

Étude

Modèles d'analytique prédictive pour l'expérience des taux de terminaison d'invalidité en assurance collective au Canada

**Commission sur les études d'expérience –
Direction de la recherche**

Mai 2019

Document 219060

This document is available in English

© 2019 Institut canadien des actuaires

Table des matières

INTRODUCTION.....	4
GOVERNANCE DU PROJET	5
ACTIVITÉ DU PROJET	5
Assureurs ayant présenté des données.....	5
Équipe du projet	6
TERMINOLOGIE.....	6
RESSOURCES EN MATIÈRE DE DONNÉES.....	7
Données externes	7
Tables désignées	8
STRUCTURE DES TABLES DÉSIGNÉES	9
EXAMENS PRÉLIMINAIRES	9
Modèles linéaires généralisés : Régression logistique	9
Modèles de survie.....	10
Procédure du biais minimal	11
ÉLABORATION DU MODÈLE	12
Description des données.....	12
Association entre les variables	13
CONSTRUCTION DU MODÈLE	14
Manipulation des données	14
Algorithme de construction du modèle.....	14
Facteurs d’ajustement – Version 2	21
APPLICATION DU MODÈLE.....	23
ÉVALUATION	24
MISES EN GARDE.....	25
CONCLUSION ET RECOMMANDATIONS.....	25
ANNEXE 1 – REVUE DE LA LITTÉRATURE.....	26
Tables des taux de terminaison de l’industrie.....	26
Analyse statistique	26
Modélisation prédictive	27
ANNEXE 2 – RECOURS À LA RÉGRESSION LOGISTIQUE POUR PRODUIRE LES TABLES DÉSIGNÉES	28
ANNEXE 3 – CODAGE DES INDUSTRIES.....	29
ANNEXE 4 – CODAGE DES DIAGNOSTICS.....	31
ANNEXE 5 – RÉSUMÉ DES FACTEURS LIÉS AUX MODÈLES.....	33
ANNEXE 6 – EXEMPLES DE L’UTILISATION DE LA VERSION 1 DU MODÈLE	36

LISTE DES TABLES

Table 1 : Forces et faiblesses des modèles de régression logistique.....	10
Table 2 : Forces et faiblesses des modèles de survie	11
Table 3 : Forces et faiblesses des modèles selon la procédure du biais minimal.....	11
Table 4 : Variables de données disponibles.....	12
Table 5 : Degré d'association – Statistique V de Cramér.....	14
Table 6 : Facteurs liés au modèle d'analytique prédictive selon la PBM – Version 1 (toutes les durées)	15
Table 7 : Exemple – Facteur PBM par rapport au facteur R/P traditionnel.....	17
Table 8 : Facteurs liés au modèle d'analytique prédictive selon la PBM – Version 2 (selon la durée)	22
Table 9 : Codes des industries.....	29
Table 10 : Correspondance des codes des industries.....	30
Table 11 : Codes des diagnostics	31
Table 12 : Correspondance des codes des diagnostics.....	32

INTRODUCTION

La présente analyse de l'expérience des taux de terminaison d'invalidité de longue durée (ILD) en assurance collective émis au Canada a été effectuée par la Direction de la recherche de l'Institut canadien des actuaires (ICA).

Le mandat de la recherche consistait à étendre les résultats de l'[Étude sur les taux de terminaison d'ILD en assurance collective de 2018](#), qui a produit des tables de terminaison d'ILD actualisées en fonction de l'expérience pour la période de 2009 à 2015.

Les nouvelles tables d'invalidité (comme les versions antérieures) varient selon le sexe, l'âge, la durée de la réclamation et la région (Québec et autres). Ce projet visait à créer des facteurs, des formules ou des méthodes qui, conjointement avec les quatre tables, pourraient être utilisés pour intégrer l'effet des variables suivantes, en paires et en groupes, dans l'estimation des taux de terminaison :

- 1) Diagnostic;
- 2) Âge;
- 3) Sexe;
- 4) Lieu de résidence;
- 5) Durée de l'invalidité;
- 6) Codes des industries;
- 7) Prestations mensuelles;
- 8) Salaire;
- 9) Situation fiscale;
- 10) Intégration du RPC/RRQ – état et montant;
- 11) Intégration des indemnités d'accident du travail;
- 12) Prestations préalables à l'ILD – oui/non et type;
- 13) Définition initiale de l'invalidité;
- 14) Durée maximale des prestations;
- 15) Délai de carence.

