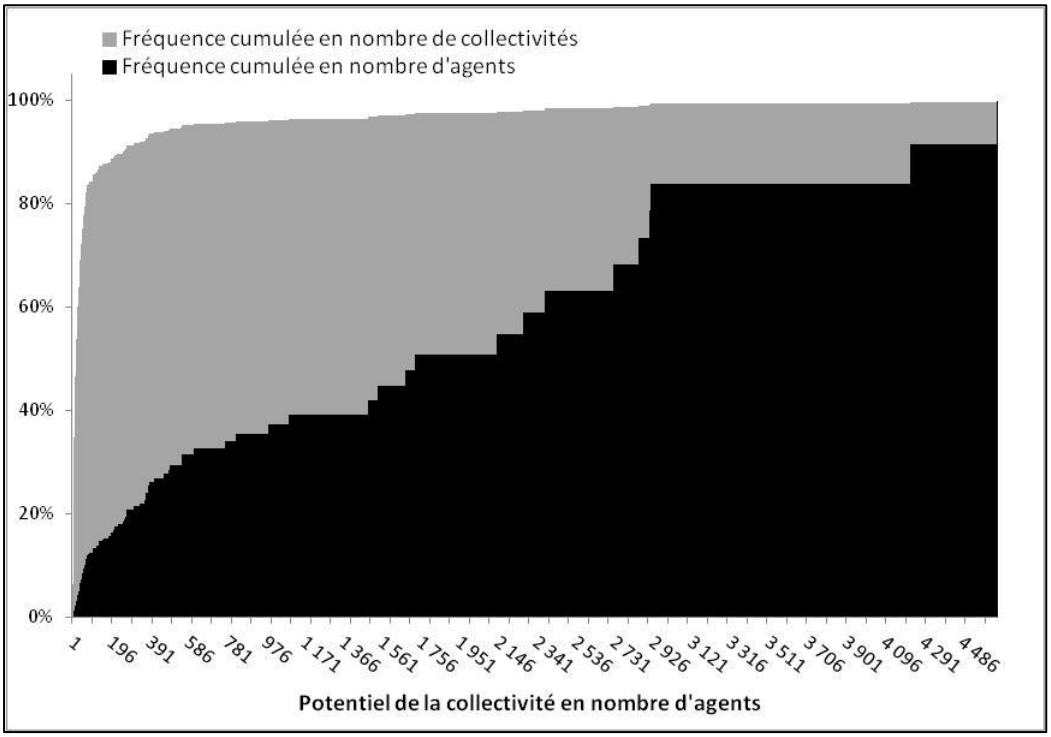


Figure 5 : Fréquences cumulées croissantes du potentiel de la collectivité en nombre de collectivités et en nombre d'agents

NB : Attention à ne pas confondre le *potentiel* de la collectivité, qui est le nombre d'agents pour une collectivité donnée, et le *nombre d'agents*, qui peut être la somme de plusieurs potentiels s'il fait référence à plusieurs collectivités. Par la suite, il faudra également bien distinguer le nombre d'agents, qui correspond au nombre total d'agents dans une ou plusieurs collectivités, du *nombre d'adhérents*, qui fera référence uniquement aux agents ayant adhéré à une convention de participation.

Nous avons défini sept classes de taille :

- A-1-10 : collectivités de 1 à 10 agents ;

- B-11-30 : collectivités de 11 à 30 agents ;

- C-31-50 : collectivités de 31 à 50 agents ;

- D-51-100 : collectivités de 51 à 100 agents ;

- E-101-500 : collectivités de 101 à 500 agents ;

- F-501-2000 : collectivités de 501 à 2000 agents ;

- G->2000 : collectivités de plus de 2 000 agents.

La table III ci-dessous donne la répartition des collectivités du portefeuille par type et par taille.

Table III : Nombre de collectivités par type et par taille

Nombre de collectivités	Potentiel de la collectivité (en nombre d'agents)							**Total**
	A-1-10	B-11-30	C-31-50	D-51-100	E-101-500	F-501-2000	G->2000	
ccas-ehpad	9	11	20	13	10			63
ccom-comagglo	2	6	10	5	8	4		35
commune	101	75	17	11	19	5	3	231
dpt-region						2	6	8
syndicat	16	1	3	5				25
autre	1	3	1	3	1			9
Total	129	96	51	37	38	11	9	**371**

Deux éléments apparaissant dans cette table seront à prendre en compte lors des analyses : moins de la moitié des types de collectivités est représentée dans les classes de taille supérieure à 500 agents, et les 'dpt-région' ne sont présents que dans les classes de taille supérieure à 500 agents. Ainsi par exemple, il ne sera pas envisageable de mettre en évidence un lien entre le taux d'adhésion et la taille de la collectivité en fonction du type de

collectivité, car dans certains cas l'« effet taille » ne pourra pas être distingué de l'« effet 'type' ».

La répartition des agents par type et taille de collectivité est présentée dans la table IV.

Table IV : Répartition du nombre d'agents en fonction du type et de la taille de la collectivité

Nombre d'agents	Potentiel de la collectivité (en nombre d'agents)							Total
	A-1-10	B-11-30	C-31-50	D-51-100	E-101-500	F-501-2000	G->2000	
ccas-ehpad	50	204	810	811	2 731			4 606
ccom-comagglo	16	101	396	300	1 862	2 699		5 374
commune	638	1 354	639	747	4 430	5 907	7 318	21 033
dpt-region						3 218	19 765	22 983
syndicat	74	18	109	318				519
autre	2	56	36	199	366			659
Total	780	1 733	1 990	2 375	9 389	11 824	27 083	**55 174**

Dans le portefeuille étudié, la classe la plus représentée en nombre d'agents pouvant adhérer à une convention de participation est la classe 'dpt-région', même si elle ne contient que 8 collectivités, réparties entre seulement deux classes de taille. A elles deux, cette classe et la classe 'commune' représentent presque 80% du potentiel total du portefeuille, mais la classe 'commune' est correctement représentée dans toutes les classes de taille.

Les autres types de collectivités sont assez bien représentés dans les différentes classes de taille, sauf pour les deux classes de plus de 500 agents, ce qui confirme que dans les analyses il ne faudra pas étudier l'interaction potentielle entre le type et la taille de la collectivité.

Nous avons donc à disposition trois variables à mettre en relation avec le taux d'adhésion : le type de collectivité, sa taille et son niveau de participation. Cependant, suite aux échanges concernant la mise en place des conventions avec la responsable des relations avec les collectivités, aux échanges réguliers avec la gestion et aux premières analyses de sinistralité, il est apparu important d'intégrer une variable supplémentaire relative à la situation sociale du territoire où se trouve la collectivité. En effet, dans certaines collectivités la prévoyance complémentaire, qui permet le maintien du traitement pour des arrêts longs (puisqu'elle n'intervient qu'à partir de trois mois d'arrêt) n'est pas la préoccupation principale des agents et les avantages que peut leur procurer ce type de couverture ne sont pas toujours connus. A cela s'ajoute l'aspect financier, la prévoyance complémentaire n'étant même pas envisagée quand la question de pouvoir subvenir aux besoins de la famille se pose.

5. *Données complémentaires*

Nous avons cherché une variable proxy[6] en consultant le site de l'INSEE (Institut National de la Statistique et des Etudes Economiques). Le taux de chômage aurait pu être un bon candidat, mais cet indicateur est souvent remis en question, et de plus cela ne semblait pas vraiment pertinent d'utiliser cette variable pour des analyses en lien avec des personnes ayant une activité professionnelle. L'utilisation des zones urbaines sensibles a également été envisagée, mais cela concerne justement les zones urbaines, et donc pas l'ensemble des possibles en termes de territoire. Enfin, la variable retenue a été la proportion d'HLM (Habitation à Loyer Modéré) parmi les habitations principales. Un fichier Excel de l'INSEE listait toutes les communes françaises avec pour chacune le nombre d'habitations principales et le nombre d'habitations principales étant des HLM en 2011. Nous avons calculé le ratio entre ces deux variables, et intégré cette proportion dans notre fichier de données grâce au code postal de la commune à laquelle est rattachée une collectivité. En moyenne dans les communes liées aux collectivités du portefeuille, 8% des habitations principales sont des HLM, avec un écart-type de 6%, cette proportion s'étendant de 0% à 36,7%. La figure 6 ci-dessous permet d'observer que 50% des collectivités ont une proportion d'HLM inférieure à 7%.

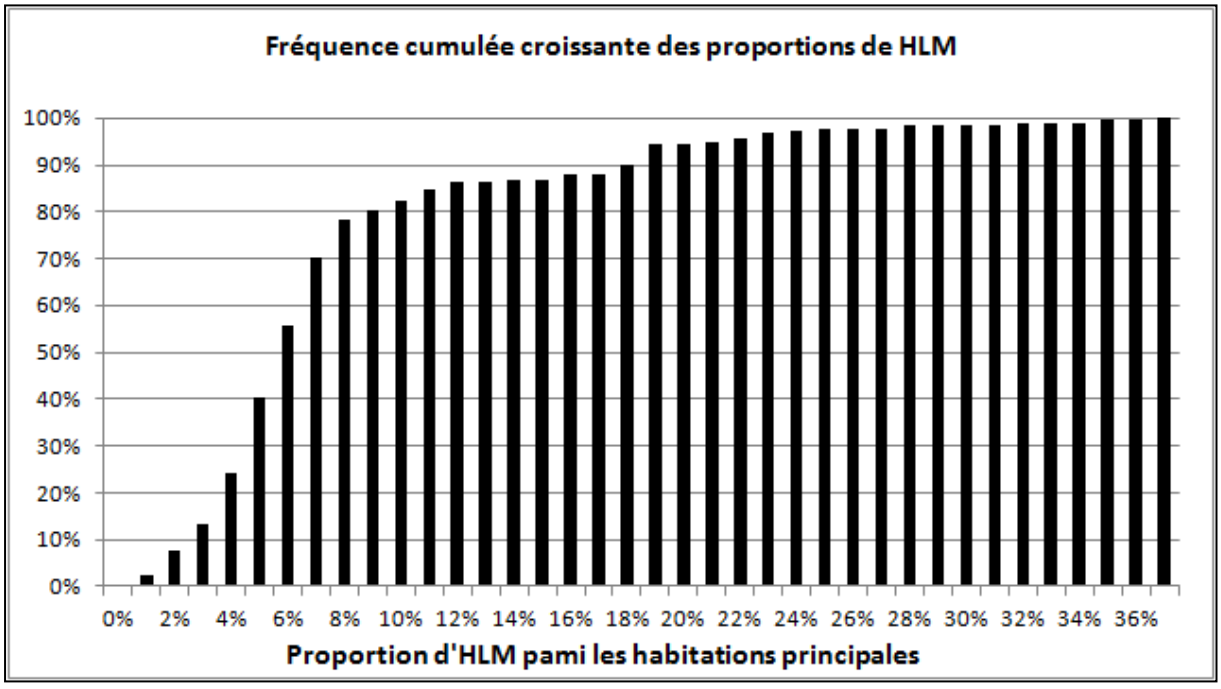


Figure 6 : Fréquence cumulée croissante de la proportion d'HLM parmi les habitations principales

[6] Une variable proxy est une variable mesurable qui peut se substituer à une variable d'intérêt non mesurable.

Ainsi, pour les **analyses** concernant le **taux d'adhésion**, nous disposons de **371 collectivités pour un total de 55 174 agents** avec les informations suivantes par collectivité :

- le **type,** constitué de 6 classes, qui est une variable qualitative ;

- le **potentiel** de la collectivité (=nombre d'agents de la collectivité pouvant adhérer à la convention de participation), constitué de 7 classes, qui est une variable qualitative ordinale ;

- le montant en euros de la **participation** par agent et par mois, qui est une variable quantitative ;

- la **proportion d'HLM** parmi les habitations principales pour la commune à laquelle est rattachée la collectivité, qui est une variable quantitative.

6. Démographie des collectivités

Pour les analyses concernant les caractéristiques des agents qui choisissent d'adhérer à une convention de participation, il nous faut, en plus des informations précédentes, des données concernant la structure démographique de la collectivité, soit une répartition par sexe et par âge des agents de la collectivité.

Pour constituer cette partie de la base de données, nous avons repris les informations contenues dans chacun des appels d'offres publiés par les collectivités du portefeuille. L'examen des données disponibles a conditionné le nombre et les bornes des classes d'âges constituées, l'objectif étant de maximiser la précision tout en conservant un maximum de collectivités dans le jeu de données. Nous avons constitué 7 classes d'âge :

- moins de 35 ans ('<35') ;
- de 35 à 39 ans ('35-39') ;
- de 40 à 44 ans ('40-44') ;
- de 45 à 49 ans ('45-49') ;
- de 50 à 54 ans ('50-54') ;
- de 55 à 59 ans ('55-59') ;
- 60 ans et plus ('60+').

Les données démographiques étaient disponibles pour 329 collectivités, et nous avons pu obtenir les informations pour 7 collectivités supplémentaires en les contactant directement.

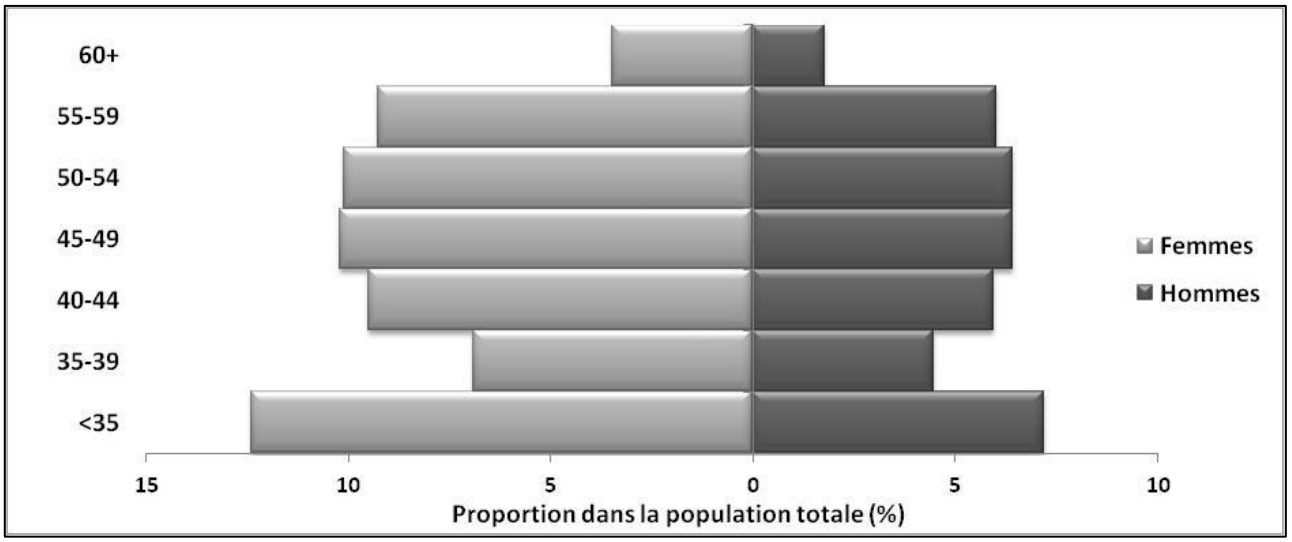


Figure 7 : Pyramide des âges de l'ensemble des agents des collectivités du portefeuille

Nous avons ainsi pu répartir par sexe et par classe d'âge les agents de **336 collectivités** (soit 75% des collectivités du portefeuille) regroupant **54 041 agents** (soit 96% des agents du portefeuille). La figure 7 présente la pyramide des âges correspondante.

La classe des moins de 35 ans est la plus représentée, mais elle couvre environ 15 années tandis que les autres classes n'en couvrent que 5.

L'information concernant d'autres caractéristiques individuelles n'est pas disponible au niveau de la collectivité. Nous concentrerons donc notre analyse des choix individuels d'adhésion sur le sexe et l'âge des agents.

B/ Données concernant les adhérents

Les données utilisées pour cette partie sont issues de la base des adhérents et de la base des sinistres.

Le fichier constitué possède une ligne par adhérent, avec pour chaque adhérent les informations suivantes :

- le sexe ;
- la date de naissance ;
- la collectivité de rattachement ;
- le numéro de police (un numéro unique par convention de participation et par collectivité) ;
- la date d'effet de la convention de participation ;
- la date d'adhésion de l'agent à la convention de participation ;

- la date de résiliation de l'agent, le cas échéant ;
- la date de fin de la période d'observation, qui est le 31/07/2014 ;
- la date de début de première indemnisation, qui correspond à la date du premier passage à demi-traitement depuis l'adhésion de l'agent.

Le fichier comprend au total 33 021 adhérents répartis sur 451 collectivités.

Plusieurs contrôles de cohérence ont été effectués lors de la construction du fichier, notamment par exemple :

- vérification de l'absence de dates de naissance aberrantes en calculant l'âge à la date d'adhésion ; pour les âges inférieurs à 16 ans et supérieurs à 67 ans, une vérification et correction éventuelle a été demandée au service gestion ;

- vérification de la cohérence des années dans toutes les dates, les dates d'effet de convention, d'adhésion, de résiliation, de début de première indemnisation, de fin de première indemnisation devant se situer entre le 01/01/2012 et le 31/07/2014 ;

- recherche de l'information concernant le sexe si absente et vérification de la cohérence du sexe avec le premier chiffre du numéro de sécurité sociale ;

- vérification que la date d'adhésion était bien identique ou postérieure à la date d'effet de la convention de participation et antérieure à la date de résiliation le cas échéant ;

- vérification que la date de début de première indemnisation était bien postérieure à la date d'adhésion et antérieure à la date de fin de première indemnisation

- vérification que la collectivité de rattachement pour un numéro de police correspondait bien à la même collectivité que celle ayant le même numéro de police dans le fichier de données des collectivités.

Ces vérifications et corrections nous ont permis de disposer de données suffisamment fiables et complètes pour nos analyses.

1. Population assurée

Dans cette partie sont présentées les données utilisées pour l'analyse du taux d'adhésion et des caractéristiques individuelles liées au choix d'adhérer ou non à la convention de participation.

Pour l'analyse du taux d'adhésion, l'information du portefeuille d'individus nécessaire est le nombre d'adhérents par collectivité de rattachement, qui sera relié au nombre d'agents dans la collectivité, pour tester s'il existe une relation entre le taux d'adhésion et les caractéristiques de la collectivité. A partir du fichier constitué, il suffit de faire la somme des adhérents par numéro de police, puis de la rattacher au fichier contenant les caractéristiques des collectivités, lui aussi contenant l'information du numéro de police, qui sert d'identifiant unique à une collectivité pour une convention de participation.

Pour l'ensemble des agents des collectivités, nous disposons de la répartition par sexe et par tranche d'âge de 5 ans (sauf les classes <35 et 60+ qui présentent des étendues supérieures à 5 ans). Pour déterminer quelles sont les caractéristiques individuelles des agents qui choisissent d'adhérer aux conventions de participation, il nous faut donc répartir les adhérents par sexe et selon les mêmes classes d'âge. Le sexe est renseigné pour chaque adhérent, ainsi que sa date de naissance. Nous pouvons donc calculer l'âge des adhérents à la date de leur adhésion et les répartir selon les mêmes classes d'âges que l'ensemble des agents de la collectivité correspondante.

NB : Toutes les analyses intégrant l'âge des adhérents concerneront l'**âge à l'adhésion**. La **variable âge** sera donc considérée **fixe dans le temps**. En effet, concernant les choix d'adhésion, ce qui nous intéresse est l'âge de l'agent au moment où il décide d'adhérer à la convention de participation. De plus, l'âge des adhérents au moment de leur adhésion est proche de leur âge au moment où la collectivité transmet les données car il y a toujours moins d'un an entre la publication de l'appel d'offres et la date de début de convention. Peu d'agents changent donc de classe d'âge entre la publication des données et leur adhésion. Concernant l'incidence de l'incapacité indemnisée, ce qui nous intéresse est de déterminer si le délai entre l'adhésion de l'agent et son premier arrêt indemnisé au titre de la prévoyance complémentaire a un lien avec l'âge auquel il a choisi d'adhérer. Enfin, la durée maximale d'observation n'excédant pour le moment pas deux ans et demi, l'âge en fin de période d'observation reste proche de l'âge à l'adhésion.

La répartition des adhérents par sexe et selon les mêmes classes d'âge que les agents des collectivités permet de construire les pyramides des âges selon le même modèle que pour l'ensemble des agents des collectivités ayant fourni des données démographiques.

Que l'on prenne en compte les adhérents des 451 collectivités (fig. 8a), ou juste les adhérents des collectivités pour lesquelles la structure démographique était disponible (336 collectivités, fig. 8b), leur répartition par âge et sexe est similaire.

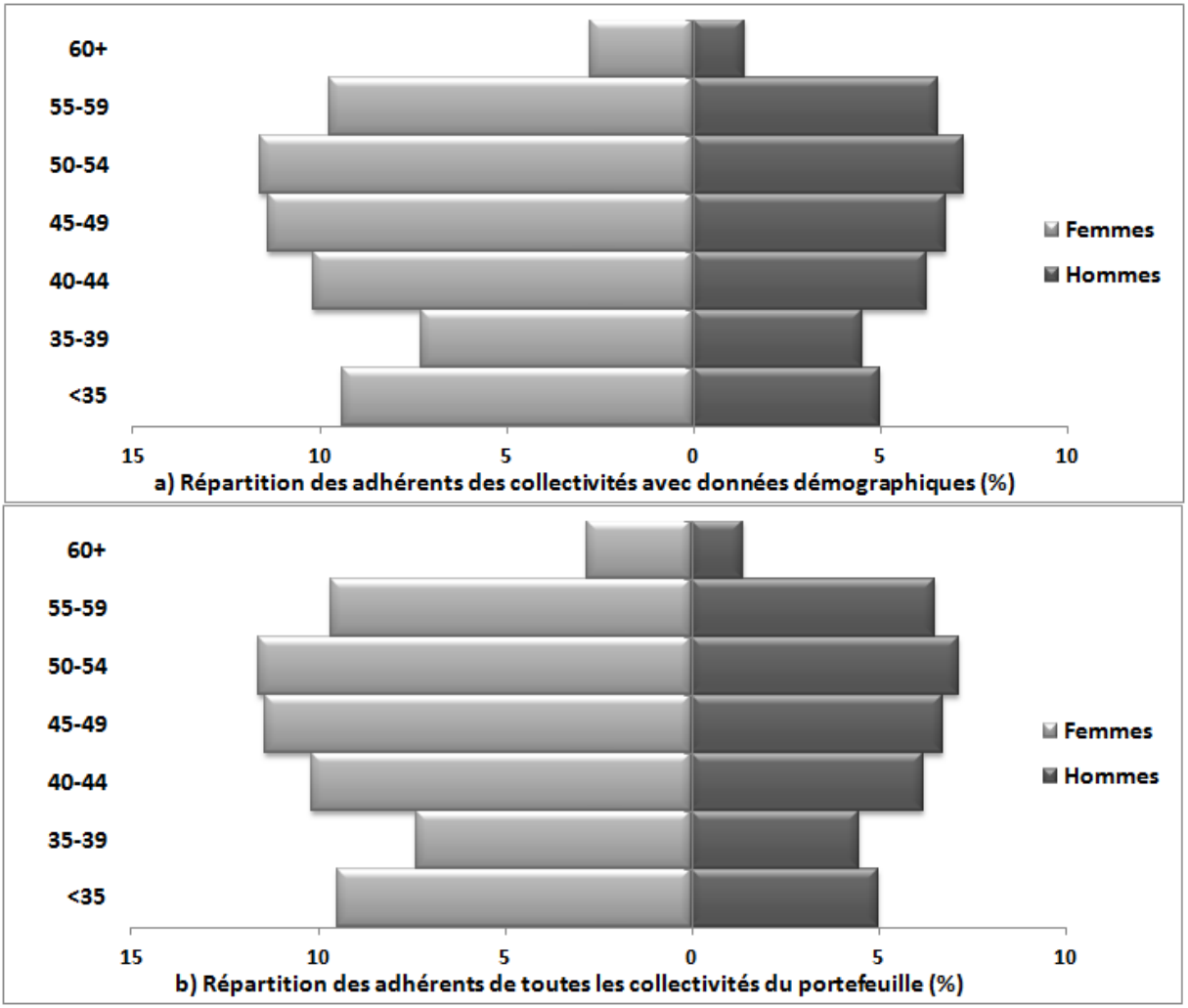


Figure 8 : Pyramides des âges des adhérents ; a) pour l'ensemble des collectivités ayant fourni des données démographiques pour leur agents et b) pour l'ensemble des collectivités ayant signé une convention de participation.

La première observation que nous pouvons faire si nous comparons ces pyramides des âges à celle de la figure 7 p. 34 est que les adhérents semblent plus âgés que les agents des collectivités mais que la répartition entre hommes et femmes ne semble pas différente. Il y aurait donc potentiellement un lien entre le choix d'adhésion et l'âge de l'agent, mais pas avec son sexe. Ces questions seront testées statistiquement dans la partie III.