L'ICA a fait appel aux cabinets Fraser Group et Denis Garand et Associés pour assurer la gestion de l'étude. Il leur a confié le mandat suivant :

- Effectuer une revue de la littérature.
- Examiner différentes techniques d'analytique qui pourraient être utilisées pour créer les outils d'analytique prédictive souhaités.
- Créer un outil pratique qu'un actuaire peut appliquer après avoir lu le rapport.
- Préparer de la documentation pertinente, notamment le présent rapport.

GOUVERNANCE DU PROJET

Président de la Direction de la recherche :

Keith Walter

Composition du groupe chargé de la surveillance du projet (GSP) :

Frank Reynolds (président)

Jean-François Blais

Pierre-Philippe Carle-Mossdorf

Erin Crump

Rhys DeGrave

Lina Forner

Tim Griffin

Kateri Laneuville

Stella-Ann Ménard

Keith Walter (liaison avec la Direction de la recherche)

ACTIVITÉ DU PROJET

Le projet de recherche est fondé sur la base de données qui avait été créée pour l'élaboration des tables de terminaison d'ILD en assurance collective (2005 à 2015). Les données de cette étude provenaient de 16 assureurs représentant environ 99 % du marché canadien des contrats collectifs d'assurance invalidité de longue durée.

Assureurs ayant présenté des données

- Assomption Vie
- Croix Bleue, Assurance-vie
- Co-operators, Compagnie d'assurance-vie
- Financière Desjardins
- Empire Vie
- Assurance-vie Équitable
- Great-West, Compagnie d'assurance-vie
- Industrielle Alliance
- Humania
- La Capitale
- Manuvie
- Pacific Blue Cross (assurance-vie en Colombie-Britannique)
- RBC, Assurance-vie
- SSQ
- Sun Life
- Wawanesa, Assurance-vie

Équipe du projet

L'ICA a fait appel aux cabinets Fraser Group et Denis Garand et Associés pour assurer la gestion de l'étude. Les chefs de projet étaient Ken Fraser et Denis Garand, tous deux FICA.

L'équipe de projet comprenait également :

- Donna Swiderek, AICA, du cabinet Denis Garand et Associés;
- Clayton Zaluski, FICA; Stephen Swenarchuk, AICA; et Merv Worden, FICA, du cabinet Worden Zaluski Consulting Actuaries;
- Taehan Bae, Ph.D., AICA, de l'Université de Regina.

TERMINOLOGIE

Cette section présente les termes clés utilisés tout au long de la présente étude.

R/P signifie *Réel à Prévu* et désigne normalement le ratio du nombre réel de terminaisons de réclamations au nombre prévu obtenu en appliquant une table de référence à l'exposition.

Toute occupation et *propre occupation* font référence aux définitions de l'invalidité utilisées dans le contrat d'assurance ILD. Pour *propre occupation*, l'invalidité se définit comme étant l'incapacité du demandeur d'exécuter les fonctions essentielles de sa *propre occupation*, et pour *toute occupation*, l'incapacité du demandeur d'exécuter les fonctions de *toute occupation* pour laquelle l'employé possède la formation, les études ou l'expérience nécessaires.

Changement de la définition de l'invalidité (CDI) fait renvoi à la disposition de la plupart des contrats d'assurance ILD en vertu de laquelle la définition d'invalidité passe de *propre occupation* à la plus stricte de *toute occupation* après une période d'invalidité initiale (habituellement deux ans). Ainsi, une personne peut être admissible à des prestations d'invalidité pendant un certain temps, puis ne plus y être admissible même si les preuves d'ordre médical ou professionnel n'ont pas changé.

Les *tables désignées* désignent les tables de terminaison d'ILD en assurance collective publiées dans la plus récente étude de l'ICA (2009 à 2015), *Étude sur les taux de terminaison d'invalidité de longue durée en assurance collective* ([document 219012](#)).

Exposition a son sens actuariel habituel et s'entend des réclamations qui sont actives et donc « exposées » à une terminaison éventuelle. L'exposition se quantifie sous forme de nombre de réclamations (plutôt que sous forme de montant des prestations).

ILD signifie *assurance invalidité de longue durée*. Dans la présente étude, le sigle fait exclusivement référence à la garantie offerte sur une base collective. Cette question est abordée de façon plus détaillée à la rubrique Ressources en matière de données.

ICD signifie *prestations d'invalidité de courte durée*, également connu sous le nom de garanties d'indemnités hebdomadaires.

Le terme *rétablissement* est utilisé dans la présente étude pour désigner toute terminaison qui n'est pas attribuable à la mortalité. Même si ce terme comprend le sens en langage simple (c.-à-d. les demandeurs se sont médicalement rétablis de leur blessure ou maladie et ont repris le

travail), le *rétablissement* implique également, dans la présente étude, une situation où la société d'assurance a mis fin à une réclamation pour une raison autre que le décès. Il s'agit notamment de scénarios de changement de la définition de l'invalidité où le demandeur n'est plus admissible conformément à une définition plus stricte de l'invalidité et de situations où les demandeurs abandonnent une réclamation en ne soumettant pas l'information requise et où la société d'assurance détermine que les preuves ne justifient pas la poursuite du versement des prestations.

L'*étude des tables* désigne l'étude de l'ICA sur l'expérience des taux de terminaison d'ILD pour les années 2009 à 2015, *Étude sur les taux de terminaison d'invalidité de longue durée en assurance collective* ([document 219012](#)).

Terminaison s'entend de tout événement éventuel qui met fin à une réclamation autrement active. Ainsi, une réclamation qui prend fin une fois la période maximale de prestations atteinte (p. ex. à l'âge de 65 ans) n'est pas une terminaison. Dans la présente étude, le terme *terminaison* comprend la mortalité et le rétablissement (voir ci-dessus).

RESSOURCES EN MATIÈRE DE DONNÉES

Aucune collecte ou validation de données n'était nécessaire pour ce projet puisque nous avons utilisé la base de données créée pour l'étude des tables de terminaison. Le lecteur devrait se reporter à cette étude¹ pour obtenir plus de détails sur la collecte et la validation des données et la création des tables.

La présente section résume le cadre conceptuel des données fournies pour la réalisation du projet. Les réclamations utilisées dans le projet sont caractérisées par ce qui suit :

- Elles sont associées aux polices d'assurance collective émises aux employeurs canadiens, aux fiducies à employeurs multiples et aux caisses sociales des syndicats.
- La date d'invalidité précède le 31 décembre 2015, et les réclamations étaient « en versement » pendant un certain temps entre le 1^{er} janvier 2009 et le 31 décembre 2015.

Données externes

Nous avons envisagé la possibilité d'inclure dans le projet des données externes comme les taux d'inflation, d'intérêt et de chômage. Nos recherches ont cependant révélé que ces facteurs sont demeurés relativement stables au cours de la période visée par les données d'expérience. L'intégration de ces éléments à l'étude aurait exigé beaucoup de ressources, mais aurait probablement produit peu de résultats utiles. Toutefois, il peut s'agir d'un domaine à explorer dans de futures études.

¹ Les auteurs principaux de la présente étude ont aussi signé l'étude des tables de terminaison.

Tables désignées

Le mandat du projet consistait à créer des facteurs de rajustement qui pourraient être utilisés conjointement avec les tables de l'étude publiée afin de mieux prédire l'expérience de terminaison. Ainsi, les tables désignées peuvent être considérées comme faisant partie des ressources en matière de données du projet.

La construction d'une table à partir de données brutes nécessite des décisions sur de nombreuses questions pratiques et il existe souvent des divergences entre les objectifs techniques concurrents. C'est pourquoi il est utile de mentionner les utilisations prévues qui ont guidé l'équipe du projet, comme documenté dans le rapport.

Voici les utilisations en question :

- Évaluation par les assureurs du passif des réclamations d'assurance ILD en cours au Canada dans les états financiers;
- Calcul par les assureurs du passif des réclamations dans la comptabilisation d'expérience pour des titulaires de polices spécifiques;
- Utilisation par les assureurs dans la détermination des taux manuels pour les prestations des contrats d'assurance ILD collective;
- Évaluation par les promoteurs de régimes autofinancés du passif des réclamations d'assurance ILD en cours au Canada dans les états financiers.