La table V donne le détail de la répartition des adhérents par sexe et classe d'âge à l'adhésion. L'âge moyen d'adhésion est de 45.6 ans (avec un écart-type de 9.4 ans), soit un peu plus que l'âge moyen des agents de la fonction publique territoriale qui est de 44 ans. Le portefeuille est constitué de 62.5% de femmes, dont l'âge moyen d'adhésion est de 45.5 ans, celui des hommes étant assez proche puisqu'il est de 45.8 ans. Ces observations sont concordantes avec l'observation graphique des pyramides des âges.

Table V : Répartition des adhérents par sexe et classe d'âge à l'adhésion

	Age à l'adhésion							TOTAL
	< 35 ans	35-39 ans	40-44 ans	45-49 ans	50-54 ans	55-59 ans	60 ans et +	
Femmes	3 137	2 429	3 367	3 768	3 832	3 194	927	20 654
Hommes	1 655	1 482	2 050	2 222	2 359	2 149	450	12 367
TOTAL	4 792	3 911	5 417	5 990	6 191	5 343	1 377	33 021

Pour l'étude du moment auquel les agents choisissent d'adhérer à partir de la mise en place de la convention de participation, nous définissons trois périodes d'adhésion :

- à la **date d'effet** de la convention de participation ;

- dans les **six premiers mois** suivant la date d'effet de la convention ;

- **au-delà** des six premiers mois suivant la date d'effet de la convention.

Ces trois classes ont été définies d'après les critères suivants :

- au-delà des six premiers mois d'effet de la convention de participation, l'adhésion peut être soumise à un questionnaire médical et peut être refusée ;

- les adhésions à la date d'effet de la convention résultent des campagnes de mise en place, qui ont lieu dans les mois précédant le début de la convention ;

- les adhésions au cours des six premiers mois de la convention peuvent être par exemple le résultat de campagnes d'adhésion complémentaires, ou le choix d'agents qui étaient en arrêt de travail et ont attendu de reprendre leur activité pour adhérer, ou qui simplement ont hésité à adhérer. Dans tous les cas, il est intéressant de distinguer ces agents de ceux qui choisissent d'adhérer dès le début de la convention.

NB : Le décret précise que les agents qui ne sont pas en arrêt à la date d'effet de la convention ont 6 mois pour adhérer sans questionnaire médical, mais il est compliqué de demander à la collectivité la liste des agents en arrêt de travail à la date d'effet de la convention, et surtout, un agent qui n'est pas en arrêt à la date d'effet mais qui n'a pas encore demandé à adhérer peut ensuite être en arrêt de travail plusieurs mois et faire sa demande d'adhésion sans être soumis à un questionnaire dès le jour de sa reprise d'activité si elle a lieu dans les 6 premiers mois de la convention.

La figure 9 présente la répartition des adhésions selon les trois périodes définies précédemment. Environ trois quarts des agents qui choisissent d'adhérer le font à la date d'effet de la convention, ce qui indique que les campagnes de mise en place des conventions sont efficaces.

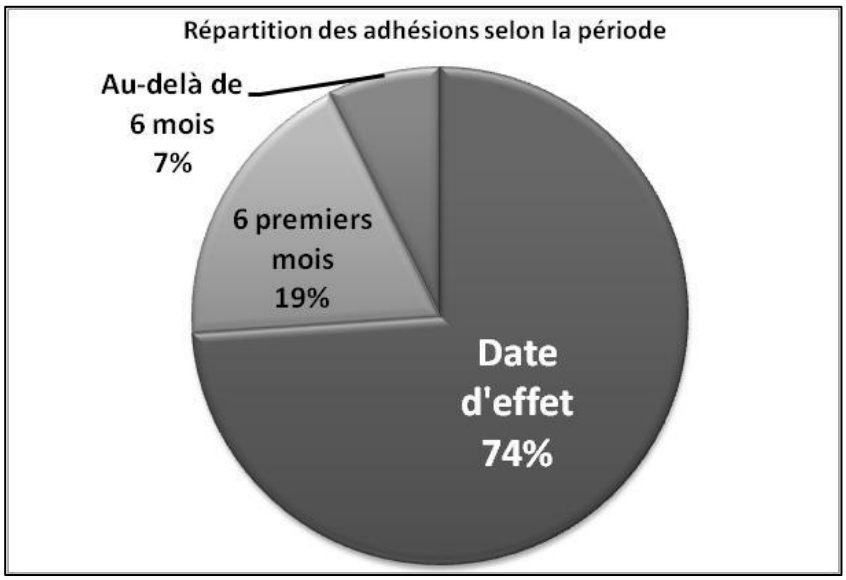


Figure 9 : Répartition des adhésions en fonction de la date d'adhésion

Le délai moyen d'adhésion est de 45 jours, soit environ un mois et demi, avec un écart-type de 109 jours, soit plus de trois mois, ce qui suggère que les agents adhérant au-delà des six premiers mois de la convention le font longtemps après. En effet, le délai d'adhésion moyen des agents qui choisissent d'adhérer au-delà des six premiers mois est de 370 jours, soit plus d'un an, avec un écart-type de 150 jours. Concernant les agents qui adhèrent au cours des six premiers mois de la convention, le délai moyen d'adhésion est de 94 jours, soit environ trois mois, avec un écart-type de 56 jours.

Si l'on s'intéresse à la répartition en fonction du sexe, il semblerait que les hommes adhèrent plus tôt que les femmes car 76% d'entre eux adhèrent à la date d'effet de la convention contre seulement 73% des femmes (fig. 10).

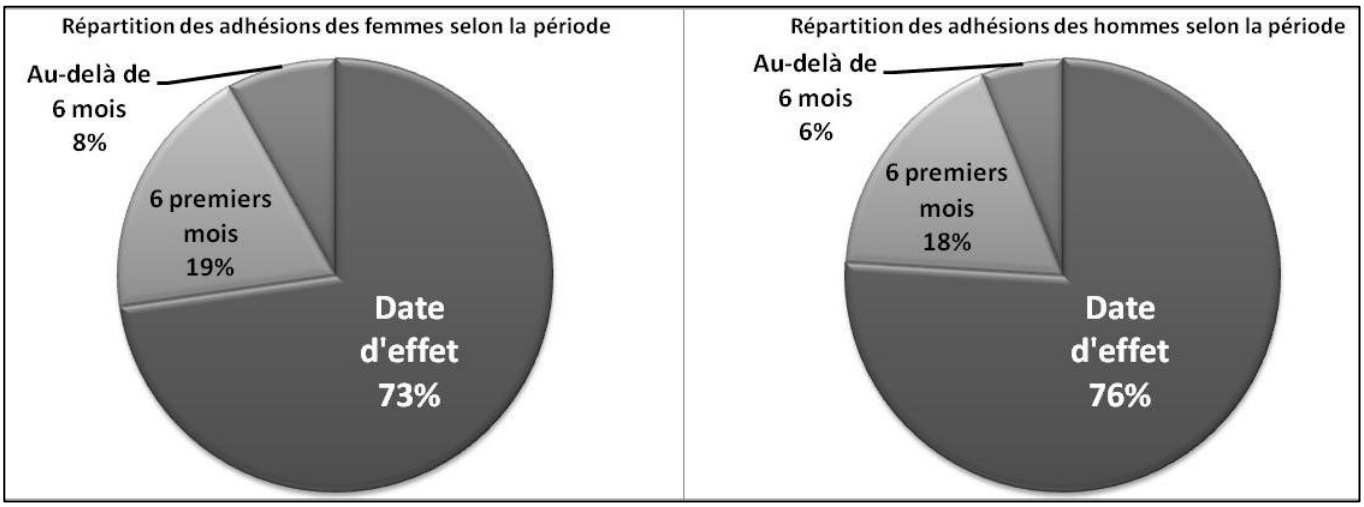


Figure 10 : Répartition des adhésions, par sexe, selon la date d'adhésion

Les relations entre la date d'adhésion et les caractéristiques individuelles telles que le sexe et l'âge seront testées dans la partie III. L'existence d'une possible relation entre l'incidence de l'incapacité indemnisée et la date d'adhésion sera abordée dans la partie IV.

2. Sinistralité

Dans cette partie sont présentées les données utilisées pour l'analyse de l'incidence de l'incapacité indemnisée et du début du maintien en incapacité indemnisée.

Pour analyser l'incidence de l'incapacité indemnisée, nous utiliserons une approche qui estime la durée entre la date d'adhésion et le premier jour en incapacité indemnisée en fonction du sexe, de l'âge à l'adhésion et de la date d'adhésion.

En plus des données décrites précédemment, nous avons donc besoin, par adhérent, du nombre de jours entre l'adhésion et le premier jour indemnisé le cas échéant.

Tous les adhérents n'ayant pas été observés (=exposés) sur des durées équivalentes, nous devrons tenir compte de la durée d'observation de chacun.

A partir des données de base, d'autres variables sont ainsi construites pour les analyses :

- la durée d'observation de l'agent, qui est le nombre de jours pendant lesquels l'agent a été observé et correspond à la différence entre la date de fin d'observation et la date de début d'observation de l'agent. La date de début d'observation est la date d'adhésion à la convention de participation et la date de fin d'observation est soit la date de fin de la période d'observation (*i.e.* 31/07/2014), soit la date de résiliation de l'agent le cas échéant ;
- la durée sans sinistre, qui est le nombre de jours entre la date d'adhésion et la date du premier jour de congé maladie indemnisé ou la date de fin d'observation de

l'adhérent (résiliation ou 31/07/2014) s'il n'a eu aucun jour de congé maladie indemnisé ;

- une variable codée 0 ou 1, qui est renseignée à 0 quand l'agent n'a eu aucun jour de congé maladie indemnisé pendant sa période d'observation, et 1 quand l'agent a eu au moins un jour de congé maladie indemnisé.

Pour les 33 021 adhérents du portefeuille, la durée d'observation moyenne est de 446 jours, soit un an et presque trois mois, variant d'un mois (27 jours) à 34 mois.

2 025 adhérents ont eu au moins un jour de congé maladie indemnisé pendant leur période d'observation, soit 6,1% d'entre eux.

Pour les adhérents ayant eu au moins un jour de congé maladie indemnisé, le délai moyen depuis leur date d'adhésion est de 272 jours, soit un peu plus de 9 mois (l'écart-type étant de 192 jours), la moitié d'entre eux étant en congé maladie indemnisé dans les 7 premiers mois suivant leur adhésion (fig. 11).

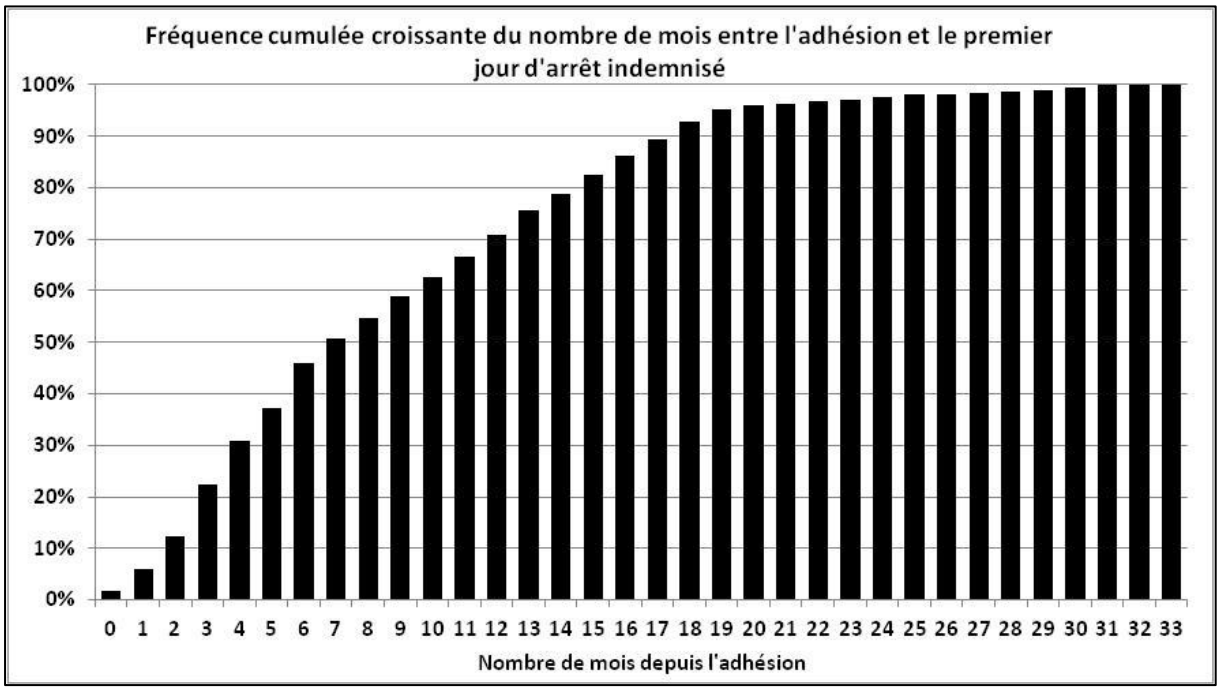


Figure 11 : Fréquence cumulée croissante du nombre de mois entre l'adhésion et le premier jour d'arrêt en congé maladie indemnisé pour les adhérents ayant eu au moins un jour de congé maladie indemnisé au cours de la période d'observation

Pour l'étude du début du maintien en incapacité indemnisée, nous calculons le nombre de jours en congé maladie indemnisé du premier congé maladie indemnisé au cours de la période d'observation. Ce nombre de jours correspond à la différence entre la date de fin de congé maladie indemnisé et sa date de début ou, si l'adhérent n'a pas repris le travail, la différence entre la date de fin de la période d'observation (soit le 31/07/2014) et la date de début de congé maladie indemnisé.

NB : Pour cette analyse, si l'agent a résilié son adhésion mais est toujours en arrêt de travail, nous ne prenons pas en compte la date de résiliation comme date de fin car l'article 7 de la loi Evin impose de continuer à indemniser l'adhérent.

Parmi ces 2 025 congés maladie indemnisés, 503 sont encore en cours à la fin de la période d'observation et 1 522 ne le sont plus. Ainsi, plus de 75% des observations ne sont pas censurées. Concernant ces 1 522 congés non censurés (*i.e.* ayant une date de fin avant la fin de la période d'observation), la durée moyenne est de 97 jours, soit un peu plus de 3 mois, avec un écart-type de 127 jours, soit plus de 4 mois. Ces congés durent entre 1 jour et plus de 2,5 ans.

L'observation de la figure 12 indique que la moitié des congés maladie indemnisés et terminés (*i.e.* non censurés) avant la fin de la période d'observation (durée totale 2 ans et 7 mois) durent moins de 2 mois.

NB : Cette médiane augmentera avec l'allongement de la période d'observation puisque l'incapacité peut durer jusqu'à plus de 5 ans.

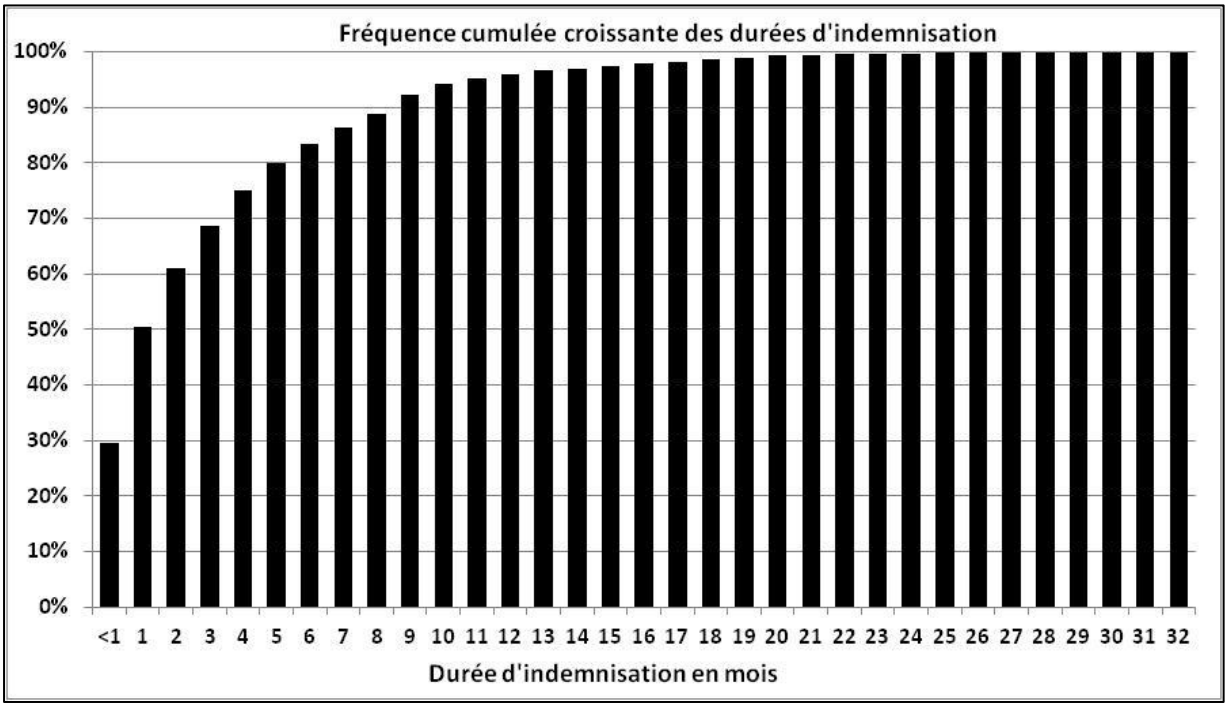


Figure 12 : Fréquence cumulée croissante de la durée du premier congé maladie indemnisé après l'adhésion pour les adhérents ayant repris le travail avant la fin de la période d'observation.

Une analyse plus précise de ces informations sera présentée dans la partie IV.

Partie III : Etude de l'adhésion

Comme précisé précédemment, le premier élément à étudier pour déterminer quel pourrait être l'effet du caractère facultatif sans sélection médicale (des conventions de participation mises en place dans le cadre du décret du 08/11/2011) est le taux d'adhésion, à mettre en relation avec différentes caractéristiques des collectivités. En effet, plus le taux d'adhésion est élevé, plus les risques peuvent être mutualisés.

Ensuite, dans les cas où moins de la totalité des agents adhèrent à la convention de participation, il est important de déterminer quels sont les agents qui choisissent d'adhérer ou non en fonction de leur âge et de leur sexe et quand ils choisissent de le faire le cas échéant, ainsi que mettre en évidence les caractéristiques des collectivités pour lesquelles la structure démographique de la population adhérente s'éloigne de celle de la population initiale.

A/ Taux d'adhésion

1. Analyse descriptive

Le taux d'adhésion est le rapport entre le nombre d'adhérents et le nombre d'agents à périmètre équivalent. Il s'agit donc d'une proportion, bornée entre 0 et 1.

Parmi les 55 174 agents des 371 collectivités conservées pour l'analyse du taux d'adhésion, 31 291 ont adhéré, ce qui donne un taux d'adhésion global de 56,7%.

Si l'on s'intéresse à la moyenne du taux d'adhésion par collectivité, on obtient 80,3%, avec un écart-type de 21,5%. Cette moyenne ne tient pas compte de la taille de la collectivité, tandis que le premier taux obtenu correspond à une moyenne pondérée par le potentiel de la collectivité. Ce premier résultat laisse supposer qu'il existe un lien entre la taille de la collectivité et le taux d'adhésion, ce dernier étant certainement plus faible dans les grandes collectivités puisque la moyenne non pondérée donne un taux d'adhésion plus élevé que la moyenne pondérée.

En effet, si nous calculons le taux d'adhésion par taille de collectivité (table VI) nous constatons qu'il est plus bas dans les grandes collectivités. La moyenne du taux d'adhésion par collectivité en fonction de sa taille (i.e. la classe de potentiel à laquelle elle appartient) reste à moins de 1% d'écart du taux global, sauf pour les classes A, E et F.

Table VI : Nombre d'agents, nombres d'adhérents et taux d'adhésion par taille de collectivité

Classe de Potentiel	Nombre d'agents	Nombre d'adhérents	Taux d'adhésion global	Moyenne du taux d'adhésion par collectivité	Ecart-type du taux d'adhésion par collectivité
A-1-10	780	627	80.4%	81.9%	22.8%
B-11-30	1 733	1 447	83.5%	82.8%	19.1%
C-31-50	1 990	1 715	86.2%	86.5%	17.6%
D-51-100	2 375	1 864	78.5%	78.4%	20.5%
E-101-500	9 389	6 797	72.4%	75.1%	17.5%
F-501-2000	11 824	7 164	60.6%	65.3%	22.4%
G->2000	27 083	11 677	43.1%	43.3%	20.8%

De plus, si l'on s'intéresse à l'écart-type (fig. 13), il ne semble pas évident que les différences de taux d'adhésion entre les différentes tailles de collectivité sont significatives. Ces deux derniers points indiquent que d'autres caractéristiques de la collectivité que sa taille doivent être prises compte pour étudier la variabilité du taux d'adhésion.

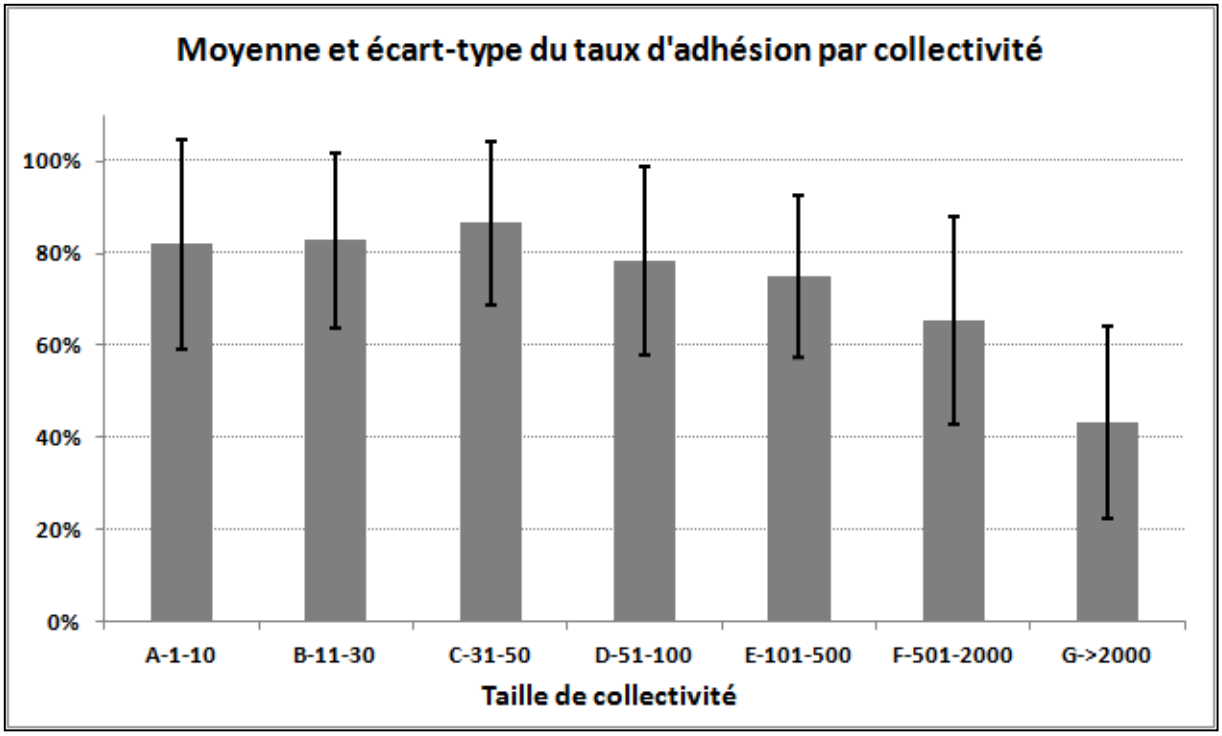


Figure 13 : Moyenne (+/- écart-type) du taux d'adhésion par collectivité en fonction de la taille de la collectivité

Si l'on procède de la même façon avec les taux d'adhésion par type de collectivité, on aboutit aux mêmes conclusions que pour la taille de la collectivité : le type de collectivité ne suffit pas à lui seul à expliquer les variations de taux d'adhésion (table VI et fig.14).

L'écart important, pour chaque type de collectivité (hormis dpt-région et syndicat), entre le taux d'adhésion global et la moyenne du taux d'adhésion par collectivité (table VI), confirme que la taille de la collectivité devra impérativement être prise en compte dans les analyses.