Les auteurs de l'étude des tables ont décrit comme suit le scénario dans lequel les tables désignées peuvent être considérées comme cadrant le mieux. Les éléments de ce modèle comprennent ce qui suit :

- Régimes d'avantages sociaux du personnel;
- Garantie souscrite sur une base assurée (c.-à-d. excluant les régimes d'avantages sociaux non assurés (RASNA));
- Employés canadiens;
- Groupes de diverses tailles;
- Niveaux d'adhésion élevés;
- Émission principalement garantie avec souscription individuelle de montants excédentaires;
- Délais de carence de quatre à six mois;
- Prestations payables jusqu'à 65 ans;
- Taux de remplacement élevés, mais inférieurs à 100 %;
- Définition d'invalidité basée sur la propre occupation pendant deux ans;
- Dispositions générales de l'industrie pour l'invalidité récurrente, les limites de toutes les sources, la réadaptation, etc.;
- Pratiques types de l'industrie en matière de gestion des réclamations, telles que l'intervention précoce.

Les utilisateurs sont avertis de la nécessité d'envisager des ajustements face à une situation qui varie considérablement du scénario type, par exemple, de longs délais de carence et des dispositions variantes des contrats.

STRUCTURE DES TABLES DÉSIGNÉES

Cette section résume la structure des tables désignées. Les facteurs élaborés dans le projet visent à ajuster les valeurs de la table de base.

Dans les tables désignées, les valeurs de terminaison sont établies en ajoutant des facteurs provenant de deux tables de composantes :

- Table de base;
- Ajustement pour le CDI.

Les taux sont présentés séparément pour :

- Nombre total de terminaisons;
- Terminaisons attribuables au décès (mortalité);
- Terminaisons attribuables à d'autres raisons (rétablissement).

Les tables se présentent en quatre sections divisées comme suit :

- Québec par rapport au reste du Canada;
- Femmes par rapport aux hommes.

Chaque section renferme ce qui suit :

- Les valeurs sélectes des 120 premiers mois de l'invalidité. Elles indiquent les terminaisons mensuelles pour des groupes d'âge de cinq ans (âge de début de l'invalidité) pour les mois cinq à 60. Les terminaisons annuelles sont fournies pour les prochains cinq ans.
- Les valeurs ultimes pour les durées au-delà de 10 ans. Ces valeurs sont fournies selon le sexe et l'âge atteint et aucune distinction n'est faite entre le Québec et le reste du Canada.

Les taux sont indiqués sous forme de valeurs mensuelles pour les 60 premiers mois et de valeurs annuelles par la suite.

EXAMENS PRÉLIMINAIRES

L'utilisation de plusieurs techniques d'analytique prédictive a été envisagée dans le projet, à savoir les suivantes :

- Modèles linéaires généralisés : Régression logistique;
- Modèles de survie;
- Procédure du biais minimal.

Dans chaque cas, nous avons tenté de construire un modèle dans le cadre d'une technique donnée et d'utiliser le modèle pour produire les outils souhaités. Les modèles ont été construits à l'aide des plateformes Python et R et évalués selon les critères décrits dans les sections qui suivent.

Modèles linéaires généralisés : Régression logistique

Les modèles linéaires généralisés (MLG) sont largement utilisés dans l'industrie des assurances IARD en raison de leur pouvoir prédictif. Ils ont également été utilisés aux États-

Unis pour élaborer des tables de mortalité. En utilisant une fonction de lien « logit » et en supposant que le terme d'erreur suit la loi binomiale, il est possible de construire un type particulier de MLG, appelé modèle de régression logistique.

Les modèles de régression logistique (RL) servent à prédire une variable dépendante binaire : 0 ou 1. Un modèle de RL peut être construit pour une étude des taux de terminaison d'ILD où :

- 0 signifie que la réclamation n'a pas pris fin (non-terminaison);
- 1 signifie que la réclamation a pris fin (terminaison).

Le modèle de RL produit une équation servant à prédire la terminaison probable des réclamations d'après un nombre quelconque de variables indépendantes. Nous avons relevé les forces et les faiblesses suivantes lors de la mise à l'essai et de la construction de modèles de RL.

Table 1 : Forces et faiblesses des modèles de régression logistique

Forces	Faiblesses
<ul style="list-style-type: none"> • Le pouvoir prédictif est bien documenté. • Produit facilement des statistiques utiles (valeurs p, intervalles de confiance). • Permet un « lissage » inhérent, éliminant la nécessité de méthodes algébriques. 	<ul style="list-style-type: none"> • Complexité. • L'objectif consiste à élaborer des facteurs, des formules ou des méthodes pour ajuster les valeurs des tables désignées; le modèle de RL n'atteint pas cet objectif. • Il est difficile de rendre compte de l'effet du CDI.

Tout compte fait, il a été décidé de ne pas appliquer de modèle de RL puisqu'il n'atteindrait pas l'objectif de l'étude. Ce type de modèle produirait de nouvelles valeurs pour les taux de base et il serait difficile de rattacher les résultats aux tables désignées. Toutefois, nous avons examiné l'utilisation d'un modèle de RL à titre de méthode pour établir les tables de taux de terminaison de base (selon l'âge, le sexe, la durée des réclamations et Québec/hors-Québec), ce qui est décrit de façon détaillée à l'annexe 2.

Modèles de survie

Les modèles de survie servent à estimer la répartition de la variable aléatoire T (période précédant la terminaison). Des méthodes non paramétriques et semi-paramétriques peuvent être utilisées, et nous avons examiné chaque possibilité.

La méthode non paramétrique la plus courante est l'estimateur produit-limite de Kaplan-Meier. Cette méthode est fondée sur le nombre de réclamations actives qui précèdent immédiatement chaque terminaison observée.

Le modèle à risque proportionnel de Cox est un modèle semi-paramétrique couramment utilisé pour analyser la durée en raison de sa capacité d'intégrer la fonction de risque de

base et une combinaison linéaire de variables prédictives. Il permet d'effectuer des tests formels des hypothèses relatives à l'importance des variables prédictives.

Nous avons constaté les forces et les faiblesses suivantes relativement aux modèles de survie :

Table 2 : Forces et faiblesses des modèles de survie

Forces	Faiblesses
<ul style="list-style-type: none"> • Efficacité théorique. • Permet de mesurer le pouvoir prédictif d'un ensemble de variables. • Permet un « lissage » inhérent, éliminant la nécessité de méthodes algébriques. 	<ul style="list-style-type: none"> • Requier beaucoup de calculs. • Il est difficile d'étalonner l'ajustement du modèle à des durées spécifiques. • Il est difficile d'isoler l'effet du CDI. • L'hypothèse de proportionnalité du modèle à risque proportionnel de Cox est invalidée dans l'ensemble de données.

Procédure du biais minimal

La procédure du biais minimal (PBM) est une méthode de modélisation prédictive couramment utilisée pour la tarification d'assurance. Un modèle multiplicateur selon la PBM génère des facteurs pour chaque variable prédictive.

Le processus est itératif et requiert la saisie des montants réels et prévus. La valeur prévue est une terminaison probable, comme déterminé dans les tables désignées. La valeur réelle est 1 si la réclamation a pris fin et 0 si elle n'a pas pris fin. Les montants réels et prévus doivent être établis pour chaque réclamation selon chaque durée.

Nous estimons que la PBM présente les forces et les faiblesses suivantes :

Table 3 : Forces et faiblesses des modèles selon la procédure du biais minimal

Forces	Faiblesses
<ul style="list-style-type: none"> • Extrait connu. • Peut s'appliquer directement aux taux de terminaison des tables désignées. • Répond facilement à l'objectif de l'étude. 	<ul style="list-style-type: none"> • Degré de perfectionnement moindre par rapport aux autres techniques. • Ne produit pas de mesures statistiques de l'ajustement. • Les résultats supposent implicitement que la valeur de chaque variable a le même effet sur toutes les durées.

Après avoir évalué les résultats des examens préliminaires, nous avons décidé de mettre l'accent sur la PBM, car elle est la mieux adaptée pour atteindre les objectifs de l'étude.

ÉLABORATION DU MODÈLE

La présente section décrit la façon dont nous avons utilisé la PBM pour élaborer les facteurs d'ajustement recommandés dans le présent rapport.

Description des données

Les variables suivantes, qui pourraient être utilisées dans un modèle prédictif, ont été établies à partir des données de l'étude des tables de terminaison. Comme indiqué dans la table 4 ci-dessous, nous avons retenu certaines variables et en avons exclu d'autres.

Table 4 : Variables de données disponibles

Variable de données	Compris dans le modèle	Raison
Âge	Non	Compris dans les tables désignées
Sexe	Non	Compris dans les tables désignées
Durée de l'invalidité	Non	Compris dans les tables désignées
Définition initiale de l'invalidité	Non	Compris dans les tables désignées
Province	Oui	
Diagnostic	Oui	
Industrie	Oui	
Prestations mensuelles	Oui	
Prestations préalables à l'ILD	Oui	
Délai de carence	Oui	
Durée maximale des prestations	Non	Compris dans la mesure de l'exposition
Salaire	Non	Corrélié avec les prestations mensuelles
Situation fiscale	Non	Piètre qualité des données
Intégration RPC/RRQ	Non	Piètre qualité des données
Indemnités d'accident du travail	Non	Piètre qualité des données

Les variables comprises dans les tables désignées sont exclues du projet afin d'éviter le double comptage. Les montants des prestations mensuelles sont habituellement calculés en

pourcentage du salaire mensuel. Les prestations mensuelles sont incluses et le salaire² a été exclu pour éviter le double comptage.