Table VI : Nombre d'agents, nombres d'adhérents et taux d'adhésion par type de collectivité

Type de collectivité	Nombre d'agents	Nombre d'adhérents	Taux d'adhésion global	Moyenne du taux d'adhésion par collectivité	Ecart-type du taux d'adhésion par collectivité
commune	21 033	13 637	64.8%	82.8%	19.4%
ccom-comagglo	5 374	4 279	79.6%	87.7%	13.0%
ccas-ehpad	4 606	3 311	71.9%	75.2%	21.7%
dpt-region	22 983	9 397	40.9%	40.0%	19.2%
syndicat	519	412	79.4%	78.5%	27.0%
autre	659	255	38.7%	62.9%	32.5%

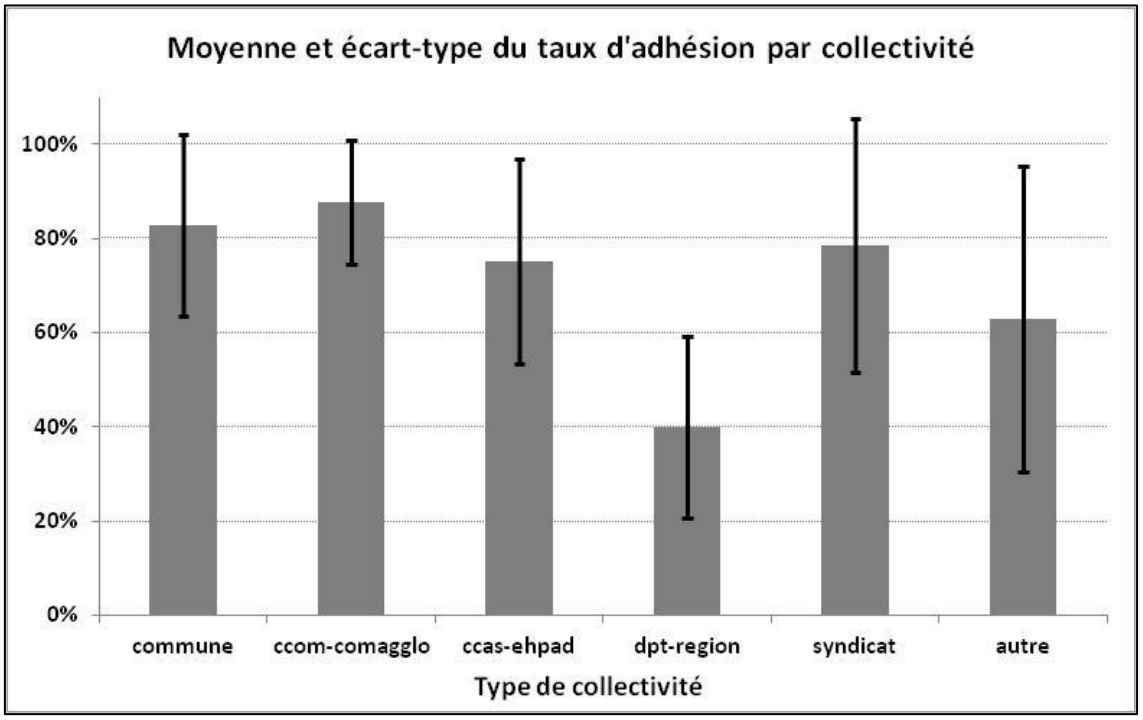


Figure 14 : Moyenne (+/- écart-type) du taux d'adhésion par collectivité en fonction du type de collectivité

Les deux autres variables qui nous intéressent pour l'étude du taux d'adhésion, soit le montant de participation mensuelle par agent et la proportion d'HLM parmi les habitations principales, sont quantitatives.

La représentation graphique des nuages de points peut nous donner quelques indications concernant la relation susceptible d'exister entre le taux d'adhésion et la participation d'une part (fig. 15 a) et entre le taux d'adhésion et la proportion d'HLM d'autre part (fig. 15 b).

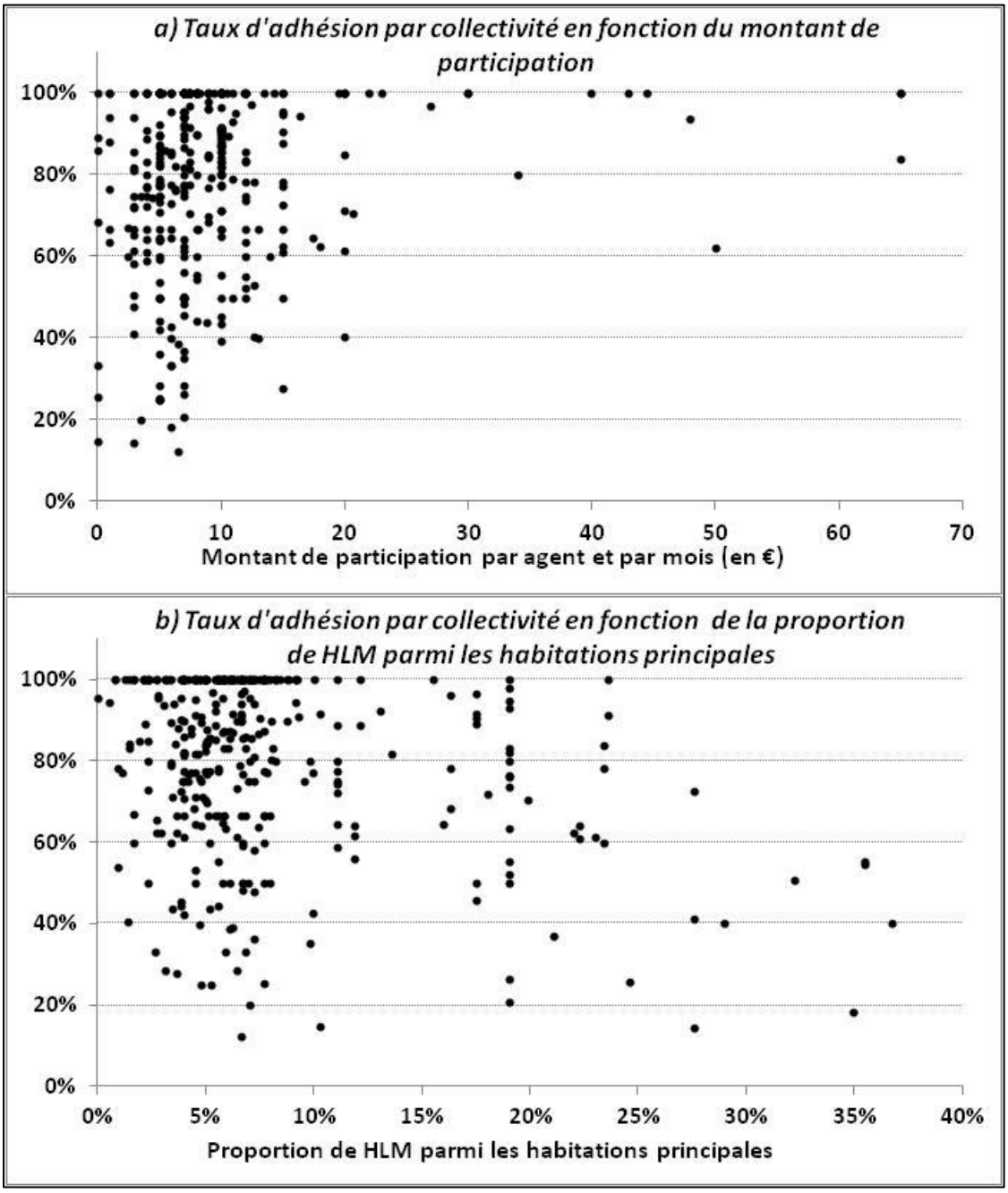


Figure 15 : Nuages de points du taux d'adhésion en fonction du montant de participation mensuel par agent (a) et de la proportion d'habitations principales qui sont des HLM dans la commune de rattachement de la collectivité (b).

La figure 15 permet de faire l'hypothèse qu'il pourrait exister :

- une relation positive entre le taux d'adhésion et le montant mensuel de participation par agent, une participation plus élevée étant ainsi liée à un taux d'adhésion plus élevé

- une relation négative entre le taux d'adhésion et la proportion d'HLM, une proportion d'HLM plus élevée étant liée à un taux d'adhésion plus faible.

Cependant, une fois encore il apparaît, ici au vu des nuages de points, qu'une variable seule ne suffit pas à expliquer la variabilité du taux d'adhésion.

Il ressort donc de cette analyse descriptive du taux d'adhésion en fonction des différentes variables d'intérêt prises une à une, que ces quatre variables (taille, type, participation et proportion de HLM) devront être prises en compte simultanément pour comprendre au mieux leurs relations avec le taux d'adhésion.

Cependant, nous souhaitons également obtenir des résultats qui soient les plus simples possibles à interpréter afin d'être utilisables en pratique.

2. Présentation de la méthode utilisée

* Choix de la méthode *

Une méthode pour caractériser la relation entre une variable, dite variable dépendante ou variable expliquée, et plusieurs autres variables, dites indépendantes ou explicatives, est la **régression linéaire**.

Le principe de la régression linéaire est d'expliquer (et/ou prédire) la valeur moyenne d'une variable Y continue (la variable dépendante) à l'aide de p variables X_1, …, X_p, le modèle de régression linéaire multiple s'écrivant :

$$y_i = \beta_0 + \sum_j \beta_j x_{ij} + \varepsilon_i \qquad i = 1, \dots, n \; ; j=1, \dots, p$$

Où :

- y_i $(i=1, \dots, n)$ est la $i^{ème}$ réalisation de la variable aléatoire Y dont les valeurs sont conditionnées par celles des variables explicatives X_j (j=1, …, p) et de la composante aléatoire ε_i ;
- x_{ij} est la valeur de la variable X_j pour la $i^{ème}$ observation ;
- les β_j sont les paramètres du modèle de régression, non aléatoires mais inconnus et donc à estimer, β_0 étant communément appelé '*intercept*' et correspondant à l'ordonnée à l'origine ;
- les ε_i $(i=1, \dots, n)$ sont des termes aléatoires d'erreur, inconnus mais supposés indépendants et suivre une loi normale $N(0, \sigma)$ avec σ inconnu.

Ce type de modèle est appelé modèle gaussien. Si on pose $E(Y_i)=\mu_i$ alors :

$$\mu_i = \beta_0 + \sum_j \beta_j x_{ij} \qquad (\text{car } E(\varepsilon_i)=0)$$

Le principal inconvénient auquel nous devons faire face dans notre cas est que le modèle gaussien va donner des valeurs dans R pour la variable expliquée, alors que le **taux d'adhésion, qui est une proportion, ne peut prendre des valeurs que dans $[0,1]$**.

La solution est apportée par le **modèle linéaire généralisé**, qui est une généralisation du modèle gaussien à d'autres distributions de la famille exponentielle (ex : Poisson, Gamma, Binomiale). Dans ce type de modèle, on cherche les valeurs des paramètres β_j permettant d'écrire une fonction g de la moyenne μ de Y comme combinaison linéaire des $\mathrm{X_j}$, soit :

$$g(\mu) = \beta_0 + \sum \beta_j \, \mathrm{X_j}$$

La relation linéaire entre μ et les variables explicatives $\mathrm{X_j}$ se fait donc à l'aide de cette fonction g, appelée **fonction de lien**, dont le choix dépend de la nature des données et de la distribution de la variable expliquée.

Le taux d'adhésion dans une collectivité donnée peut être vu comme la probabilité de réalisation d'un événement, *i.e.* l'adhésion, avec un nombre d'expositions, *i.e.* le nombre d'agents dans la collectivité soit le potentiel de cette collectivité, et un nombre de réalisations de cet événement, *i.e.* le nombre d'adhérents. Nous pouvons ainsi considérer que la variable expliquée Y suit une loi binomiale. Le modèle linéaire généralisé qui correspond est alors la **régression logistique**, avec une fonction de lien g appelée **logit** et définie par :

$$g : [0 \quad 1] \rightarrow R$$

$$x \rightarrow \ln\left(\frac{x}{1-x}\right)$$

Son inverse est :

$$g^{-1} : R \rightarrow [0 \quad 1]$$

$$x \rightarrow \frac{\exp(y)}{1+\exp(y)}$$

Ainsi, en notant p_i la probabilité que l'adhésion se réalise sachant les valeurs x_{ij} des $\mathrm{X_j}$, $\mathrm{P(Y=1|X1=x_{i1}, \ldots, Xp=x_{ip})}$, le modèle de régression logistique s'écrit :

$$p_i = \frac{\exp\left(\beta_0 + \beta_1 \, x_{i1} + \ \ldots + \beta_p \, x_{ip}\right)}{1+\exp\left(\beta_0 + \beta_1 \, x_{i1} + \ \ldots + \beta_p \, x_{ip}\right)} = \frac{1}{1+\ e^{-(\beta_0+\beta_1 \, x_{i1}+ \ \ldots+\beta_p \, x_{ip})}} \qquad i = 1, \ldots, n$$

Dans ce type de modèle, l'**estimation des paramètres** β_j se fait par la méthode du **maximum de vraisemblance**.

Dans le cas d'une régression logistique simple (par mesure de simplification des écritures), la fonction de vraisemblance peut s'écrire :

$$L(y,\beta) = \prod_i p_i^{y_i}(1-p_i)^{1-y_i} \qquad avec\ y_i = 0\ ou\ 1\ , y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}\ et\ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Maximiser une fonction équivaut à maximiser son logarithme puisque le logarithme est une fonction strictement croissante. En passant au logarithme népérien et en utilisant le fait que $\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$, on aboutit à :

$$\ln L(y,\beta) = \sum_i \left(y_i[\beta_0 + \beta_1 x_i] - \ln(1 + e^{\beta_0+\beta_1 x_i})\right)$$

L'équation à résoudre pour maximiser la vraisemblance est :

$$\frac{\partial \ln L\,(y,\beta)}{\partial \beta} = 0 \quad (a)$$

Le vecteur $\hat{\beta} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix}$, solution de l'équation (a), est l'estimation du vecteur des paramètres $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$. La résolution se fait par itérations d'algorithmes, implémentés dans les logiciels de statistiques.

La **validation du modèle**, soit le fait de déterminer si le modèle apporte plus d'information qu'un modèle sans variable, se fait par le test du rapport de vraisemblance (LR= 'likelihood ratio').

Sous l'hypothèse H_0, aucun des paramètres estimés n'est significativement différent de 0.

Sous l'hypothèse H_1, au moins un des paramètres estimés est différent de 0.

Soit L_0 la vraisemblance du modèle sans variable et L_1 celle du modèle à valider, alors :

$$LR = 2[lnL_1 - lnL_0] = -2\ln\left(\frac{L_0}{L_1}\right)$$

Sous H_0, LR suit une loi du χ^2 à p degrés de liberté, p étant le nombre de paramètres du modèle à valider (une variable explicative qualitative à i modalités comptant pour *i-1 ddl* puisqu'elle génère *i-1* paramètres).

Le test de la **significativité de chacun des paramètres** du modèle, soit déterminer si un β_j donné est significativement différent de 0, se fait grâce à la statistique de Wald (W), avec :

$$W_j = \frac{\hat{\beta}_j^2}{\widehat{Var}(\hat{\beta}_j)}$$

Sous l'hypothèse H_0 : le paramètre β_j n'est pas significativement différent de 0

Sous l'hypothèse H1 : le paramètre est significativement différent de 0

Sous H_0, W_j suit une loi du χ^2 à 1 degré de liberté.

Une particularité intéressante de la régression logistique est qu'elle permet de calculer des rapports de cotes (OR = 'odds-ratio'). L'odds-ratio est une mesure de dépendance entre deux variables et appartient à $[0; +\infty[$. Lorsqu'il est égal à 1, il y a indépendance.

Dans le cas d'une variable explicative quantitative X, l'odds-ratio permet de quantifier l'impact sur la probabilité p_i de l'augmentation d'une unité de X. En effet, dans une régression logistique simple, pour chaque valeur x_i de X, la probabilité estimée $\hat{p}_i$ vérifie :

$$logit(\hat{p}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Pour $x_i = x_i + 1$ la probabilité estimée $\hat{p}_i'$ vérifie :

$$logit(\hat{p}_i') = \hat{\beta}_0 + \hat{\beta}_1 (x_i + 1)$$

La différence de ces quantités donne $\hat{\beta}_1$, d'où :

$$OR = \frac{\hat{p}_i' / (1 - \hat{p}_i')}{\hat{p}_i / (1 - \hat{p}_i)} = exp(\hat{\beta}_1)$$

Il est également possible de calculer des odds-ratios entre deux modalités d'une même variable explicative qualitative, ce qui permet par exemple dans notre cas de quantifier l'écart de la probabilité d'adhésion entre deux types de collectivités. Dans ce cas, en prenant $\hat{\beta}_{ja}$ l'estimation du paramètre pour le type de collectivité a et $\hat{\beta}_{jb}$ pour le type de collectivité b, l'odds-ratio s'écrira, toutes choses égales par ailleurs (i.e. en contrôlant pour les autres paramètres du modèle) :

$$OR = exp(\hat{\beta}_{jb} - \hat{\beta}_{ja})$$

S'il est supérieur à 1, la probabilité d'adhérer est plus élevée dans la collectivité de type b que dans la collectivité de type a, s'il est inférieur à 1, la probabilité d'adhérer dans la collectivité de type b est plus faible.

* Surdispersion *

Dans les cas où la variance résiduelle est supérieure à la variance spécifiée par le modèle, on parle de **surdispersion** (ce phénomène s'observe dans les régressions binomiales et de poisson, variance et moyenne étant liées). L'origine de la surdispersion peut être un mauvais choix de la fonction de lien, ou également une mauvaise spécification du modèle comme par exemple la non prise en compte de variables ou d'interactions ayant un effet significatif ou encore la présence d'outliers dans les données (Le Rest, 2013 ; Carruthers *et al.* 2008).

Dans notre cas, il est évident que nous ne pouvons pas intégrer dans le modèle tous les paramètres pouvant influer sur la probabilité d'adhésion car nous n'avons pas connaissance de l'ensemble de ces paramètres (à part dans le cas de données expérimentales où tous les paramètres sont connus et contrôlés, il est peu probable d'être exhaustif). Par conséquent, il est fort probable que nous constations l'existence de surdispersion lors de l'ajustement du modèle.

Un moyen pour détecter la surdispersion dans un modèle est d'utiliser la déviance du modèle et le Khi-deux de Pearson.

La déviance d'un modèle M est une mesure de son écart au modèle saturé et s'écrit :

$$D_M = ln\left[\left(\frac{L_{sat}}{L_M}\right)^2\right] = 2(lnL_{sat} - lnL_M)$$

Avec L_M la vraisemblance du modèle M et L_{sat} la vraisemblance du modèle saturé (i.e. reconstituant parfaitement les données), qui est égale à 1.

On obtient donc $D_M = -2\ (lnL_M)$.

Le Khi-deux de Pearson est quant à lui défini par :

$$\chi_P^2 = \sum {(y_i - \hat{\mu}_i)^2}\Big/{V(\hat{\mu}_i)}$$

Si le modèle est correct, la déviance et le khi-deux de Pearson suivent approximativement une loi du χ^2 à (nb d'observations – k) *ddl,* k étant fonction du nombre de variables dans le modèle et du nombre de modalités par variable.

Dans les deux cas, si le rapport entre la valeur et le nombre de degrés de liberté est supérieur à 1, il existe de la surdispersion.

La présence de surdispersion ne change pas la valeur des $\hat{\beta}_j$ mais si on ne corrige pas ce problème de surdispersion dans les résultats, des variables peuvent sembler avoir un effet significatif alors que ce n'est pas le cas (un ou des $\hat{\beta}_j$ apparaîtraient significativement différents de 0 alors qu'ils ne le seraient pas, la surdispersion rendant le test de Wald trop significatif).

Une solution pour corriger les résultats pour le problème de surdispersion est de multiplier la fonction de variance $V(\mu)$ par un paramètre de dispersion φ. Il est souvent conseillé d'utiliser, pour estimer ce coefficient, la valeur d'un des deux rapports cités ci-dessus (Déviance/ddl ou Khi-deux de Pearson/ddl).

Cependant, dans les cas d'échantillons de tailles différentes, il est préférable d'utiliser la méthode de Williams (Williams, 1982 ; Hardin & Hilb, 2012). Nous sommes dans ce type de situation car les collectivités étudiées ont des potentiels très différents.

Le principe de cette méthode est d'estimer φ par itérations successives en approximant la valeur du Khi-deux de Pearson pour le modèle complet par sa valeur attendue en l'absence de surdispersion.

Une fois $\hat{\varphi}$ obtenu, une pondération ayant pour valeur $\frac{1}{\left(1+\hat{\varphi}(n_i-1)\right)}$ est appliquée à l'observation i, n_i étant la taille de l'échantillon de cette observation, ce qui correspond dans notre cas au potentiel de la collectivité.

En pratique, pour utiliser cette méthode (implémentée notamment dans SAS dans la *proc logistic*), il faut en premier lieu ajuster un modèle complet en spécifiant l'utilisation de la méthode de Williams, ce qui donnera la valeur de $\hat{\varphi}$. Ensuite, il est possible d'utiliser cette valeur de $\hat{\varphi}$ pour ajuster des modèles avec moins de paramètres.

* <u>Qualité d'ajustement du modèle aux données (GOF : Goodness Of Fit)</u> *

Les tests couramment utilisés pour évaluer l'ajustement d'un modèle gaussien aux données ne sont pas utilisables dans le cas d'une régression logistique (Rotella, 2013). Or il ne suffit pas de déterminer si les variables explicatives du modèle ont une incidence sur la probabilité de réalisation de l'événement étudié (cf validation du modèle), il faut déterminer si le modèle s'ajuste bien aux données. Cette étape est d'autant plus nécessaire si l'on veut comparer plusieurs modèles entre eux (Cooch & White, 2014).

Les tests de la déviance et du Khi-deux de Pearson (cf 'surdispersion') peuvent être une façon de vérifier si le modèle s'ajuste correctement aux données mais dans le cas où les résultats sont corrigés pour la surdispersion, ils ne sont plus utilisables.

Un test d'ajustement très couramment utilisé dans le cas de la régression logistique est le test de Hosmer & Lemeshow (1980). Le principe de ce test est de classer les données par ordre croissant des probabilités calculées par le modèle, puis de les partager en plusieurs groupes (la plupart du temps 10). Le test du Khi-deux (à nb de groupes -2 *ddl*) est ensuite utilisé pour comparer les effectifs observés dans chaque classe aux effectifs théoriques. Cependant, ce test est peu puissant et surtout, on peut arriver à des conclusions opposées si l'on change le nombre de groupes (Allison, 2013 ; Hosmer *et al.*, 1997). Ainsi, il est possible dans certains cas de conclure que le modèle s'ajuste correctement aux données en prenant 10 groupes (test du Khi-deux non significatif) et de conclure l'inverse si on choisit 9 ou 11 groupes (test du Khi-deux significatif) !

Plusieurs autres tests d'ajustement pour la régression logistique ont été proposés (Hallett, 1999 ; Allison, 2014a). L'un d'entre eux, le test de Stukel (Stukel, 1988), est plus puissant (Hosmer *et al.*, 1997) et plus adapté pour détecter un mauvais ajustement lié à la non prise en compte d'une ou plusieurs interactions (Allison, 2014b).

Le principe du test de Stukel repose sur la généralisation du modèle logistique en ajoutant 2 paramètres supplémentaires (z_a et z_b) dans le modèle, qui permettent une asymétrie du modèle. Le modèle est le modèle logistique linéaire habituel quand z_a=0 et z_b=0.

Les 2 paramètres supplémentaires sont calculés de la façon suivante :

Si on note g_i le prédicteur linéaire du modèle ajusté, soit $g_i = X_i'\hat{\beta}$, où X_i est le vecteur des valeurs des variables du modèle pour l'observation i et $\hat{\beta}$ le vecteur des coefficients estimés, alors :

$$z_a = g^2 \; si \; g \geq 0, \qquad z_a = 0 \; sinon$$

$$z_b = g^2 \; si \; g < 0, \qquad z_b = 0 \; sinon$$

La nullité des deux paramètres peut être testée par un test du score avec pour hypothèse nulle H_0 : les coefficients des deux paramètres sont égaux à 0. Sous l'hypothèse H_0, la statistique de ce test suit une loi du χ^2 à 2 *ddl*.

* <u>Sélection du modèle</u> *

Un critère souvent utilisé pour sélectionner un modèle parmi plusieurs modèles candidats est l'AIC, ou Critère d'Information d'Akaike (Akaike, 1974). En effet, la vraisemblance d'un modèle augmente mécaniquement quand on ajoute des paramètres dans le modèle, et l'AIC, qui est un critère parcimonieux, pénalise les modèles ayant plus de variables. Il s'écrit :

$$AIC = -2lnL + 2k$$

Avec L la vraisemblance maximisée et k le nombre de paramètres du modèle.

Le meilleur modèle est celui ayant l'AIC le plus faible, une différence entre deux modèles étant usuellement considérée non nulle quand $|\Delta AIC| \geq 2$.

Dans le cas où de la surdispersion a été identifiée, le critère à utiliser n'est pas l'AIC mais le QAIC, qui est de la forme :

$$QAIC = \frac{-2lnL}{\hat{\varphi}} + 2k$$

avec $\hat{\varphi}$ l'estimation du paramètre de dispersion et k le nombre de paramètres du modèle.

Une autre méthode qui peut être utilisée est la sélection '*backward*', qui consiste à ajuster le modèle complet et retirer une à une les variables non significatives, en commençant par celle ayant la p-value la plus élevée, puis en ajustant le nouveau modèle et ainsi de suite. La procédure inverse est appelée sélection '*forward*', et celle en dérivant qui consiste à ajouter et enlever des variables jusqu'à trouver le meilleur modèle est la sélection '*stepwise*'. Ces méthodes utilisent souvent également le critère AIC.

Un autre critère, qui lui donne une idée du pouvoir prédictif du modèle, est l'aire sous la courbe de ROC (Receiver Operating Characteristics), parfois aussi appelée AUC. Cette courbe s'obtient en représentant graphiquement la sensibilité du modèle en fonction de (1-spécificité du modèle) pour chaque valeur du seuil de décision (généralement fixé à 0,5) entre 0 et 1. La sensibilité du modèle correspond à sa capacité à prédire un événement, et la spécificité à sa capacité à prédire un non-événement.

Plus l'AUC est élevée, plus le modèle a une bonne capacité de prédiction, le maximum étant 1, et 0,5 correspondant au hasard.

3. Application au taux d'adhésion

Pour analyser la relation entre le taux d'adhésion et les quatre variables retenues pour caractériser la collectivité (type, taille, participation et proportion de HLM), nous avons donc ajusté un modèle logistique ayant pour variable expliquée le taux d'adhésion dans une collectivité donnée. Cette proportion est renseignée en indiquant le nombre d'adhérents *vs* le potentiel de la collectivité. Tous les modèles de cette partie ont été ajustés à l'aide de la 'proc logistic' du logiciel SAS, et le seuil de significativité utilisé est toujours $\alpha = 5\%$.

Les variables explicatives sont à la fois des variables qualitatives (type et potentiel de la collectivité) et des variables quantitatives (participation et proportion d'HLM).

Nous souhaitions obtenir un modèle qui soit le plus simple possible (facilité d'interprétation et éventuellement d'utilisation ultérieure des résultats) tout en étant correctement ajusté aux données et avec un pouvoir prédictif acceptable.

* Modèle complet et sélection du modèle final*

Nous avons en premier lieu ajusté un modèle que nous avons défini comme complet afin de pouvoir déterminer si nous observions de la surdispersion. Le modèle complet inclut les quatre variables, ainsi que l'interaction entre les variables deux à deux. En effet, il est important de tenir compte de l'interaction entre les variables même si cela complexifie l'interprétation des résultats : il est tout à fait possible par exemple que la relation entre le taux d'adhésion et le montant de participation, si elle existe, diffère selon le type de collectivité. La seule interaction double que nous n'avons pas intégrée dans le modèle est celle entre le type et le potentiel de la collectivité car, comme nous l'avons précisé dans la description des données, certains types de collectivités sont trop peu représentés entre les différentes classes de taille. Nous n'avons pas non plus intégré dans le modèle d'interactions triples car elles sont beaucoup trop complexes à interpréter.

Après ajustement, nous avons regardé en premier les résultats des tests de la déviance et du Khi-deux de Pearson (table VII).

Table VII : Résultat des tests de la Déviance et du Khi-deux de Pearson pour le modèle complet sans correction pour la surdispersion

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	**Value**	**DDL**	**Value/DF**	**Pr > Khi-2**
Ecart	3593.2229	334	10.7582	<.0001
Pearson	3392.2517	334	10.1564	<.0001

Les p-values sont très inférieures à 0.05, ce qui indique une distribution observée significativement différente d'une distribution du ^{2}, et si l'on s'intéresse au rapport entre soit la Déviance soit le Khi-deux de Pearson et le nombre de ddl, dans les deux cas les valeurs sont très supérieures à 1, ce qui indique de la surdispersion.

Il faut donc intégrer une correction pour tenir compte de cette surdispersion et ne pas risquer de considérer comme ayant un effet significatif des variables qui n'en ont pas. Pour pouvoir ensuite faire une comparaison, la table VII présente les résultats de l'analyse de type III des effets de ce modèle.

Table VIII : Tests de significativité des effets dans le modèle complet sans prise en compte de la surdispersion

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	588.4674	<.0001
HLM2	1	5.9372	0.0148
HLM2*typecoll2	5	200.3150	<.0001
classpot	6	423.5359	<.0001
Participation	1	14.6970	0.0001
Participati*classpot	6	593.3798	<.0001
HLM2*Participation	1	170.6371	<.0001
HLM2*classpot	6	140.8789	<.0001
Participat*typecoll2	5	177.4009	<.0001

La notation des variables est la suivante et restera le même dans la suite du document :

- typecoll2 : type de collectivité, les modalités étant détaillées dans la partie II.A ;

- HLM2 : proportion d'HLM parmi les habitations principales dans la commune de rattachement de la collectivité ;

- classpot : potentiel de la collectivité, i.e. nombre d'agents susceptibles d'adhérer ;

- Participation : montant de participation mensuel par agent de la collectivité.

D'après la table VIII, les effets de toutes les variables et de toutes leurs interactions apparaissent hautement significatifs (p<0.0001).

Si maintenant on ajuste le même modèle en tenant compte de la surdispersion par la méthode de Williams, la Déviance et le Khi-deux de Pearson prennent les valeurs présentées dans la table IX

Table IX : Déviance et du Khi-deux de Pearson pour le modèle complet avec correction par la méthode de Williams pour la surdispersion (attention, les résultats des tests ne peuvent plus être utilisés pour valider l'ajustement du modèle)

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	**Value**	**DDL**	**Value/DF**	**Pr > Khi-2**
Ecart	362.3555	334	1.0849	0.1372
Pearson	333.9959	334	1.0000	0.4898

Le deuxième élément à vérifier est l'ajustement du modèle aux données. Le résultat du test de Stukel est présenté dans la table X.

Table X : Résultat du test de Stukel pour le modèle complet avec prise en compte de la surdispersion

Linear Hypotheses Testing Results				
	Label	**Wald Chi-Square**	**DDL**	**Pr > Khi-2**
	Test 1	1.3679	2	0.5046

La distribution de la statistique de Stukel n'est pas significativement différente d'une distribution du χ^2 (p>0.05), nous pouvons donc considérer que le modèle complet avec prise en compte de la surdispersion s'ajuste bien aux données.

Le $\hat{\varphi}$ obtenu avec ce modèle devra être renseigné dans tous les autres modèles ajustés puisqu'ils devront désormais tous tenir compte de la surdispersion. Sa valeur est donnée par la variable de pondération renseignée dans les informations du modèle (table XI).

Table XI : Informations du modèle complet avec prise en compte de la surdispersion

Model Information		
Table	STATMEM.DATPARTICIPATION	
Variable de réponse (Evénements)	adher	adher
Variable de réponse (Expériences)	potentiel	potentiel
Variable de pondération	1 / (1 + 0.166252 * (potentiel - 1))	potentiel
Modèle	logit binaire	

Ainsi, $\hat{\varphi} = 0.166252$.

On peut donc maintenant regarder les résultats des tests de validation du modèle (table XII) et de significativité des effets (table XIII).

Table XII : Test de validation du modèle complet

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DDL**	**Pr > Khi-2**
Rapport de vrais	140.1383	36	<.0001

Le test du rapport de vraisemblance indique qu'au moins un des β_j du modèle est significativement différent de 0 (p<0.0001) et donc que le modèle apporte plus d'information qu'un modèle sans variable.

Table XIII : Tests de significativité des effets dans le modèle complet

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	21.2892	0.0007
HLM2	1	0.2787	0.5975
HLM2*typecoll2	5	5.0214	0.4133
classpot	6	12.0527	0.0608
Participation	1	3.3959	0.0654
Participati*classpot	6	11.9022	0.0642
HLM2*Participation	1	0.0611	0.8048
HLM2*classpot	6	1.8000	0.9371
Participat*typecoll2	5	8.1081	0.1504

Maintenant que la surdispersion est prise en compte, nous constatons que les variables du modèle n'ont pas toutes un effet significatif.

Nous pouvons ensuite procéder à la sélection backward en retirant une à une les variables non significatives en commençant par celle ayant la p-value la plus élevée. Ici, il s'agit de l'interaction entre le potentiel de la collectivité et la proportion d'HLM (HLM2*classpot, p=0.9371).

On supprime donc cette interaction et on ajuste un nouveau modèle, qui a un test de validation significatif (p<0.001) et dont les effets sont présentés dans la table XIV.

Table XIV : Tests de significativité des effets du premier modèle intermédiaire (M1)

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	24.7214	0.0002
HLM2	1	0.4486	0.5030
HLM2*typecoll2	5	7.2993	0.1993
classpot	6	17.9314	0.0064
Participation	1	3.7215	0.0537
Participati*classpot	6	12.2004	0.0576
Participat*typecoll2	5	8.1970	0.1457
HLM2*Participation	1	0.1250	0.7237

La p-value la plus élevée est celle de l'interaction entre la proportion d'HLM et le montant de participation (p=0.7237). On supprime cette interaction du modèle et on ajuste le modèle M2 (table XV), avec un test de validation significatif (p<0.0001).

Table XV : Tests de significativité des effets du deuxième modèle intermédiaire (M2)

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	24.8062	0.0002
HLM2	1	2.8961	0.0888
HLM2*typecoll2	5	7.2451	0.2030
classpot	6	18.2221	0.0057
Participation	1	4.0788	0.0434
Participati*classpot	6	12.3917	0.0538
Participat*typecoll2	5	8.6537	0.1237

La p-value la plus élevée est celle de l'interaction entre le type de collectivité et la proportion d'HLM (p=0.2030), que l'on supprime pour obtenir un troisième modèle intermédiaire (table XVI), dont le test de validation est significatif (p<0.0001).

Table XVI : Tests de significativité des effets du troisième modèle intermédiaire (M3)

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	21.9741	0.0005
HLM2	1	4.0530	0.0441
classpot	6	20.3570	0.0024
Participation	1	4.3677	0.0366
Participati*classpot	6	12.5645	0.0505
Participat*typecoll2	5	8.6323	0.1247

Le modèle M4 est ajusté après suppression de l'interaction entre le type de collectivité et le montant de participation (p=0.1247). Le test de validation de M4 est significatif (p<0.0001) et les tests de significativité des effets sont présentés dans la table XVII.

Table XVII : Tests de significativité des effets du quatrième modèle (M4)

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	28.9370	<.0001
HLM2	1	3.9913	0.0457
classpot	6	24.9989	0.0003
Participation	1	6.0191	0.0142
Participati*classpot	6	13.7068	0.0331

Pour ce modèle M4, les effets des variables sont tous significatifs. Ce modèle est donc un bon candidat pour décrire la relation entre le taux d'adhésion et les caractéristiques de la collectivité. Il nous reste à vérifier que ce modèle s'ajuste bien aux données. Le test de Stukel n'est pas significatif (p=0.9333), donc nous pouvons considérer que le modèle s'ajuste bien aux données.

Dans la mesure où une seule interaction est significative dans le modèle candidat, nous ajustons également le modèle additif et vérifierons si M4 est mieux ou non que ce modèle additif. En effet, les résultats sont plus faciles à interpréter en absence d'interaction.

Le test de validité du modèle additif est significatif (p<0.001) et nous pouvons considérer que le modèle s'ajuste bien aux données car le test de Stukel n'est pas significatif (p=0.1334), même si sa p-value est beaucoup plus faible que celle du modèle M4.

Les variables du modèle additif ont toutes un effet significatif (table XVIII).

Table XVIII : Tests de significativité des effets du modèle additif

Type 3 Analysis of Effects			
Effect	**DDL**	**Wald Chi-Square**	**Pr > Khi-2**
typecoll2	5	26.2644	<.0001
HLM2	1	3.8818	0.0488
classpot	6	13.6294	0.0341
Participation	1	8.5633	0.0034

Pour déterminer quel modèle conserver, nous comparons les QAIC et les AUC (aire sous la courbe de ROC) des deux modèles (table XIX et fig. 16).

Table XIV : Comparaison de M4 et du modèle additif

	QAIC	AUC
Additif	575.4	0.735
M4	573.1	0.763

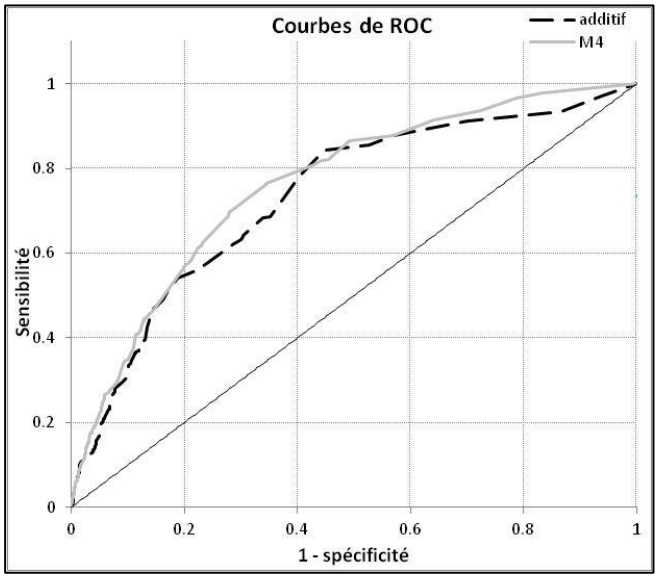


Figure 16 : Courbes de ROC de M4 et du modèle additif

Le QAIC du modèle M4 est inférieur à celui du modèle additif, avec $|\Delta QAIC| = 2{,}3 > 2$, et l'aire sous la courbe de ROC de M4 est supérieure à celle du modèle additif (on voit sur la figure 16 qu'à une exception près, la courbe de ROC de M4 est toujours au-dessus de celle du modèle additif). Avec une AUC de 0.763, le pouvoir prédictif de M4 est acceptable.

Nous **conservons donc le modèle** M4, **qui contient toutes les variables et l'interaction entre le potentiel de la collectivité et le montant de participation**.

* Analyse du modèle final *

La figure 17 illustre les effets des deux variables qui n'entrent pas dans l'interaction : le type de collectivité et la proportion d'HLM parmi les habitations principales.

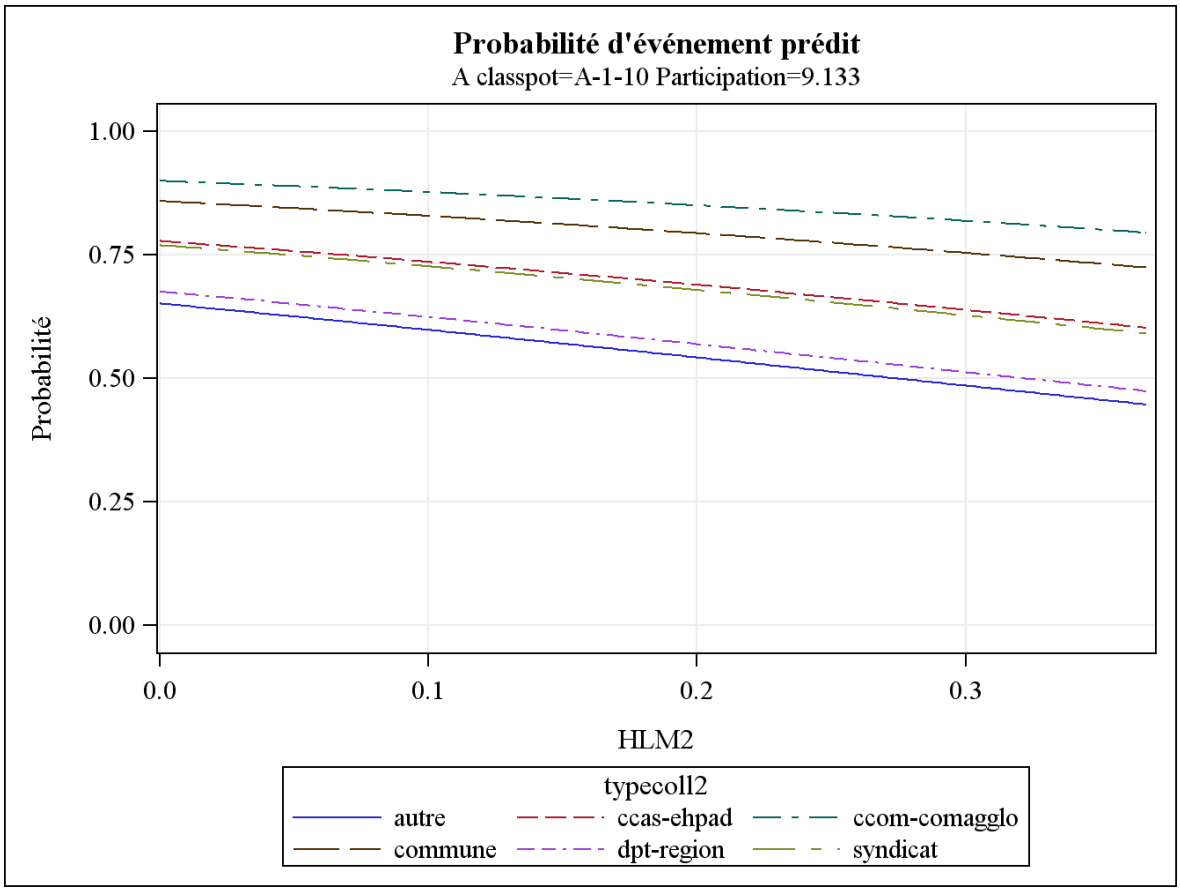


Figure 17 : Valeurs prédites du taux d'adhésion en fonction de la proportion d'HLM dans les habitations principales par type de collectivité, pour les collectivités de 1 à 10 agents, avec une participation de 9,13€ par mois et par agent (correspond au montant moyen pour l'ensemble des collectivités)

NB : Dans l'analyse des résultats, nous n'utiliserons les odds-ratios que pour qualifier la tendance de la relation avec une variable quantitative et déterminer s'il existe des écarts significatifs de taux d'adhésion, et leur sens mais pas leur valeur, entre les différentes modalités des variables qualitatives. Nous préférons attendre d'avoir un peu plus de données (de nouvelles conventions de participation prendront effet au 01/01/2015) pour

dans un premier temps vérifier que le modèle retenu est toujours valable et dans un deuxième temps quantifier les tendances et les écarts.

L'hypothèse faite dans l'analyse descriptive des données est validée : le **taux d'adhésion diminue quand la proportion d'HLM dans les habitations principales augmente** ($\beta_{HLM2} = -2.281 < 0 \; et \; OR_{HLM2} = 0.102 < 1$). La pente est la même pour tous les types de collectivités car il n'y a pas d'interaction significative entre la proportion d'HLM et le type. Elle sera également identique pour toutes les classes de taille car là non plus il n'y a pas d'interaction significative entre les deux variables.

La figure 17 permet également d'illustrer le fait que le taux d'adhésion diffère entre les types de collectivités. Trois groupes semblent se distinguer :

- les ccom-comagglo et les communes avec les plus hauts taux ;

- les ccas-ehpad et syndicats avec des taux intermédiaires ;

- les dpt-région avec les plus faibles taux.

Nous ne nous intéresserons pas au taux d'adhésion du type 'autre', qui n'a pas de valeur en soi mais a été créé uniquement pour conserver les données pour les autres variables.

Pour déterminer si les différences de taux d'adhésion entre les types de collectivités sont significatives, nous utilisons les odds-ratios entre les types deux à deux. La figure 18 présente ces odds-ratios et leurs intervalles de confiance. Si la valeur 1 est incluse dans l'intervalle de confiance, le taux d'adhésion ne diffère pas significativement entre les deux types de collectivités, s'il n'est pas dans l'intervalle, la différence est significative.

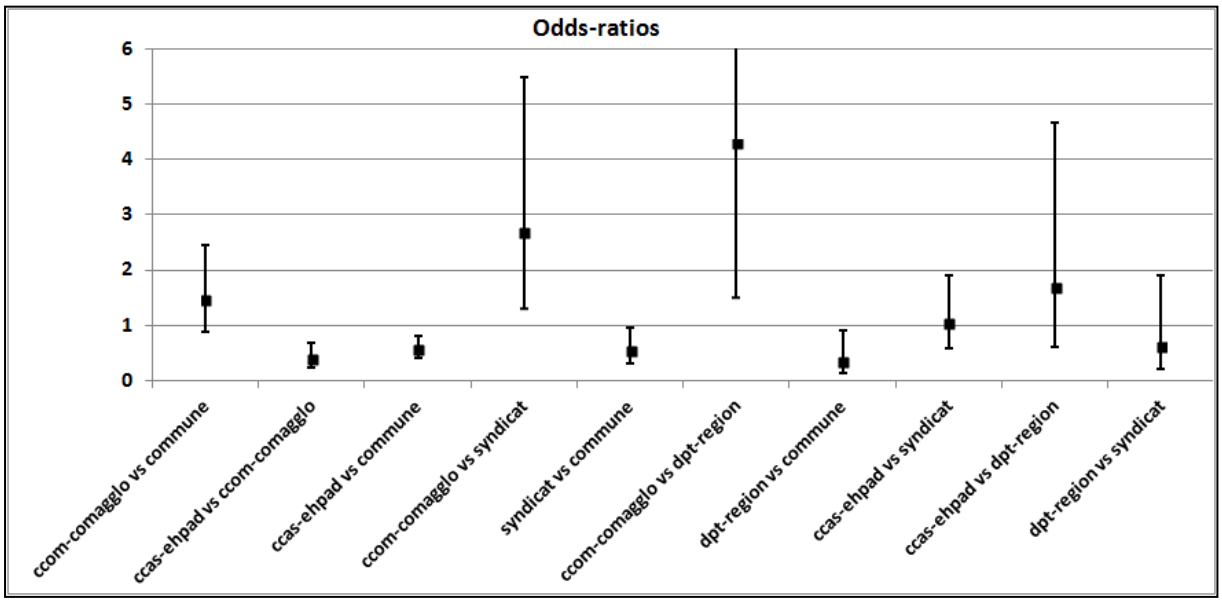


Figure 18 : Odds-ratios (et IC à 95%) entre les différents types de collectivités

Si l'on reprend le groupe commune et ccom-comagglo, on voit bien que l'intervalle de l'OR ccom-comagglo *vs* commune (à gauche dans la figure) inclut 1, et qu'il n'y a donc pas de différence significative entre les taux d'adhésion dans ces deux types de collectivités.

Ensuite, concernant le groupe ccas-ehpad et syndicat, les intervalles de confiance des OR avec les communes et ccom-comagglo n'incluent pas 1, et selon le sens de spécification l'OR est >1 ou <1. En revanche, l'intervalle de l'OR entre ccas-ehpad et syndicat inclut 1, donc le taux d'adhésion ne diffère pas significativement entre les syndicats et les ccas-ehpad, mais ces deux types de collectivités ont des taux d'adhésion plus faibles que les communes et ccom-comagglo.

Les intervalles de confiance des OR des communes et ccom-comagglo *vs* les dpt-région n'incluent pas 1 : le taux d'adhésion est significativement plus faible dans les dpt-région que dans les communes et ccom-comagglo.

Enfin, les intervalles de confiance des OR entre le groupe ccas-ehpad + syndicat et dpt-région contiennent 1, donc il n'existe pas de différence significative entre les taux d'adhésion de ces types de collectivités.

En résumé, **deux grands groupes se distinguent : les communes et ccom-comagglo, avec des taux d'adhésion significativement plus élevés que les ccas-ehpad, syndicats et dpt-région**.

L'effet des deux autres variables, participation et potentiel de la collectivité, est plus complexe à interpréter car il existe une interaction significative.

La figure 19 représente les valeurs prédites par le modèle du taux d'adhésion en fonction du montant de participation, pour les différents types de collectivités (le type ccom-comagglo n'est pas représenté car très similaire à la commune) et une proportion d'HLM dans les habitations de 8,4% (moyenne pour l'ensemble des collectivités).

Dans l'ensemble, le taux d'adhésion augmente quand le montant de participation augmente, ce qui confirme l'hypothèse faite dans l'analyse descriptive, avec des relations différentes selon le potentiel de la collectivité.