La durée maximale des prestations est utilisée pour calculer les expositions et a été exclue de l'étude pour éviter le double comptage.

Les données relatives à la situation fiscale, l'intégration RPC/RRQ et aux indemnités d'accident du travail n'ont pas été fournies par toutes les sociétés et contenaient de nombreux éléments inconnus et suspects. Ces champs ont été exclus afin de maintenir l'intégrité des résultats.

Au total, six variables étaient assorties de données relativement robustes et n'avaient pas été intégrées à la structure des tables désignées :

- Province;
- Diagnostics;
- Industries;
- Prestations mensuelles;
- Prestations préalables à l'ILD;
- Délai de carence.

Ces six variables ont été incluses dans l'analyse.

Association entre les variables

Puisque le but du modèle proposé est de démêler les effets des variables prédictives corrélées, nous avons examiné les variables afin de déterminer l'indépendance ou la corrélation.

À la première étape, à l'aide du test chi carré, l'hypothèse d'indépendance entre chaque paire de variables prédictives a été rejetée puisque les valeurs p étaient toutes proches de zéro³.

La deuxième étape consistait à évaluer le degré de corrélation entre les variables. Le degré d'association entre les variables peut être mesuré au moyen de la statistique V de Cramér⁴ qui est semblable à la statistique r^2 , mais elle peut être utilisée avec les variables catégoriques. Comme pour la statistique r^2 , une valeur de 0,00 signale l'absence d'association et une valeur de 1,00 dénote une association parfaite.

² Les données salariales dans l'étude des tables n'étaient pas disponibles pour près de 30 % des données.

³ Il est cependant à noter qu'une faible valeur p ne signifie pas nécessairement qu'il existe une forte dépendance entre les variables, mais plutôt que l'hypothèse d'indépendance n'est pas appropriée compte tenu des données. Dans un grand ensemble de données, même une association faible peut donner lieu à une valeur p très faible.

⁴ La statistique V de Cramér se calcule comme suit :

$$\sqrt{\frac{\chi^2}{N(k-1)}}$$

où

χ^2 est la statistique du chi carré de Pearson;

N est la taille de l'échantillon en question dans le test;

k est le nombre inférieur de catégories de l'une ou l'autre variable.

Une valeur de 0,25 est un seuil type qui signale une association modérée, tandis que les valeurs inférieures à 0,15 dénotent une association très faible.

La table suivante montre que le niveau d'association varie de 0,055 (Diagnostic – Délai de carence) à 0,253 (Industrie – Délai de carence).

Table 5 : Degré d'association – Statistique V de Cramér

	Délai de carence	Prestations préalables à l'ILD	Prestations mensuelles	Diagnostic	Province
Industrie	0,25	0,18	0,18	0,06	0,09
Délai de carence		0,18	0,16	0,06	0,09
Prestations préalables à l'ILD			0,18	0,09	0,22
Prestations mensuelles				0,07	0,14
Diagnostic					0,08

Puisque certaines paires de variables présentent une association quelconque, on s'attend à ce qu'un modèle d'analytique prédictive ajoute de la valeur en déterminant la contribution relative de chaque variable.

CONSTRUCTION DU MODÈLE

Manipulation des données

Le modèle selon la PBM nécessite l'extension des données sur les réclamations afin de créer un enregistrement pour chaque mois d'exposition de la réclamation. La table de données a été transformée, passant de 485 000 enregistrements de réclamations à 10,8 millions enregistrements de mois d'exposition de la réclamation. Ce processus a été mené à terme en utilisant une logique conforme à celle de l'étude des tables, mais avec une différence notable. L'étude des tables a calculé l'exposition précise jusqu'à une journée donnée, tandis que selon la PBM, l'exposition est calculée pour chaque mois de la durée. Il en a résulté de légers écarts par rapport à l'étude des tables, qui sont présentés ci-dessous.