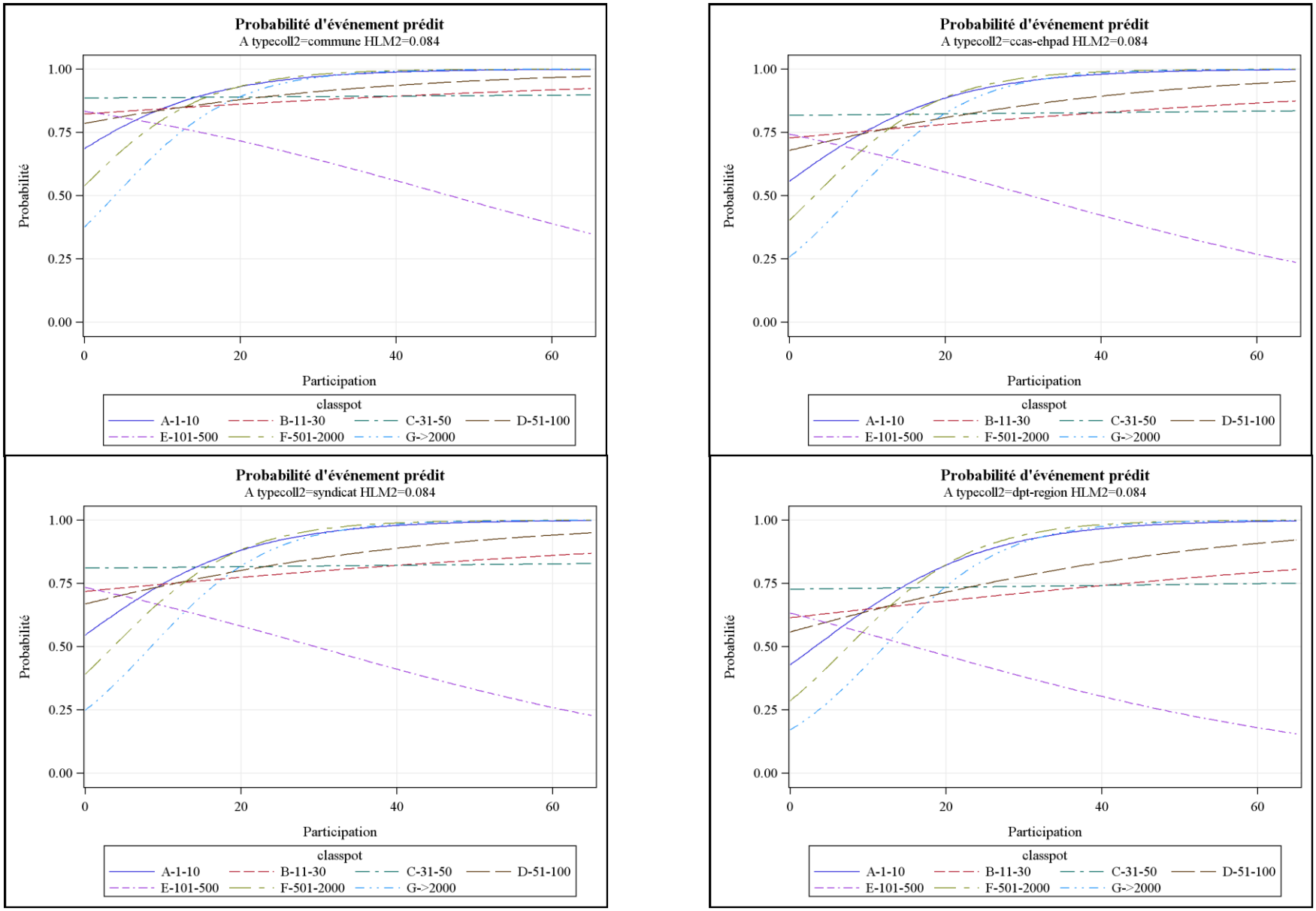


Figure 19 : Taux d'adhésion prédit par potentiel en fonction de la participation pour différents types de collectivités

Cependant, un élément surprenant concerne la classe E-101-500 : le taux d'adhésion prédit par le modèle diminue quand le montant de participation augmente, ce qui est totalement contre-intuitif. Cette classe est correctement représentée puisqu'elle contient 38 collectivités. Il est possible qu'un (ou des) outlier n'ai pas été identifié dans les nuages de points de l'analyse descriptive de l'ensemble des données. Nous allons donc vérifier si cette tendance peut être due à un outlier.

La figure 20 représente le nuage de points du taux d'adhésion en fonction du montant de participation pour la classe E-101-500 uniquement.

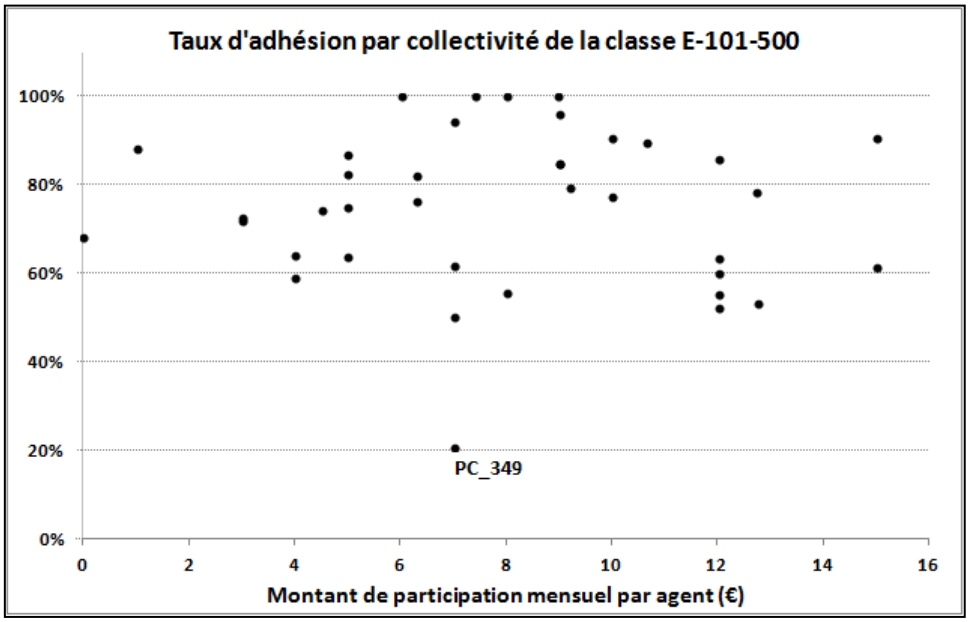


Figure 20 : Taux d'adhésion par collectivité en fonction du montant de participation

D'après le nuage de points, une collectivité semble s'éloigner des autres : la PC_349. Nous ajustons donc le modèle sans cette collectivité pour déterminer s'il est toujours valable, s'ajuste toujours aux données et surtout si cette tendance négative persiste.

Le test de validation du modèle est significatif (p<0.0001) et le modèle s'ajuste toujours aux données (p=0.9546). La tendance quant à elle reste négative (fig. 21).

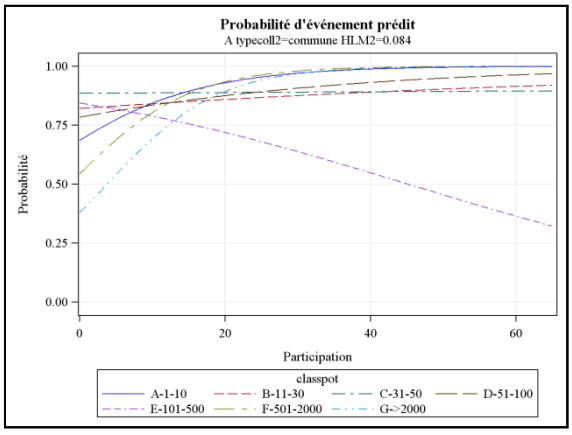


Figure 21 : Taux d'adhésion prédit, après suppression de la collectivité PC_349

Un autre élément à prendre en compte est que le montant de participation observé pour cette classe de potentiel ne dépasse pas 15€, donc il est possible que la forme de la relation change au-delà de 15€. Il nous faudra pour cela des données supplémentaires pour conclure sur cette particularité. Mais nous pouvons pour le moment conserver le modèle tel quel.

Dans la mesure où 96% des collectivités participent à hauteur de 20€ ou moins, même si 5 classes de potentiel sur 7 ont des collectivités qui participent au-delà de 20€, nous nous concentrerons sur l'intervalle [0, 20] pour comparer la relation taux d'adhésion–participation entre les différentes classes de taille.

Si l'on s'intéresse de nouveau à la figure 19, on constate que le taux d'adhésion augmente plus vite avec la participation pour les très petites collectivités (A-1-10) et pour les très grandes (F-501-2000 et G->2000). Il augmente un peu moins rapidement pour les classes B-11-30 et D-51-100, et semble même ne pas augmenter avec l'augmentation de la participation pour la classe C-31-50.

Cette différence ne vient pas du fait que le taux d'adhésion le plus faible serait plus bas pour les classes A-1-10, F-501-2000 et G->2000 que pour les classes B-11-30, C-31-50 et D-51-100 car dans les deux groupes les taux les plus faibles sont autour de 15%. Les taux les plus élevés sont quant à eux de 100% pour toutes les classes (sur la totalité de l'étendue du montant de participation) sauf pour les classes F et G (92% et 79% respectivement), ce qui tendrait plutôt à réduire la pente, donc l'inverse de ce qui est observé. Ces différences de pentes ne proviennent donc très certainement pas d'un artefact lié aux données.

S'il existe un lien de causalité entre les deux (pour rappel ce type de modèle met en évidence des liens significatifs entre des variables mais en aucun cas ne permet d'affirmer que l'une des variables est la cause de l'autre), alors le montant de participation a un impact plus important sur le taux d'adhésion dans les très petites et les très grandes collectivités que dans les tailles intermédiaires.

Cette différence entre les différentes tailles de collectivités semble se réduire avec l'augmentation du montant de participation, à part pour la classe E-101-500 qui serait 'atypique'. La figure 22 illustre ce phénomène : elle représente, pour trois montants de participation, 5€, 10€ et 20€, le taux d'adhésion prédit par le modèle en fonction de la proportion d'HLM. On voit bien qu'à part pour la classe E-101-500, les courbes se rapprochent avec l'augmentation du montant de participation.

Cette observation se confirme avec le calcul des odds-ratios et de leurs intervalles de confiance en fixant le montant de participation (fig. 23). Pour 5€ de participation, on observe 7 différences significatives : la classe G->2000 a un taux d'adhésion significativement inférieur à toutes les autres classes de potentiel sauf F-501-2000, et la classe C-31-50 a en plus un taux d'adhésion significativement supérieur à A-1-10 et F-501-2000. Quand la participation atteint 20€, la seule différence significative est entre A-1-10 et E-101-500.

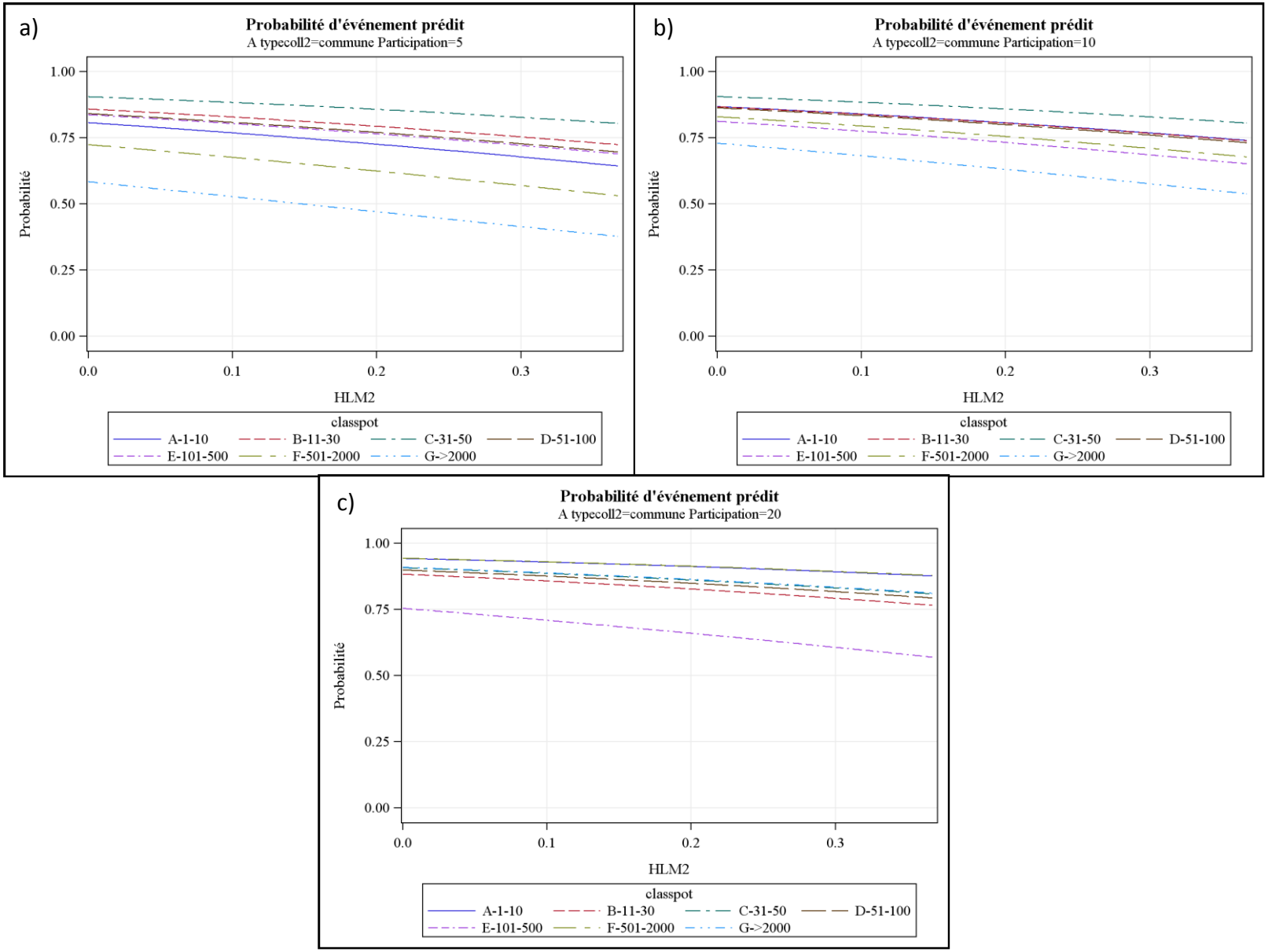


Figure 22 : Taux d'adhésion prédit en fonction de la proportion d'HLM pour une commune pour 3 niveaux de participation : 5€ (a), 10€ (b) et 20€ (c)

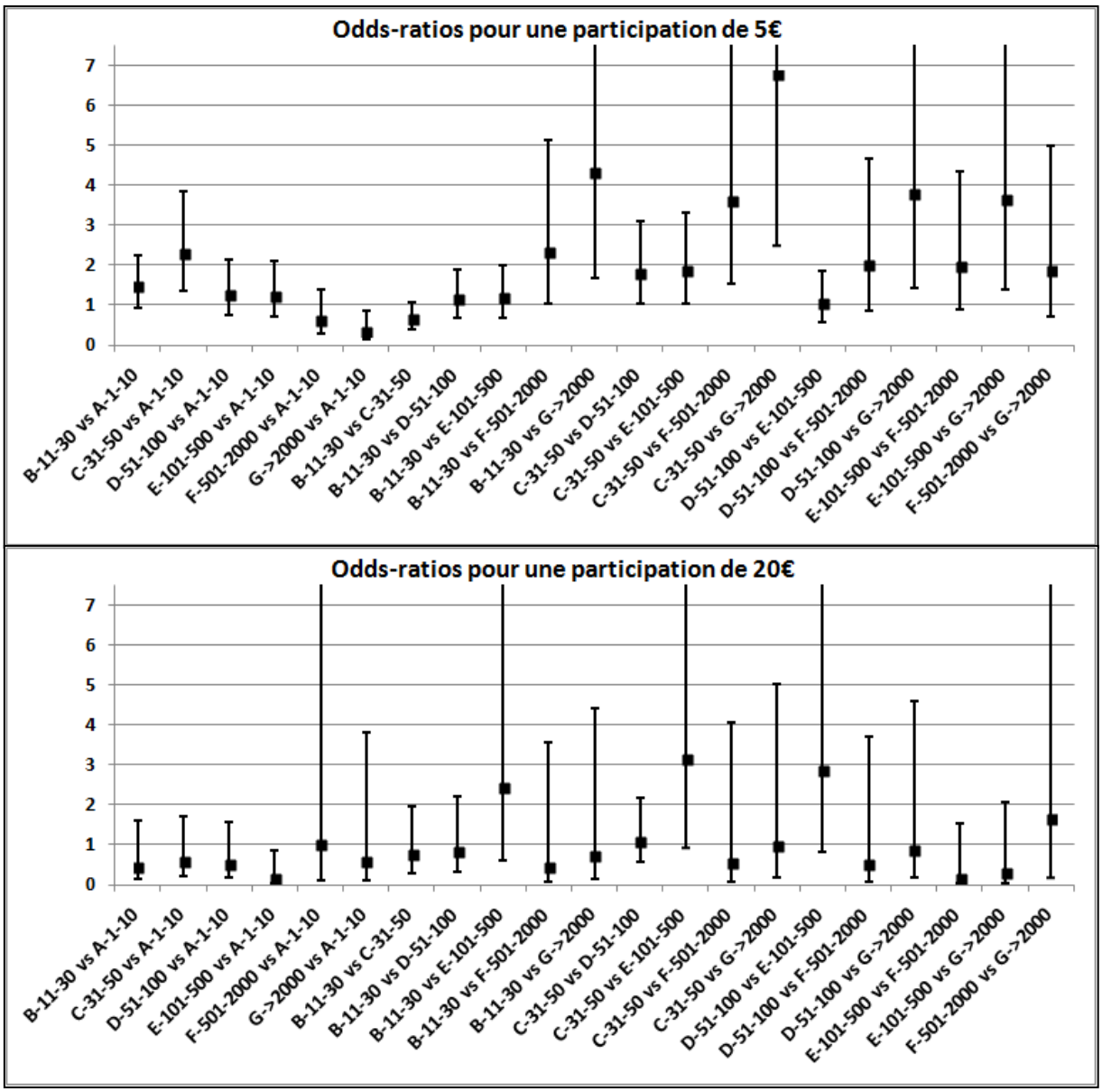


Figure 23 : Odds-ratios entre les différentes classes de potentiel pour 5€ et 20€ de participation

L'augmentation du montant de participation tendrait donc à réduire les écarts de taux d'adhésion entre les différentes tailles de collectivités.

4. Synthèse concernant le taux d'adhésion

En synthèse de l'analyse du taux d'adhésion en fonction de différentes caractéristiques de la collectivité nous pouvons retenir les points suivants :

- le taux d'adhésion a un lien significatif avec les quatre caractéristiques sélectionnées, soit le type de collectivité, le potentiel de la collectivité, le montant mensuel de participation par agent et la proportion d'HLM dans les habitations principales, ainsi qu'avec l'interaction entre le potentiel et le montant de participation, les prédictions données par le modèle étant exactes dans 76% des cas (AUC) ;

- le taux d'adhésion diminue quand la proportion d'HLM dans les habitations principales augmente ;

- concernant le type de collectivité, deux groupes se distinguent : les communes, communautés de communes et communautés d'agglomération d'une part, qui ont un taux d'adhésion significativement plus élevé que les CCAS (et CIAS), les EHPAD, les syndicats et les conseils généraux et régionaux d'autre part ;

- la relation entre le taux d'adhésion et le montant de participation varie selon le potentiel de la collectivité. Ainsi, le taux d'adhésion augmente plus fortement avec le montant de participation dans les très petites collectivités, *i.e.* 10 agents et moins, et dans les très grandes, i.e. 500 agents et plus, que dans les collectivités de taille intermédiaire. Il semble même diminuer avec l'augmentation du montant de participation dans les collectivités de 101 à 500 agents, mais ce point sera à approfondir quand nous disposerons de plus de données ;

- Avec l'augmentation du montant de participation les écarts de taux d'adhésion entre les différentes tailles de collectivités se réduisent, et sont presque inexistants pour une participation de 20€ par mois et par agent.

B/ Caractéristiques individuelles liées à l'adhésion

Les deux caractéristiques individuelles que nous avons retenues pour étudier le choix et le délai d'adhésion sont le sexe et l'âge, qui sont à la fois les informations les plus faciles à obtenir et les premières à prendre en compte dans une analyse au niveau de l'individu.

Les agents des collectivités et les adhérents sont répartis selon les mêmes classes de sexe et d'âge, ce qui permet de comparer les deux populations.

Les délais d'adhésion ont été répartis en trois catégories : à la date d'effet de la convention, dans les six premiers mois suivant la date d'effet de la convention et au-delà des six premiers mois.

1. Choix d'adhérer ou non

Les questions que nous nous sommes posées pour déterminer s'il existait des différences individuelles entre les agents qui choisissent d'adhérer et ceux qui choisissent de ne pas adhérer sont :

- les femmes choisissent-elles plus d'adhérer que les hommes ou inversement ?

- certaines classes d'âges adhèrent-elles plus que d'autres ?

Nous avons, pour répondre à ces questions, calculé pour chaque collectivité la proportion d'agents ayant adhéré par sexe et par classe d'âge, soit en d'autres termes le taux d'adhésion par sexe et par classe d'âge.

Nous souhaitons déterminer si, au sein d'une collectivité donnée, le taux d'adhésion diffère selon deux modalités d'une même caractéristique individuelle. Dans le cas du sexe par exemple, nous voulons déterminer si au sein d'une même collectivité, les femmes adhèrent significativement (ou pas) plus ou moins que les hommes, et donc ont un taux d'adhésion significativement différent de celui des hommes.

La non prise en compte des différentes caractéristiques de la collectivité n'est pas un problème ici puisque nous cherchons simplement à déterminer s'il existe une différence entre deux taux d'adhésion au sein d'une même collectivité, nous ne cherchons ni à estimer un niveau d'adhésion ni à comparer des collectivités entre elles.

Un type de test qui peut s'appliquer à ce genre de question est un test de comparaison de deux échantillons appariés. Il reste à déterminer si nous utilisons un test paramétrique ou non paramétrique.

Dans le cas de distributions normales et d'homoscédasticité des variances nous pouvons utiliser un test paramétrique, dans le cas contraire, non.

La figure 24 présente la distribution des taux d'adhésion des deux sexes, et nous permet de constater que les distributions ne sont pas normales, donc que nous devons utiliser un test non paramétrique.

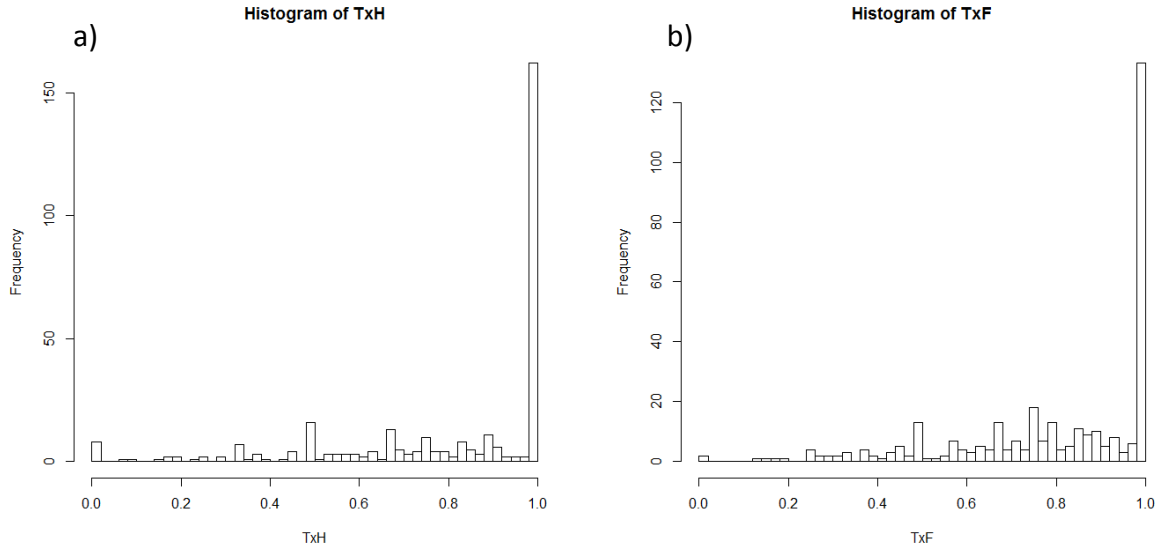


Figure 24 : Distributions de fréquence des taux d'adhésion par sexe (a : hommes et b :femmes) et par collectivité

Le test de Wilcoxon pour échantillons appariés peut être utilisé. Il s'agit d'un test de rang, dont l'hypothèse nulle H_0 est l'égalité des moyennes entre les deux groupes, qui procède de la façon suivante :

1. calcul de l'écart entre les deux valeurs appariées ;
2. classement par ordre de croissant des valeurs absolues des différences, sans tenir compte des valeurs nulles ;
3. attribution du rang dans le classement à chaque différence (ou de la moyenne des rangs en cas d'égalité de différence) ;
4. attribution de w_+ à la somme des rangs des différences strictement positives et w_- à la somme des rangs des différences strictement négatives, puis vérification que :

$$w_+ + w_- = \frac{N(N+1)}{2}$$

avec *N* le nombre de différences non nulles.

5. Soit $w = min\{w_+, w_-\}$; soit W la variable aléatoire associée à w.
   - si $N \leq 25$, on utilise la table de Wilcoxon pour la décision de rejeter H_0 ou non ;
   - si $N \geq 25$, alors sous H_0 W suit approximativement une loi normale $N(\mu, \sigma)$ avec :

$$\mu = \frac{N(N+1)}{4}, \quad \sigma = \sqrt{\frac{N(N+1)(2N+1)}{24}}.$$

Dans cette partie les analyses sont faites avec le logiciel R par simplicité d'utilisation pour le test de Wilcoxon.

La figure 25 représente le taux d'adhésion moyen par sexe : il ne semble pas y avoir de différence entre les taux d'adhésion chez les hommes et chez les femmes.