La PBM exige que toutes les variables soient catégoriques. Les variables continues, comme le montant des prestations, ont été regroupées en tranches ou en fourchettes à cette fin. Les fourchettes ont été déterminées en fonction de l'exposition et d'une cohérence raisonnable avec l'analyse de l'étude des tables.

Algorithme de construction du modèle

La première étape consistait à ajuster les terminaisons prévues, de sorte que le ratio global R/P soit égal à 1,00. Pour ce faire, les terminaisons prévues ont été multipliées par un facteur fixe. Dans notre étude, ce facteur fixe (la « variable de pondération ») était de 1,02, ce qui indique que les terminaisons réelles sont 2 % plus élevées que les terminaisons prévues établies à partir des tables désignées. Cette différence est attribuable au fait que l'exposition est calculée en

fonction du mois le plus près dans ce projet, comparativement à l'étude des tables où elle a été calculée jusqu'à une journée donnée.

Le modèle selon la PBM est construit en suivant un processus itératif, dans le cadre duquel des variables prédictives sont ajoutées au fur et à mesure. Ainsi, à chaque étape, le logiciel de la PBM⁵ ajuste les réclamations prévues de manière que le ratio R/P soit égal à 1,00 pour chaque catégorie associée à une variable indépendante. Les ratios R/P convergent éventuellement à 1,00, ce qui signifie que d'autres ajustements ne sont pas nécessaires.

Le facteur final provient du résultat des facteurs d'ajustement utilisés dans chaque itération de la PBM.

Les données de sortie de la PBM est un ensemble de facteurs multiplicateurs d'ajustement qui peuvent être utilisés pour modifier les taux de terminaison prévue pour les réclamations ayant une caractéristique donnée. Par exemple, la catégorie de l'industrie Travail de bureau et services professionnels est associé à un facteur de l'industrie de 1,018 selon la PBM (voir la table 6). Pour évaluer une réclamation, les taux de terminaison des tables désignées seraient multipliés par 1,018 si le demandeur appartient à la catégorie Travail de bureau et services professionnels. D'autres facteurs seraient appliqués en fonction de chacune des variables prédictives.

Les facteurs de toutes les variables sont présentés dans la table suivante.

Table 6 : Facteurs liés au modèle d'analytique prédictive selon la PBM – Version 1 (toutes les durées)

INDUSTRIE		PBM	EXPOSITION	
Industrie lourde		1,038	1 332 695	12,3 %
Fabrication		0,994	1 522 271	14,1 %
Commerce de gros et de détail		1,019	1 265 085	11,7 %
Travail de bureau et services professionnels		1,018	1 616 981	14,9 %
Santé, éducation, services sociaux		1,025	1 312 515	12,1 %
Autres services (secteur privé)		0,990	875 876	8,1 %
Administration publique		0,920	1 802 829	16,7 %
Inconnu		1,017	1 088 653	10,1 %

DÉLAI DE CARENCE		PBM	EXPOSITION	
0 à 3 mois		0,948	2 101 997	19,4 %
4 mois		1,018	4 772 036	44,1 %
5 à 6 mois		1,012	2 746 656	25,4 %
Plus de 6 mois		0,968	1 196 216	11,1 %

⁵ Logiciel créé exclusivement pour le projet.

PRESTATIONS PRÉALABLES À L'ILD		PBM	EXPOSITION	
Notre régime d'ICD		1,181	1 936 321	17,9 %
Autre ou aucun		0,939	8 880 584	82,1 %

MONTANT DES PRESTATIONS		PBM	EXPOSITION	
Inconnu		1,082	460 056	4,3 %
Moins de 1 499 \$		1,009	2 022 371	18,7 %
1 500 à 1 999 \$		0,975	2 102 488	19,4 %
2 000 à 2 499 \$		1,003	1 877 541	17,4 %
2 500 à 3 249 \$		1,018	2 111 827	19,5 %
Plus de 3 250 \$ ⁶		0,969	2 242 622	20,7 %

DIAGNOSTIC		PBM	EXPOSITION	
Troubles mentaux		1,026	3 072 438	28,4 %
Troubles du système musculo-squelettique		0,900	2 371 654	21,9 %
Néoplasmes (tumeurs, surtout cancers)		1,236	1 016 681	9,4 %
Troubles circulatoires		0,854	877 167	8,1 %
Maladies du système nerveux		0,526	1 200 137	11,1 %
Accidents		1,219	772 068	7,1 %
Toutes les autres causes non déterminées		1,049	1 336 964	12,4 %
Non déclaré ou inconnu		1,059	169 796	1,6 %