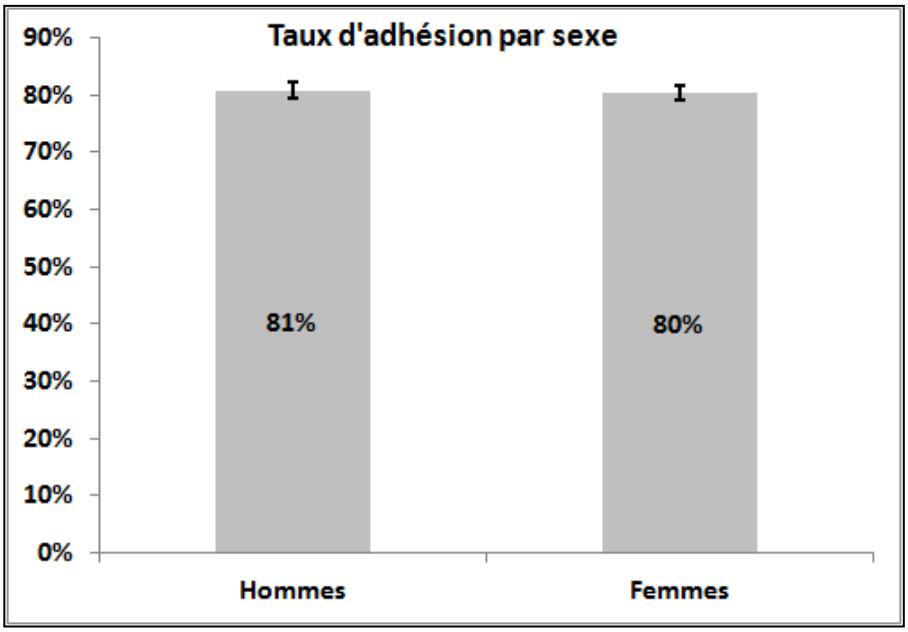


Figure 25 : Taux d'adhésion en fonction du sexe (moyenne par collectivité +/-se)

Il est possible de donner un peu de puissance au test en précisant le sens de la relation attendue. Dans le cas présent nous allons donc tester si le taux d'adhésion moyen des hommes est supérieur à celui des femmes.

Le test de Wilcoxon donne une p-value de 0.4739, donc il n'existe pas de différence significative entre le taux d'adhésion chez les hommes et le taux d'adhésion chez les femmes. Les hommes ne choisissent pas plus que les femmes d'adhérer.

Comme nous n'avons pas mis en évidence de différence entre les sexes, nous comparons les classes d'âges entre elles sans faire de distinction par sexe.

La figure 26 donne les taux d'adhésion moyens par collectivité en fonction de la classe d'âge.

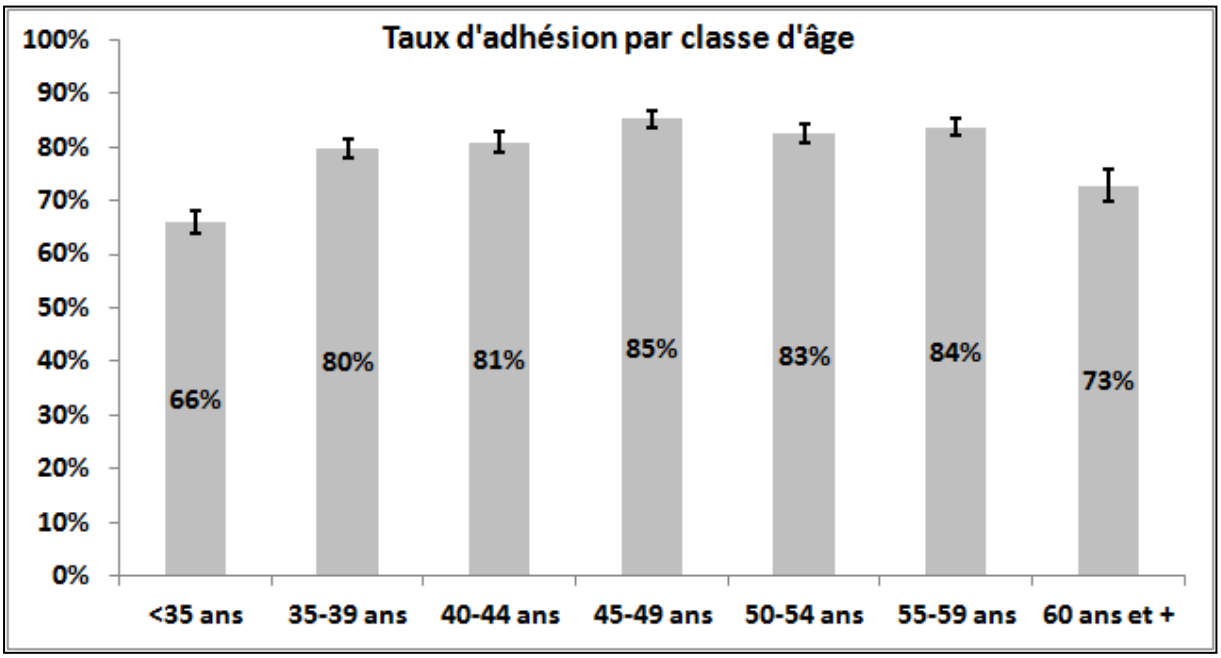


Figure 26 : Taux d'adhésion en fonction de l'âge (moyenne par collectivité +/-se)

Concernant les classes d'âge, nous les comparons deux à deux uniquement entre classes adjacentes (choix lié au caractère ordinal de la variable), en orientant également les tests. Graphiquement, les classes <35 ans et 60 ans et + semblent avoir des taux d'adhésion inférieurs aux autres.

Les résultats des tests de Wilcoxon sont présentés dans la table XV.

Table XV : Sens et résultats des tests de Wilcoxon sur les différences de taux d'adhésion entre classes d'âge

Test	p-value	Significativité
Taux <35 ans < Taux 35-39 ans	< 0.0001	***
Taux 35-39 ans < Taux 40-44 ans	0.044	*
Taux 40-44 ans < Taux 45-49 ans	0.068	NS
Taux 45-49 ans > Taux 50-54 ans	0.396	NS
Taux 50-54 ans < Taux 55-59 ans	0.602	NS
Taux 55-59 ans > Taux 60 ans et +	< 0.0001	***

Les résultats des tests indiquent que les agents de moins de 35 ans adhèrent significativement moins aux conventions de participation que les agents de 35 à 39 ans, qui

eux-mêmes adhèrent significativement moins que les agents de 40 à 44 ans. Les agents de 60 ans et plus adhèrent moins que les agents de 55 à 59 ans.

Concernant les choix d'adhésion en fonction des caractéristiques individuelles les plus accessibles, *i.e.* sexe et âge, nous pouvons retenir que :

- Aucun des deux sexes ne choisit plus d'adhérer que l'autre ;
- Les plus jeunes, *i.e.* moins de 35 ans, et dans une moindre mesure probablement (limite de significativité du test et écart moins important) les agents de 35 à 39 ans, ainsi que les plus âgés, *i.e.* 60 ans et plus, adhèrent moins que les autres aux conventions de participation.

2. Choix de la période d'adhésion

Pour analyser les liens entre les caractéristiques individuelles et la date d'adhésion, pour les agents qui ont choisit d'adhérer, nous procéderons de la même façon que pour l'étude du choix d'adhésion. Comme nous l'avons vu dans la description des données, une large majorité des agents qui choisissent d'adhérer le font à la date d'adhésion (environ les trois quarts). Il est donc inutile de comparer les taux d'adhésion entre périodes d'adhésion pour un même sexe ou une même classe d'âge. En revanche, des tendances devraient pouvoir se distinguer si nous comparons, au sein d'une même période d'adhésion, les taux d'adhésion entre sexes et entre classes d'âge.

Ainsi, nous testerons les différences de taux d'adhésion entre sexes et classes d'âges par période d'adhésion. Nous nous concentrerons sur l'adhésion à la date d'effet de la convention de participation et dans les six premiers mois car ce sont ces deux périodes qui nous intéressent le plus du fait de l'absence de sélection médicale pour les agents qui sont en activité, et parce que la dernière période représente moins de 10% des adhésions.

La figure 27 présente les moyennes par sexe, de proportions d'adhérents ayant choisi d'adhérer à la date d'effet de la convention de participation d'une part et dans les six premiers mois de la convention d'autre part.

Comme pour le choix d'adhérer ou non, il ne semble pas y avoir de différence significative entre les deux sexes quant à la période d'adhésion.

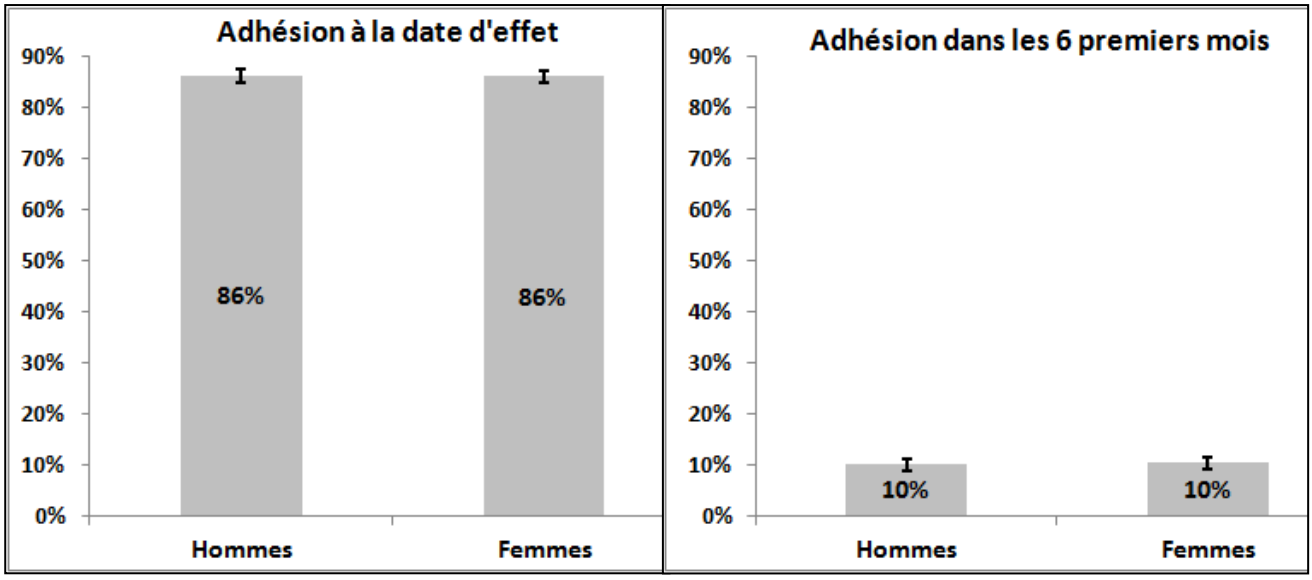


Figure 27 : Proportions, par sexe, d'adhérents à la date d'effet de la convention et dans les six premiers mois de la convention (moyennes par collectivité +/-se)

Nous utilisons de nouveau le test de Wilcoxon pour échantillons appariés en précisant le sens dans le test lorsque graphiquement nous observons un écart d'au moins 1%.

Le test n'est significatif ni pour l'adhésion à la date d'effet de la convention (p=0.494) ni pour l'adhésion dans les six premier mois (p=0.371). Il n'y a donc pas de différence significative entre les sexes pour le choix de la période d'adhésion. En conséquence, nous ne distinguerons pas les sexes dans les analyses concernant l'âge.

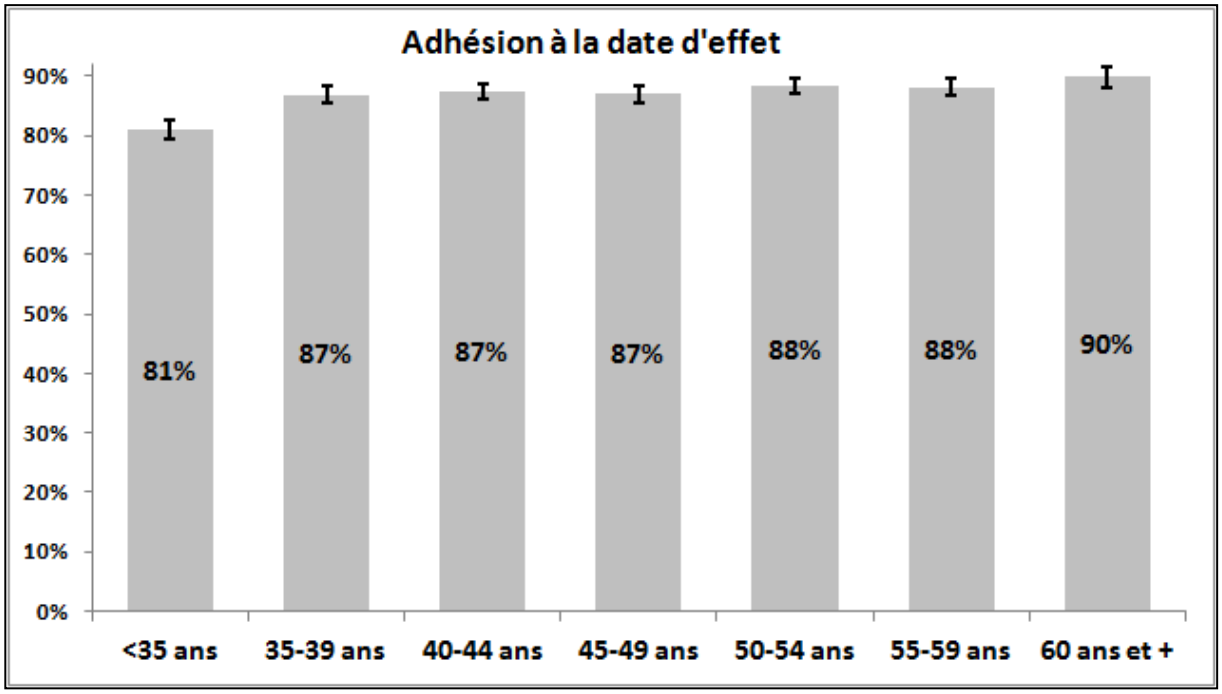


Figure 28 : Proportions, par tranche d'âge, d'adhérents ayant adhéré à la date d'effet de la convention (moyenne par collectivité +/-se)

A l'observation de la figure 28, nous pouvons faire l'hypothèse que la proportion d'adhérents de moins de 35 ans qui choisissent d'adhérer à la date d'effet de la convention plutôt qu'à une autre période est inférieure à celle des 35-39 ans et que celle des 60 ans et plus est supérieure à celle des 55-59 ans.

Les résultats des tests de Wilcoxon (table XVI) confirment ces hypothèses.

Table XVI : Résultats des tests de Wilcoxon de comparaisons entre classes d'âge des proportions d'adhérents à la date d'effet de la convention

Test	p-value	Significativité
Taux <35 ans < Taux 35-39 ans	< 0.0001	***
Taux 35-39 ans <> Taux 40-44 ans	0.377	NS
Taux 40-44 ans <> Taux 45-49 ans	0.199	NS
Taux 45-49 ans < Taux 50-54 ans	0.162	NS
Taux 50-54 ans <> Taux 55-59 ans	0.509	NS
Taux 55-59 ans < Taux 60 ans et +	0.011	*

Ainsi parmi les agents qui choisissent d'adhérer à la convention de participation, la proportion qui adhère dès la date d'effet serait plus faible chez les de moins de 35 ans que dans les âges intermédiaires et celle des 60 ans et + plus élevée

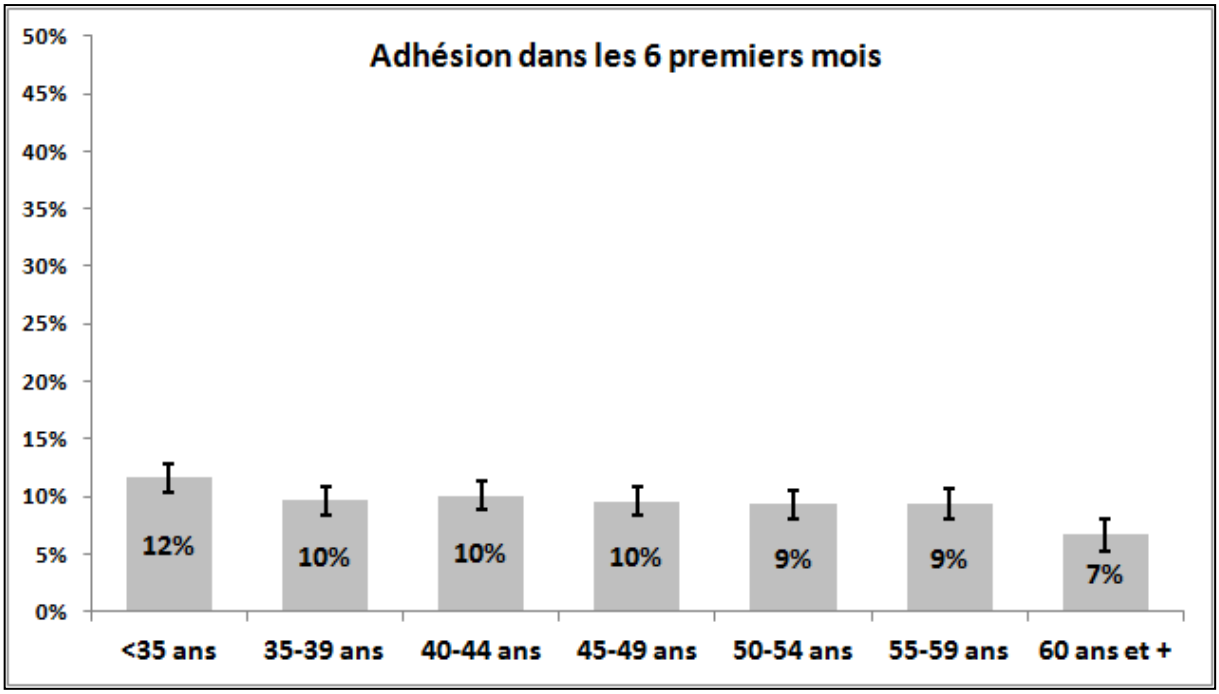


Figure 29 : Proportions, par tranche d'âge, d'adhérents ayant adhéré dans les six premiers mois de la convention (moyenne par collectivité +/-se)

Si maintenant nous nous posons les mêmes questions mais concernant les adhésions au cours des six premiers mois de la convention (figure 29), les hypothèses que nous faisons sont inverses : la proportion d'adhérents de moins de 35 ans choisissant d'adhérer dans les six premiers mois serait supérieure à celle des 35-39 ans, et la proportion de 60 ans et plus serait inférieure à celles des 55-59 ans.

Parmi les tests non orientés, celui entre la classe 40-44 ans et 45-49 ans donne un résultat significatif (p=0.01). Nous cherchons donc à connaître le sens de la différence en utilisant une décimale pour les moyennes calculées et obtenons 10,1% pour la classe 40-44 ans et 9,5% pour la classe 45-49 ans. Le test donne alors une p-value de 0.005. L'ensemble des résultats est donné dans la table XVII.

Table XVII : Résultats des tests de Wilcoxon de comparaisons entre classes d'âge des proportions d'adhérents dans les six premiers mois de la convention

Test	p-value	Significativité
Taux <35 ans > Taux 35-39 ans	0.007	**
Taux 35-39 ans <> Taux 40-44 ans	0.923	NS
Taux 40-44 ans > Taux 45-49 ans	0.005	**
Taux 45-49 ans > Taux 50-54 ans	0.385	NS
Taux 50-54 ans <> Taux 55-59 ans	0.742	NS
Taux 55-59 ans > Taux 60 ans et +	0.004	**

Les hypothèses sont confirmées : les moins de 35 ans choisissent plus que les 35-39 ans d'adhérer dans les six premiers de la convention plutôt qu'à une autre période et les 60 ans et plus choisissent moins que les 55-59 ans d'adhérer dans les six premiers mois plutôt qu'à une autre période. Une différence dans les âges intermédiaires apparaît : les 40-44 ans choisissent plus que les 45-49 ans d'adhérer dans les six premiers mois plutôt qu'à une autre période.

Ainsi, concernant les choix de la période d'adhésion :

- Nous n'avons mis en évidence aucune différence significative entre les hommes et les femmes ;
- Nous avons mis en évidence le fait que parmi les agents qui choisissent d'adhérer, même si la majorité adhère à la date d'effet de la convention, les plus jeunes tendraient plus à adhérer dans les six premiers mois que les autres et les plus âgés tendraient plus à adhérer dès la date d'effet que les autres.

C/ synthèse générale concernant l'adhésion

Dans cette partie, nous souhaitions déterminer d'une part si le caractère facultatif sans sélection médicale des conventions de participation mises en place dans le cadre du décret du 08 novembre 2011 entraînait des disparités d'adhésion entre collectivités et entre individus.

Nous avons pu mettre en évidence des différences significatives, à la fois entre collectivités et entre individus. Nous pouvons donc conclure que le caractère facultatif des conventions de participation entraîne des disparités, que ce soit à l'échelle de la collectivité ou à l'échelle individuelle. Concernant l'absence de sélection médicale, ces premiers résultats ne nous permettent pas de poser une hypothèse forte, nous devrons les mettre en relation avec la sinistralité.

D'autre part nous souhaitions qualifier ces disparités. Il ressort que :

- Les collectivités du type commune, communauté de commune et communauté d'agglomération ont des taux d'adhésion significativement plus élevés que les autres types ;
- Le taux d'adhésion est lié à la situation sociale du territoire où se trouve la collectivité : en effet il diminue quand la proportion d'HLM dans les habitations principales augmente ;
- Le montant de participation a très probablement un effet sur le taux d'adhésion. Même si les analyses statistiques ne permettent pas de mettre en évidence des relations de cause à effet, dans la mesure où dans la plupart des collectivités le taux d'adhésion augmente avec le montant de participation, il est raisonnable de penser qu'un montant de participation plus élevé incite à adhérer, avec un effet plus fort dans les très petites et les très grandes collectivités ;
- Le taux d'adhésion varie avec la taille de la collectivité, tendant à être plus faible dans les grandes collectivités, mais les écarts se réduisent avec l'augmentation du montant de participation, s'annulant quasiment quand il atteint 20 € mensuels par agent, ce montant couvrant très souvent la totalité ou presque du montant de la cotisation d'incapacité. Ce résultat va également dans le sens d'un effet du montant de la participation sur le taux d'adhésion.
- Au niveau individuel, aucune différence significative n'a été mise en évidence entre les hommes et les femmes : ni sur le choix d'adhérer ou non, ni sur la période d'adhésion ;
- En revanche, les plus jeunes agents (moins de 35 ans) et les plus âgés (60 ans et plus) tendent à moins adhérer que les âges intermédiaires, et parmi ceux qui

choisissent d'adhérer, les plus jeunes tendraient à adhérer plus dans les six premiers mois qu'à la date d'effet par rapport aux autres, tandis que l'inverse est observé pour les 60 ans et plus. Ces observations pourraient s'expliquer par le fait que les agents de 60 ans et plus, dans la mesure où il sont près de l'âge de retraite, ne considèrent pas avantageux pour eux d'adhérer, mais ceux qui adhèrent le font sans délai. Concernant les moins de 35 ans, la prévoyance peut ne pas entrer dans leurs préoccupations principales, et une partie de ceux qui choisissent d'adhérer le fait après un délai de réflexion supplémentaire.

Partie IV : Etude de la sinistralité

Pour déterminer si l'absence de sélection médicale pendant un délai allongé peut avoir un impact sur le profil de risque des agents qui choisissent d'adhérer et donc sur la mutualisation des risques, il est nécessaire de relier les caractéristiques individuelles des agents qui adhèrent aux caractéristiques individuelles des adhérents les plus indemnisés.

Nous ne pourrons bien entendu pas répondre totalement à cette question car nous avons trop peu de recul pour un risque long tel que l'incapacité (et *a fortiori* l'invalidité et la perte de retraite) des territoriaux (incapacité qui, pour rappel, peut durer jusqu'à 8 ans en cas d'arrêt continu, et plus encore en cas d'arrêt discontinu). Nous pourrons néanmoins dégager quelques tendances concernant les adhérents potentiellement les plus indemnisés, et comparer leur profil à celui des agents qui choisissent plus ou moins que d'autres, et plus ou moins tôt, d'adhérer aux conventions de participation.

Nous nous posons les questions suivantes :

- Certaines caractéristiques de sexe, d'âge et la date d'adhésion ont-elles un lien avec une incidence plus précoce de l'incapacité indemnisée ? En d'autres termes, certains adhérents sont-ils en congé maladie indemnisé plus tôt que d'autres selon leur sexe, leur âge et/ou le moment où ils choisissent d'adhérer ?

- Certaines caractéristiques de sexe, d'âge et la date d'adhésion ont-elles un lien avec le maintien en incapacité indemnisée ? En d'autres termes, certains adhérents restent-ils en congé maladie indemnisé plus longtemps que d'autres selon leur sexe, leur âge et/ou le moment où ils choisissent d'adhérer ?

NB : Ne sont pris en compte que l'incidence et le maintien en incapacité indemnisée, soit les périodes où les adhérents sont à demi-traitement, car il n'est pas possible d'avoir connaissance avec certitude et précision des périodes d'arrêt de travail non indemnisées.

A/ Méthode utilisée

La même méthode sera utilisée pour les deux types de questions. En effet, dans les deux cas nous nous intéressons à une durée : la durée entre l'adhésion et le premier congé maladie indemnisé d'une part, et la durée de ce congé maladie indemnisé d'autre part.

En d'autres termes, nous cherchons à estimer la probabilité de survenue à plusieurs instants d'un événement : le congé maladie indemnisé dans un cas, et la reprise du travail dans l'autre cas.

De plus, nous sommes en présence de données censurées : l'événement auquel nous nous intéressons peut ne pas s'être encore réalisé au moment où nous cessons d'observer l'individu (dans notre cas il s'agit soit de la date d'extraction des données, soit de la date de résiliation de l'adhérent), et nous ne savons pas quand il se réalisera.

Nous pouvons donc utiliser une méthode de la famille des modèles de durée.

Nous souhaitons étudier la probabilité de survenue de l'événement en fonction de variables explicatives qui sont le sexe, l'âge et la date d'adhésion : la question qui nous intéresse est de mettre en évidence des différences significatives selon ces variables, mais pas de connaître précisément la fonction de risque (='hazard' en anglais, donc appelée aussi fonction de hasard). Un type de modèle que nous pouvons utiliser est donc le **modèle de Cox** (1972).

Il s'agit d'un modèle de survie **semi-paramétrique à risques proportionnels** (Timsit *et al.*, 2005) : il n'est pas nécessaire de faire une hypothèse quant à la forme de la fonction de survie de base, et on fait l'hypothèse que la fonction de risque du phénomène observé est proportionnelle à la fonction de risque de base[7].

Le modèle peut ainsi s'écrire :

$$h\big(t|\boldsymbol{Z}_i(t)\big) = h_0(t)\exp(\boldsymbol{Z}_i(\mathrm{t})\beta')$$

Avec :

- $h_0(t)$ le risque de base, non spécifié, correspondant à l'intensité en l'absence des variables explicatives ;
- $\boldsymbol{Z}_i$ le vecteur des variables explicatives pour un individu i ;
- β le vecteur des coefficients de la régression.

Pour pouvoir appliquer ce modèle, l'hypothèse à vérifier est que pour deux individus (ou populations d'individus), les risques sont proportionnels et donc que le rapport des risques (HR= 'Hazard Ratio') est constant au cours du temps.

Pour deux individus i et j on a ainsi :

$$HR = \frac{h(t|\boldsymbol{Z}_i)}{h\big(t|\boldsymbol{Z}_j\big)} = \frac{\exp\big(\boldsymbol{Z}_\mathrm{i}\beta'\big)}{\exp\big(\boldsymbol{Z}_\mathrm{j}\beta'\big)} = exp\big[\big(\boldsymbol{Z}_\mathrm{i} - \boldsymbol{Z}_\mathrm{j}\big)\beta'\big]$$

Un test qui peut être utilisé pour vérifier cette hypothèse est le test des résidus de Schoenfeld (1982). Ce test est à appliquer sur le modèle contenant l'ensemble des variables explicatives testées (et leurs interactions le cas échéant).

[7] Si $S(t)$ est la fonction de survie et $h(t)$ la fonction de risque, alors par définition : $h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} lnS(t)$

Le principe est de calculer, pour chaque date de réalisation d'un événement, la différence entre les caractéristiques de l'individu pour lequel l'événement s'est réalisé et une moyenne pondérée de l'ensemble des caractéristiques des individus exposés à la date *t*.

Soit $U_i(t)$ le vecteur des résidus de Schoenfeld pour l'individu i au temps *t*, alors :

$$U_i(t) = z_i(t) - \tilde{z}_t(t)$$

Avec :

- $z_i(t)$ le vecteur des valeurs des explicatives pour l'individu i au temps t ;
- $\tilde{z}_t(t) = \frac{\sum_{j \in R_t} z_j(t) exp[\hat{\beta}' z_j(t)]}{\sum_{j \in R_t} exp[\hat{\beta}' z_j(t)]}$, le vecteur des moyennes pondérées des explicatives calculées sur tous les individus exposés au temps *t*, avec R_t l'ensemble des individus exposés à t^- (juste avant *t*).

Sous l'hypothèse de PH ('Proportional Hazard'), ces résidus, standardisés, ne présentent aucune tendance au cours du temps, ce qui peut être vérifié soit par représentation graphique soit par un test.

Les méthodes utilisées pour la validation du modèle et la significativité des paramètres sont les mêmes que pour la régression logistique : maximisation de la vraisemblance et test de Wald, à la différence près que le modèle de Cox utilise une vraisemblance partielle. La vraisemblance (partielle) totale s'écrit :

$$L_{Cox}(\beta) = \prod_i \frac{exp(\beta' Z_i)}{\sum_{j \in R_t} exp(\beta' Z_j)}$$

avec $\frac{exp(\beta' Z_i)}{\sum_{j \in R_t} exp(\beta' Z_j)}$ la probabilité que l'individu i subisse l'événement en t sachant qu'un événement a eu lieu en *t*.

Les censures ne participent pas au calcul de la vraisemblance.

B/ Analyse de l'incidence de l'incapacité

Dans cette partie nous souhaitons tester si la durée entre l'adhésion et le premier congé maladie indemnisé varie en fonction :

- du sexe de l'adhérent

- de son âge au moment de l'adhésion (nous utiliserons les mêmes classes d'âge que pour l'analyse de l'adhésion)

- de sa date d'adhésion (nous utiliserons les mêmes périodes que pour l'analyse de l'adhésion)

Nous utiliserons donc la durée, en jours, entre la date d'adhésion et le premier jour de congé maladie indemnisé, la variable de censure étant la survenue ou non d'un congé maladie indemnisé.

Le logiciel utilisé pour la validation de l'hypothèse de risques proportionnels (PH) sera le logiciel R et les modèles de Cox seront ajustés avec SAS.

1. *Sélection du modèle*

Pour déterminer le modèle à retenir, nous procédons de la même façon que pour l'analyse du taux d'adhésion : par une sélection *backward*.

Nous ajustons donc le modèle complet qui comporte les trois variables explicatives que nous souhaitons tester et les interactions entre ces variables deux à deux.

Le test du rapport de vraisemblance permet de valider le modèle (table XVIII) et de tester l'hypothèse de PH.

Table XVIII : Test de validation du modèle complet

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > Khi-2**
Likelihood Ratio	133.5049	29	<.0001

Le test de l'hypothèse de PH sous R teste l'hypothèse nulle $H_0 : \beta(t) = \beta$ contre l'hypothèse alternative $H_1 : \beta(t) \neq \beta$ (R calcule la corrélation entre les résidus et une fonction du temps basée sur l'estimateur de Kaplan Meier).

La p-value de 0.327 permet de ne pas rejeter H_0 et donc de considérer que l'effet des variables ne dépend pas du temps.

La table XIX donne les résultats de l'analyse de type 3 et permet de constater que toutes les variables n'ont pas un effet significatif.

Table XIX : Tests de significativités des paramètres du modèle complet pour l'incidence de l'incapacité

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
Sexe	1	12.4668	0.0004
clage	6	10.1463	0.1186
Sexe*clage	6	25.4821	0.0003
cldadh	2	1.1804	0.5542
Sexe*cldadh	2	3.1666	0.2053
clage*cldadh	12	7.5588	0.8186

La notation des variables est la suivante et sera maintenue pour le reste du document :

- 'Sexe' : sexe de l'adhérent

- 'clage' : classe d'âge à l'adhésion à laquelle appartient l'agent (pour rappel, les classes, sont : <35, 35-39, 40-44, 45-49, 50-54, 55-59, 60+)

- 'cldadh' : période d'adhésion (à la date d'effet, dans les six premiers mois ou au-delà des six premiers mois de la convention).

La p-value la plus élevée (0,8186) est pour l'interaction entre l'âge à l'adhésion et la date d'adhésion. Nous supprimons donc cette interaction et ajustons le modèle Ci1, dont les résultats sont présentés dans la table XX.

Table XX : Tests de validation et de significativité des paramètres du modèle Ci1

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > Khi-2**
Likelihood Ratio	125.6081	17	<.0001

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
Sexe	1	14.2345	0.0002
clage	6	59.1100	<.0001
Sexe*clage	6	26.3483	0.0002
cldadh	2	7.3176	0.0258
Sexe*cldadh	2	3.9638	0.1378

L'interaction entre le sexe et la date d'adhésion n'est pas significative (p-value=0,1378), donc nous supprimons cette interaction et ajustons un modèle Ci2 dont les résultats sont présentés dans la table XXI.

Table XXI : Tests de validation et de significativité des paramètres du modèle Ci2

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > Khi-2**
Likelihood Ratio	121.4630	15	<.0001

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
Sexe	1	23.3796	<.0001
clage	6	58.3173	<.0001
Sexe*clage	6	26.2338	0.0002
cldadh	2	4.4324	0.1090

Dans le modèle Ci2, la date d'adhésion a une p-value de 0,109, indiquant un effet non significatif. Nous supprimons donc cette variable et ajustons un modèle Ci3.

Le test de validation de ce modèle est significatif (p<0.0001) et toutes les variables ont un effet significatif (table XXII).

Table XXII : Tests de validation et de significativité des paramètres du modèle Ci3

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
Sexe	1	23.5058	<.0001
clage	6	57.3537	<.0001
Sexe*clage	6	26.0330	0.0002

Nous vérifions que l'hypothèse de PH est toujours valable : la p-value est égale à 0.386, donc nous pouvons conserver ce modèle.

Il existe donc une relation significative entre l'incidence de l'incapacité indemnisée et le sexe, l'âge à l'adhésion ainsi que l'interaction entre les deux.

2. Analyse des résultats

Dans cette partie nous qualifierons le sens des effets des variables à l'aide des ratios de risques (=HR). Comme pour l'analyse du taux d'adhésion, nous n'utiliserons ces ratios que pour donner un sens à la relation, et attendrons de disposer de plus de données pour la quantifier.

La figure 30 nous permet d'étudier le HR entre femmes et hommes par âge à l'adhésion : un HR supérieur à 1 indique que les femmes sont en congé maladie indemnisé plus tôt que les hommes. Un HR inférieur à 1 indique l'inverse.

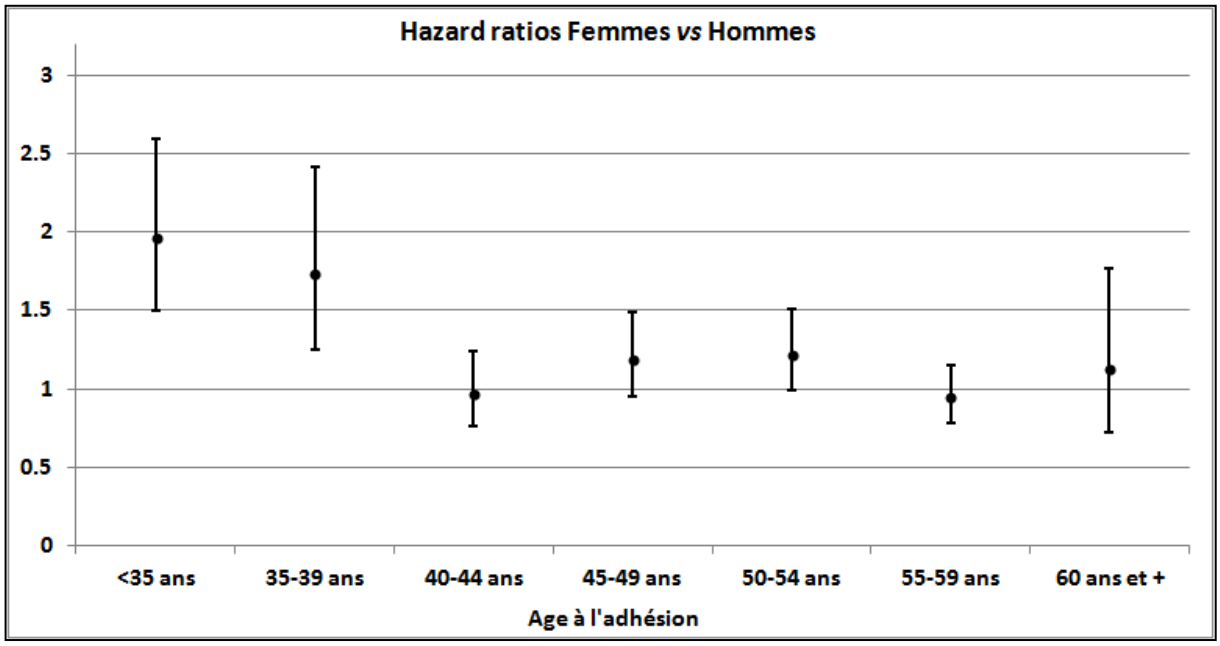


Figure 30 : Hazard-ratios et intervalles de confiance à 95% entre femmes et hommes par âge à l'adhésion

Nous constatons que les HR des adhérents de moins de 40 ans sont très différents de 1, indiquant que les femmes ayant adhéré avant 40 ans sont en congé maladie indemnisé plus tôt après leur adhésion que les hommes du même âge. Entre 45 et 54 ans, la borne inférieure de l'intervalle de confiance est très proche de 1, donc il est possible que les femmes ayant adhéré entre 45 et 54 ans soient en arrêt plus tôt que les hommes du même âge. En revanche, il n'existe pas de différence significative entre les deux sexes pour les adhésions entre 40 et 44 ans et à partir de 55 ans.

Nous nous intéressons maintenant aux HR entre classes d'âge à l'adhésion. Dans la mesure où l'interaction entre sexe et âge à l'adhésion est significative, nous dissocions l'étude des HR par sexe.

Sur les figures nous ne présentons que les HR entre deux classes contiguës pour faciliter la lecture et l'interprétation des résultats.

Chez les femmes (fig. 31), on constate que celles ayant adhéré avant 35 ans sont en arrêt plus tôt que celles ayant adhéré entre 35 et 39 ans.

Nous observons également une différence significative entre les classes 40-44 et 45-49, les femmes ayant adhéré entre 40 et 44 ans étant en congé maladie indemnisé plus tard que celles ayant adhéré entre 45 et 49 ans.

Aucun autre HR n'est significativement différent de 1. Nous pouvons donc regrouper des classes pour faciliter l'interprétation des résultats.

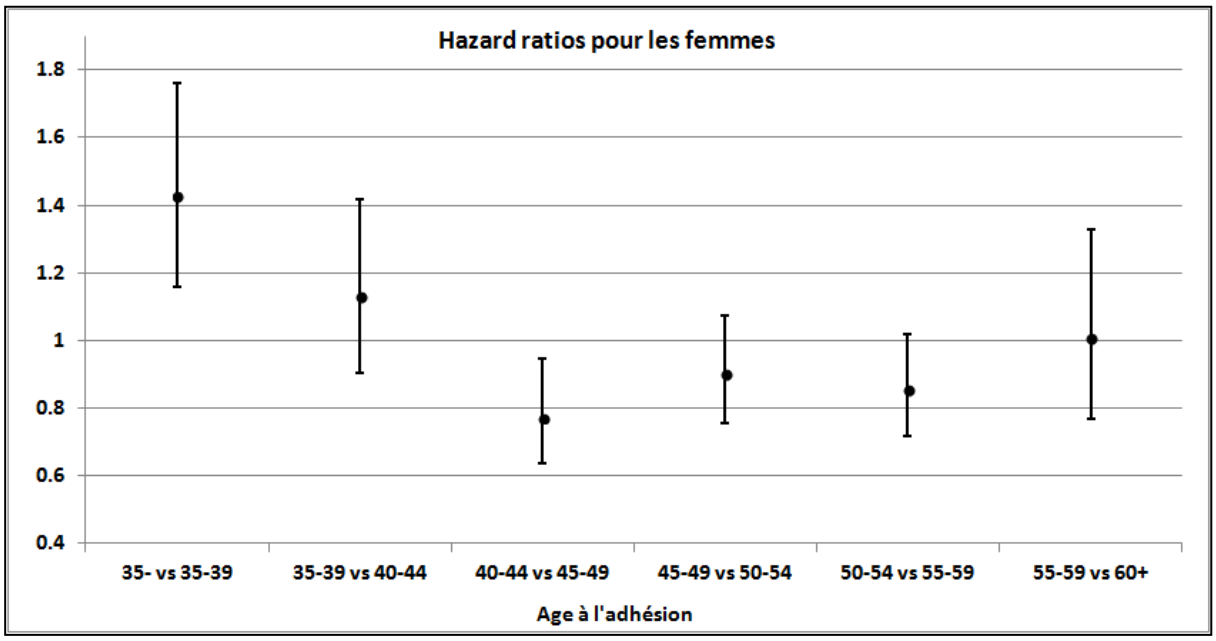


Figure 31 : Hazard ratios et intervalles de confiance à 95% entre classes d'âge à l'adhésion pour les femmes

Il y aurait donc :

- les femmes ayant adhéré avant 35 ans ;

- les femmes ayant adhéré entre 35 et 44 ans ;

- les femmes ayant adhéré à partir de 45 ans.

Cependant, la borne supérieure de l'intervalle de confiance du HR '50-54 vs 55-59' est très proche de 1, et si l'on s'intéresse au HR '45-49 vs 55-59', on constate que la borne supérieure de son intervalle de confiance est au-dessous de 1.

Nous préférons donc conserver un groupe supplémentaire et répartir les femmes de la façon suivante :

- les femmes ayant adhéré avant 35 ans, qui ont une incidence de l'incapacité indemnisée plus précoce que celles ayant adhéré entre 35 et 44 ans ;

- les femmes ayant adhéré entre 35 et 44 ans ;

- les femmes ayant adhéré entre 45 et 54 ans, qui ont une incidence de l'incapacité indemnisée plus précoce que celles ayant adhéré entre 35 et 44 ans ;

- les femmes ayant adhéré à partir de 55 ans, qui ont une incidence de l'incapacité indemnisée plus précoce que celles ayant adhéré entre 45 et 54 ans.

Nous procédons de la même façon avec les HR des hommes entre classes d'âge à l'adhésion.

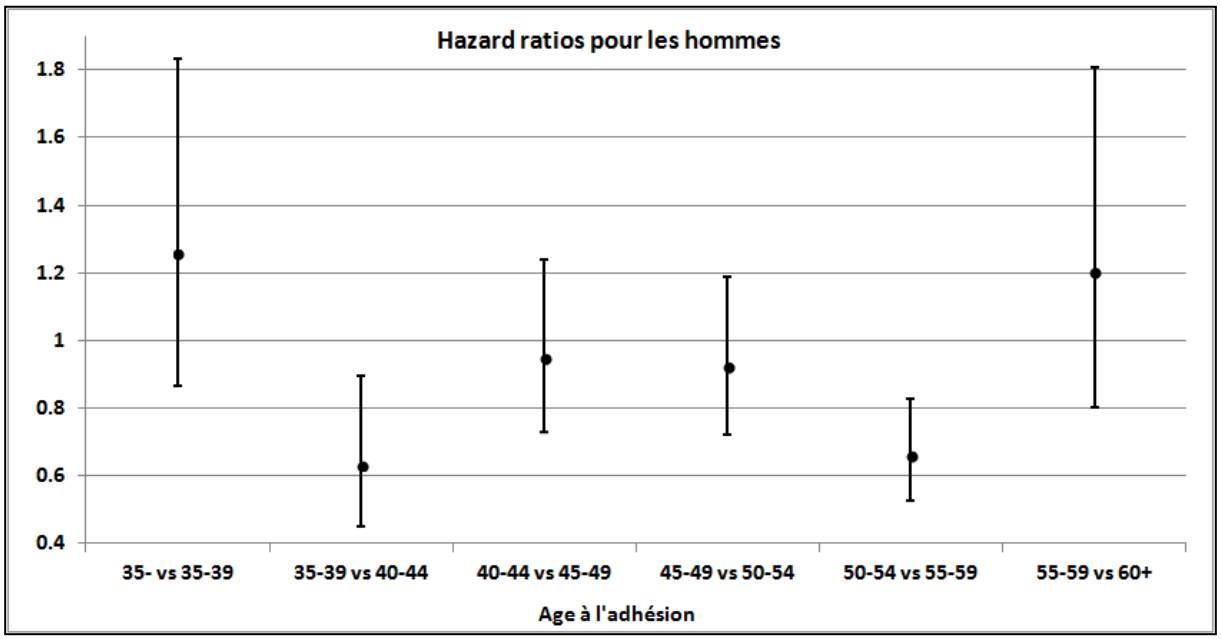


Figure 32 : Hazard-ratios et intervalles de confiance à 95% entre classes d'âge à l'adhésion pour les hommes

La figure 32 met en évidence le fait que les hommes ayant adhéré avant 40 ans sont en congé maladie indemnisé plus tard après leur adhésion que les hommes ayant adhéré entre 40 et 44 ans. Les hommes ayant adhéré entre 50 et 54 ans sont en congé maladie indemnisé plus tard que ceux ayant adhéré à partir de 55 ans.

Nous pouvons donc constituer trois groupes pour les hommes :

- les hommes ayant adhéré avant 40 ans, qui sont en arrêt plus tard après leur adhésion que ceux ayant adhéré entre 40 et 54 ans ;

- les hommes ayant adhéré entre 40 et 54 ans, qui sont en arrêt plus tard après leur adhésion que ceux ayant adhéré à partir de 55 ans.

3. Synthèse sur l'incidence de l'incapacité indemnisée

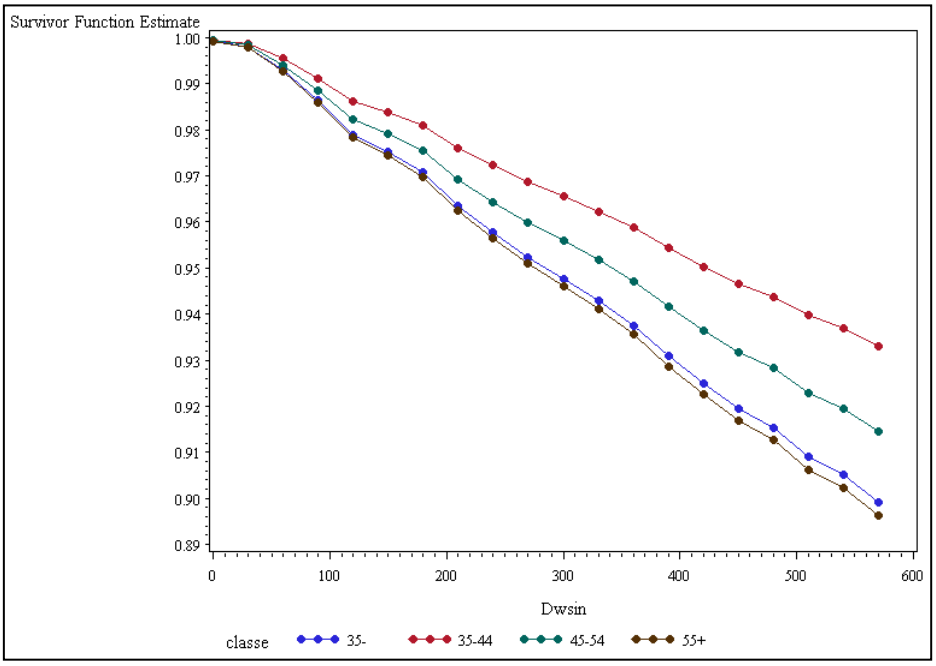
Concernant l'incidence de l'incapacité indemnisée, nous pouvons retenir que :

- Nous n'avons pas mis en évidence de relation significative avec la date d'adhésion ;
- Nous avons mis en évidence des différences significatives entre les caractéristiques individuelles étudiées : le sexe et l'âge à l'adhésion ;
- La différence entre les deux sexes semble plutôt concerner les moins âgés, les femmes ayant une incapacité indemnisée plus précoce que les hommes pour les âges d'adhésion inférieurs à 40 ans ;

- Globalement, la précocité de l'incapacité indemnisée augmente avec l'âge à l'adhésion, sauf pour les femmes de moins de 35 ans qui sont en congé maladie indemnisé plus tôt que les autres après leur adhésion.

La figure 33 illustre cela par les fonctions de survie (l'axe des abscisses correspondant ici au nombre de jours depuis l'adhésion) prédites par le modèle pour chacun des sexes par groupe d'âge retenu.

a) Femmes



b) Hommes

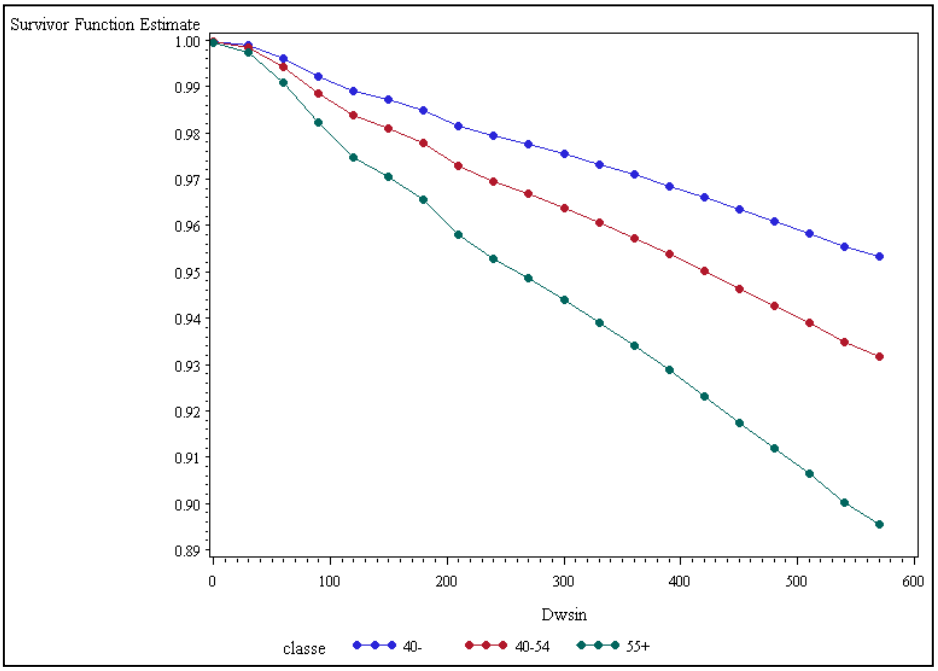


Figure 33 : Fonctions de survie par sexe et groupe d'âge à l'adhésion

C/ Analyse du début du maintien en incapacité indemnisée

Dans cette partie nous nous intéressons à la durée en incapacité indemnisée, la variable de censure étant le fait d'avoir repris le travail ou non.

Dans la mesure où parmi les 2 025 congés maladie indemnisés au cours de la période d'observation, 1 522, soit 75%, se sont terminés pendant la période d'observation, il n'est pas inutile de commencer à s'intéresser à cette question, même si bien entendu nous ne couvrons principalement qu'un peu plus de la première année de l'incapacité indemnisée.

Nous souhaitons tester les mêmes variables que pour l'incidence de l'incapacité indemnisée et procéderons de la même façon.

1. Sélection du modèle

Le modèle complet a un test de validation significatif (p<0.0001) et respecte l'hypothèse de PH (p=0.669).

Table XXIII : Tests de significativité des paramètres du modèle complet pour le maintien en incapacité indemnisée

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
Sexe	1	1.0508	0.3053
clage	6	6.5681	0.3626
Sexe*clage	6	9.0578	0.1704
cldadh	2	1.9441	0.3783
Sexe*cldadh	2	1.2048	0.5475
clage*cldadh	12	9.3628	0.6717

Les résultats des tests de significativité (table XXIII) indiquent qu'il faut supprimer des variables. Nous utilisons donc une méthode de sélection *backward* comme précédemment dans ce document.

Nous ne détaillons pas tous les résultats intermédiaires car le raisonnement est exactement le même.

Arrivés au même modèle que pour l'incidence de l'incapacité indemnisée, nous constatons que l'interaction entre sexe et âge à l'adhésion n'est pas significative (p=0.203).

Nous la supprimons donc et le modèle obtenu indique que la variable sexe seule n'est pas significative non plus (p=0.896). Nous la supprimons également et obtenons un modèle avec uniquement l'âge à l'adhésion (table XXIV).

Table XXIV : Résultats des tests de validation et de significativité de paramètres du modèle retenu pour le maintien en incapacité indemnisée

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > Khi-2**
Likelihood Ratio	45.6114	6	<.0001

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > Khi-2**
clage	6	47.9978	<.0001

L'hypothèse de PH est toujours vérifiée (p=0.274) donc nous pouvons conserver le modèle. Pour la période d'observation dont nous disposions, nous avons mis en évidence une relation significative entre le début du maintien en incapacité indemnisée et l'âge à l'adhésion.

2. *Analyse des résultats*

L'analyse des HR entre classes contigües (fig. 34) montre que la seule différence significative est entre les classes <35 ans et 35-39, les adhérents ayant adhéré plus jeunes restant moins longtemps en congé maladie indemnisé que ceux ayant adhéré plus âgés.

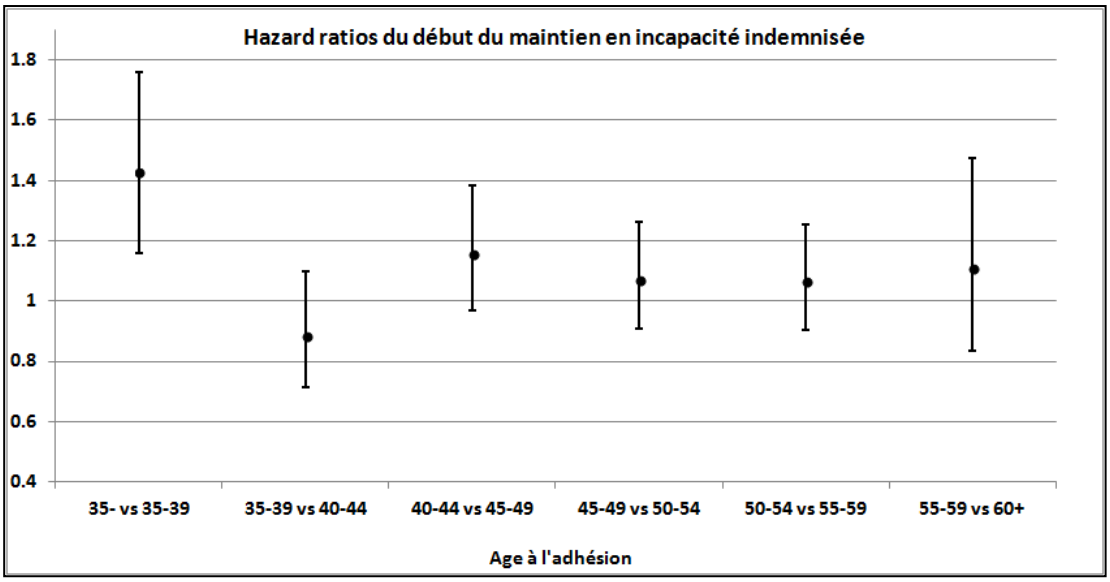


Figure 34 : Hazard-ratios entre classes d'âge à l'adhésion, de la durée des congés maladie indemnisés sur la période d'observation

Nous n'observons pas d'autres différences significatives entres classes d'âge contigües, mais si nous nous intéressons aux HR entre toutes les classes deux à deux (fig. 35), nous constatons que pour chaque classe d'âge à l'adhésion, son HR avec les autres classes augmente avec l'écart d'âge à l'adhésion.

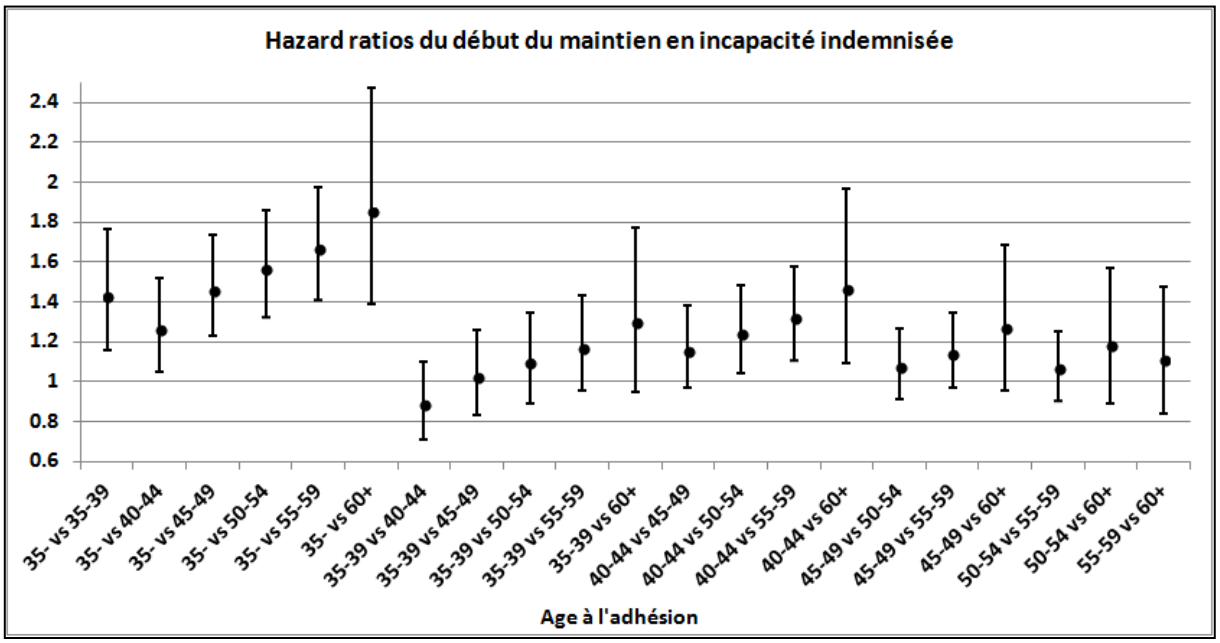


Figure 35 : Hazard-ratios entre toutes les classes d'âge à l'adhésion

Cette dernière observation laisserait supposer que la durée en incapacité indemnisée augmente avec l'âge à l'adhésion.

3. Synthèse concernant le début du maintien en incapacité

Concernant le maintien en incapacité indemnisée au cours de notre période d'observation, nous pouvons retenir que :

- Nous n'avons mis en évidence aucune relation significative entre le maintien en incapacité indemnisée au cours de la période d'observation et la date d'adhésion ;
- Nous n'avons mis en évidence aucune relation significative entre le maintien en incapacité indemnisée au cours de la période d'observation et le sexe de l'adhérent ;
- Il est possible que le maintien en incapacité indemnisée augmente avec l'âge à l'adhésion.

D/ Synthèse concernant la sinistralité

Nous souhaitions déterminer dans cette partie si la sinistralité était liée à certaines caractéristiques individuelles, pour ensuite mettre en relation ces résultats avec les caractéristiques individuelles à l'adhésion.

Nous pouvons retenir que :

- Aucun lien n'a été mis en évidence entre la date d'adhésion et le niveau de sinistralité, que ce soit pour l'incidence de l'incapacité indemnisée ou le début de son maintien ;
- La seule différence entre sexes mise en évidence concerne l'incidence de l'incapacité indemnisée chez les plus jeunes, les femmes ayant adhéré avant 40 ans étant en arrêt plus tôt après leur adhésion que les hommes du même âge ;
- Les congés maladies indemnisés tendraient à arriver plus tôt après l'adhésion et à durer plus longtemps avec l'augmentation de l'âge, la seule exception étant les femmes ayant adhéré avant 35 ans qui seraient en congé maladie plus tôt après leur adhésion que celles de 35-39 ans.

Conclusion

A/ Synthèse générale

Nous nous sommes intéressés dans cette étude à l'impact que pourrait avoir le caractère facultatif sans sélection médicale des conventions de participation, sur l'adhésion d'une part et sur la sinistralité d'autre part. En effet, l'inconvénient de ce type de contrat collectif est qu'il peut y avoir une mauvaise mutualisation des risques et donc un déséquilibre des résultats par rapport au tarif, calculé pour l'ensemble de la population susceptible d'adhérer.

Nous avons dans un premier temps étudié les taux d'adhésion en relation avec différentes caractéristiques des collectivités. Plus le taux d'adhésion est élevé, meilleure est la mutualisation des risques. Les taux d'adhésion du portefeuille étudié sont globalement corrects puisqu'en moyenne ils dépassent 50%. Mais il existe d'importantes disparités entre collectivités selon leur type, leur taille, le montant de participation dont bénéficient leurs agents et la proportion d'HLM dans les habitations principales de leur commune de rattachement.

Le taux d'adhésion :

- Tend à augmenter avec le montant de participation, de façon plus prononcée dans les très petites et les très grandes collectivités. L'écart entre les différentes tailles de collectivité se réduit avec l'augmentation du montant de participation, devenant presque nul à partir de 20€ par agent et par mois. Cette observation peut s'expliquer par le fait qu'un tel montant couvre fréquemment la totalité, ou presque, de la cotisation de la garantie incapacité.
- Est plus élevé dans les communes et communautés de communes et d'agglomération que dans les autres types de collectivités, sans relation avec le montant de participation.
- Tend à diminuer quand la proportion d'HLM dans les habitations principales augmente, confirmant l'hypothèse de départ de l'existence d'un lien entre la situation sociale de l'environnement de la collectivité et la propension à souscrire à un contrat de prévoyance complémentaire.

Ainsi, les collectivités présentant les plus forts taux d'adhésion sont les communes et les communautés de communes, plutôt de petite taille, avec un montant élevé de participation.

Dans un second temps nous nous sommes intéressés aux liens qui pouvaient exister entre l'adhésion et deux caractéristiques individuelles majeures : le sexe et l'âge. L'idée étant de

pouvoir ensuite mettre en relation ces résultats avec les résultats de l'analyse de la sinistralité (qui dans le cadre de cette étude concerne l'incidence, au cours de la période d'observation, de l'incapacité indemnisée et le début de son maintien).

Concernant l'adhésion :	Concernant la sinistralité :
- Nous n'avons pas mis en évidence de différence entre les sexes.	- Aux âges d'adhésion les plus jeunes (*i.e.* avant 40 ans), les femmes sont en congé maladie indemnisé plus tôt que les hommes.
- Les plus jeunes (<35 ans) et les plus âgés (60 ans et+) adhèrent moins que les âges intermédiaires.	- A part pour les femmes ayant adhéré avant 35 ans, dont l'incapacité indemnisée est plus précoce que pour celles ayant adhéré entre 35 et 39 ans, l'incapacité indemnisée est d'autant plus précoce et dure d'autant plus longtemps que l'âge à l'adhésion augmente.

La mise en parallèle des résultats concernant les caractéristiques individuelles liées à l'adhésion et à la sinistralité ne permet pas de mettre en évidence un lien entre le caractère facultatif sans sélection des conventions de participation et un niveau de sinistralité plus élevé. Néanmoins, la classe atypique des femmes adhérant avant 35 ans méritera d'être analysée plus en détail. Il s'agira de déterminer si la précocité de l'incidence de l'incapacité indemnisée est liée au sexe et à l'âge, ou si les femmes de moins de 35 ans qui adhèrent sont celles pour lesquelles le niveau de sinistralité est plus élevé.

Ainsi, il peut exister des problèmes de mutualisation des risques dans les conventions de participation. Mais d'après les premières tendances, l'origine serait plutôt au niveau des taux d'adhésion que des caractéristiques individuelles des agents qui choisissent d'adhérer. Cependant, comme nous l'avons précisé dès le début de l'étude, il ne peut s'agir que d'hypothèses, qui ne pourront être confirmées ou infirmées qu'avec beaucoup plus d'années de recul étant donnée la longueur du risque concerné.

B/ Perspectives

Dès que nous aurons plus de données, nous devrons valider, ou invalider, les différents modèles retenus dans les analyses de cette étude.

Nous devrons également trouver une méthode plus globale que comparer des classes deux à deux pour analyser les écarts entre la population adhérente et la population d'origine si nous voulons utiliser les résultats pour affiner les méthodes de tarification. Un test du Khi-deux permettrait de déterminer si la population adhérente est significativement différente de la population d'origine, mais elle ne permettrait pas de quantifier cet écart. De plus, avec une grande quantité de données, le test serait presque systématiquement significatif.

Une autre méthode que nous envisageons s'appuierait sur le coefficient de variation. En effet, si l'on répartit la population d'origine et la population adhérente par sexe et classe d'âge à l'adhésion pour chacune des collectivités, on peut calculer le taux d'adhésion de chaque classe obtenue par collectivité. Si dans une collectivité donnée la structure de la population adhérente est strictement identique à celle de la population d'origine, l'écart-type des taux d'adhésion doit être nul. Plus la population adhérente s'éloigne de la population d'origine, plus l'écart-type augmente. Pour standardiser l'analyse et limiter la dépendance à la valeur du taux d'adhésion, on utilise le coefficient de variation (écart-type/moyenne), que l'on utilise comme variable expliquée dans un modèle de régression. Le problème est que le coefficient de variation reste très corrélé au taux d'adhésion, mais nous approfondirons néanmoins cette piste.

Les résultats de cette étude sont cependant déjà utilisables tels quels pour différentes applications et confirment un certain nombre d'observations faites sur le terrain :

- Ils nous montrent que nous devons progresser en termes de communication et de persuasion chez les jeunes agents pour améliorer leurs taux d'adhésion ;

- Ils nous permettent de déterminer en amont, pour les nouvelles conventions, les cas où il sera nécessaire de renforcer la campagne de mise en place et éventuellement prévoir une ou des campagnes d'adhésion complémentaires pour maximiser les taux d'adhésion (*e.g.* zones où les proportions d'HLM sont élevées, collectivités avec faible participation) ;

- Ils nous permettent de sélectionner, lors des périodes de très forte activité en souscription, les appels d'offres auxquels il est préférable de répondre (*i.e.* ceux où l'on peut espérer les plus forts taux d'adhésion).

Bibliographie

Akaike, H. (1974) A new look at statistical model identification. ***IEEE Transactions on Automatic Control***, AU-19 ; pp716-722.

Allison P. (2013) Why I don't trust the Hosmer-Lemeshow test for Logistic regression ; ***Statistical horizons*** 05/03/2013 ; http://www.statisticalhorizons.com/hosmer-lemeshow.

Allison P. (2014a) Measures of fit for logistic regressions ; SAS Global Forum ; Paper 1485-2014 ; 13pp.

Allison P. (2014b) Another Goodness Of Fit test for logistic regression ; ***Statistical horizons*** 13/05/2014 ;http://www.statisticalhorizons.com/another-goodness-of-fit-test-for-logistic-regression.

Carruthers E., Lewis K., McCue T. & Westley P. (2008) Generalized linear models : model selection, diagnostics, and overdispersion, 22pp.

Collectivites-locales.gouv (2010) Maintien du régime indemnitaire dans certaines situations de congés - Octobre 2010 ; http://www.collectivites-locales.gouv.fr/maintien-regime-indemnitaire-dans-certaines-situations-conges-octobre-2010, consulté le 13/12/2014.

Cooch E.G. & White (2014) Chapter 5 : Goodness of fit testing… ; ***Program Mark : a gentle introduction*** ; 13th edition ; Cooch & White eds ; 1025pp ; pp 5.1-5.40.

Cox D.R. (1972) Regression models and life-tables. ***Journal of the Royal Statistical Society. Series B*** (Methodological), 34(2) ; pp187–220.

CSFPT (2011) La précarité dans la fonction publique territoriale ; Rapport du Conseil Supérieur de la Fonction Publique Territoriale, 116pp.

DGAFP (2013) Fonction Publique : chiffres clés 2013 ; ***Rapport Annuel de la Direction Générale de l'Administration et de la Fonction Publique***, 16pp ; http://www.fonction-publique.gouv.fr/files/files/statistiques/chiffres_cles/pdf/Chiffre_Cles_2013.pdf.

Economie.gouv (2012) Assurance prévoyance des salariés du secteur privé ; Fiches pratiques, 21-05-2012 ; http://www.economie.gouv.fr/dgccrf/Publications/Vie-pratique/Fiches-pratiques/Assurance-prevoyance-des-salaries-du-secteur-prive, consulté le 14/12/2014.

Hallett D.C. (1999) Goodness of fit tests in logistic regression ; Master of science, University of Toronto ; 129pp.

Hardin J.W. & Hilbe J.M. (2012) 11.3 Williams' procedure *; **Generalized Estimating Equations, Second Edition*** ; pp175-178.

Hosmer D.W., Hosmer T., Le Cessie S. & Lemeshow S. (1997) A comparison of goodness-of-fit tests for the logistic regression model ; ***Statisics in Medicine***, 16 ; pp965–980.

Hosmer D.W. & Lemeshow S. (1980) Goodness-of-fit tests for the multiple logistic regression model ; ***Communications in Statistics Theory Meth. A***, A9 (10) ; pp1043–1069.

iFRAP (2014) V.Le palmarès iFRAP de l'absentéisme ; ***Palmarès iFRAP***, Société civile n°142 ; pp 24-27.

Le Rest K. (2013) Méthodes statistiques pour la modélisation des facteurs influençant la distribution et l'abondance de populations ; thèse de Doctorat soutenue le 19/12/2013.

Paris A. (2012) Chapitre 2 : En assurance de personne ; ***Evolution de la notion d'aléa en droit des assurances***, Partie 2-Titre 1, Mémoire de l'Institut des Assurances de Lyon ; http://www.institut-numerique.org/evolution-de-la-notion-dalea-en-droit-des-assurances-4fd85587af9c7.

Rotella J. (2013) Testing GOF & estimating overdispersion ; ***Analysis of Population and Habitat Data : GOF & overdispersion*** ; Course reading ; 7pp.

Schoenfeld D. (1982) Partial Residuals for The Proportional Hazards Regression Model ; ***Biometrika***, 69(1) ; pp239-241.

Stukel T.A. (1988) Generalized logistic models ; ***Journal of the American Statistical Association***, 83 ; pp426–431.

Timsit J.F., Alberti C. & Chevret S. (2005) Le modèle de Cox ; ***Revue des Maladies Respiratoires***, 22(6) ; pp1058-1064.

Williams D. A. (1982), Extra-binomial Variation in Logistic Linear Models ; ***Applied Statistics***, 31 ; pp144–148.

Sources de données publiques

INSEE, Base chiffres clés : logement 2011
http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=base-cc-logement-2011

Mémoires d'actuariat consultés

Bargain D. (2011) Modélisation du taux d'absentéisme en entreprise.

Pérol M. & Santina O., della (2014) Influence de la conjoncture économique sur la fréquence d'entrée en arrêt de travail.

Rose N. (2009) Provisionnement en assurance non-vie : utilisation de modèles paramétriques censurés.

Verchay J.B. (2008) Le provisionnement des garanties en complément au statut des agents territoriaux.

Annexe 1 : Liens internet du décret du 08 novembre 2011 et des éléments associés

1. Décret n°2011-1474 du 8 novembre 2011 relatif à la participation des collectivités territoriales et de leurs établissements publics au financement de la protection sociale complémentaire de leurs agents :

http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20111110&numTexte=31&pageDebut=18895&pageFin=18899

2. Arrêtés associés :

- Arrêté du 8 novembre 2011 relatif aux critères de choix des collectivités territoriales et des établissements publics en relevant dans le cas d'une convention de participation

http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20111110&numTexte=71&pageDebut=18934&pageFin=18935

- Arrêté du 8 novembre 2011 relatif aux majorations de cotisations prévues par le décret n° 2011-1474 du 8 novembre 2011 relatif à la participation des collectivités territoriales et de leurs établissements publics au financement de la protection sociale complémentaire de leurs agents

http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20111110&numTexte=69&pageDebut=18933&pageFin=18933

- Arrêté du 8 novembre 2011 relatif à l'avis d'appel public à la concurrence publié au Journal officiel de l'Union européenne pour le choix des organismes en cas de convention de participation

http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20111110&numTexte=72&pageDebut=18935&pageFin=18943

- Arrêté du 8 novembre 2011 relatif à la composition du dossier de demande d'habilitation des prestataires habilités à délivrer les labels pour les contrats et règlements ouvrant droit à participation à la protection sociale complémentaire des agents de la fonction publique territoriale

http://www.legifrance.gouv.fr/jopdf/common/jo_pdf.jsp?numJO=0&dateJO=20111110&numTexte=70&pageDebut=18933&pageFin=18934

3. Circulaire d'application :

http://circulaire.legifrance.gouv.fr/pdf/2012/06/cir_35387.pdf

Annexe 2 : codes d'implémentation du test de Stukel sous SAS et du test des résidus de Schoenfeld sous R

1. Test de Stukel

Les lignes de commandes ci-dessous correspondent au modèle logistique retenu pour l'étude du taux d'adhésion, avec le test d'ajustement de Stukel :

```
proc logistic data=statmem.Datparticipation ;
class classpot(ref='A-1-10') typecoll2(ref='commune') / param=ref ;
model adher/potentiel = classpot|participation typecoll2 hlm2 / scale=williams(0.166252) ;
output out=af xbeta=xbf;
data bf;
set af;
zaf=xbf**2*(xbf>=0);
zbf=xbf**2*(xbf<0);
num=1;
proc logistic data=bf;
class  classpot(ref='A-1-10') typecoll2(ref='commune') / param=ref ;
model adher/potentiel = classpot|participation typecoll2 hlm2 zaf zbf / scale=williams(0.166252) ;
test zaf=0,zbf=0;
run;
```

2. Test des résidus de Schoenfeld

Les lignes de commandes ci-dessous correspondent au modèle complet de l'analyse de l'incidence de l'incapacité indemnisée avec le test de l'hypothèse de PH (risques propotionnels) :

```
coxarret=coxph(Surv(Dwsin, sin) ~ Sexe*clage + cldadh + cldadh:Sexe + cldadh:clage)
summary(coxarret)
res.c=cox.zph(coxarret)
res.c
```