PROVINCE		PBM	EXPOSITION	
Colombie-Britannique		0,999	1 286 554	11,9 %
Alberta		1,189	1 100 046	10,2 %
Saskatchewan		1,242	281 222	2,6 %
Manitoba		1,112	359 099	3,3 %
Ontario		0,966	4 299 717	39,7 %
Québec ⁷		0,976	2 352 523	21,7 %
Ailleurs au Canada		0,906	1 137 744	10,5 %

⁶ Les groupes de données ont été choisis pour faire en sorte que chaque catégorie soit suffisamment fiable et pour réduire la complexité du modèle final. La table 25 de l'étude des tables a révélé que les ratios R/P variaient légèrement jusqu'à ce que les montants mensuels atteignent 15 000 \$.

⁷ Il faut souligner que le facteur PBM indiqué est appliqué aux taux de terminaison du Québec dans les tables désignées. Bien que les facteurs d'ajustement soient semblables en Ontario et au Québec, cela n'indique pas que les comportements liés à la terminaison soient identiques dans les deux provinces.

Il y a lieu d'examiner de plus près la variable Province. Les tables désignées renferment des valeurs distinctes pour le Québec et le reste du Canada. Ces tables présentent les valeurs prévues utilisées pour construire le modèle selon la PBM. Par conséquent, le facteur Province pour le Québec du modèle selon la PBM (0,976 dans la table) ne représente pas l'écart entre le Québec et le reste du Canada. Étant donné que les tables désignées tiennent déjà compte des différences majeures entre le Québec et le reste du Canada, le facteur Province du modèle selon la PBM rajuste seulement l'écart résiduel. Il est inférieur à 1,00 puisqu'une petite partie de l'écart entre le Québec et le reste du Canada est prise en compte par les autres variables du modèle.

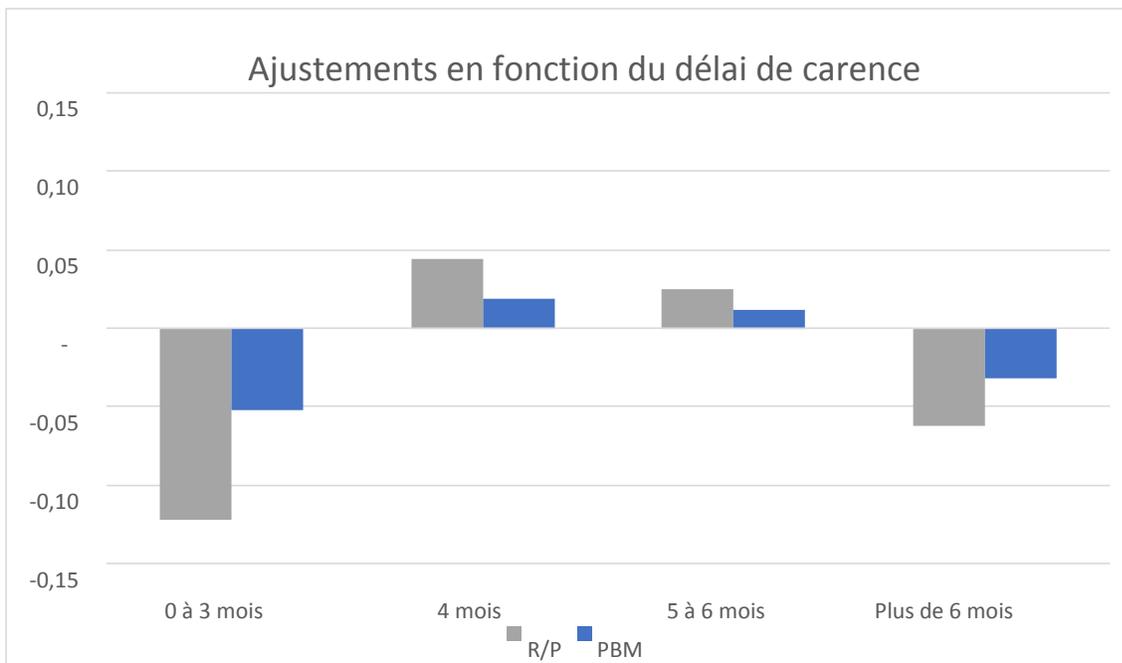
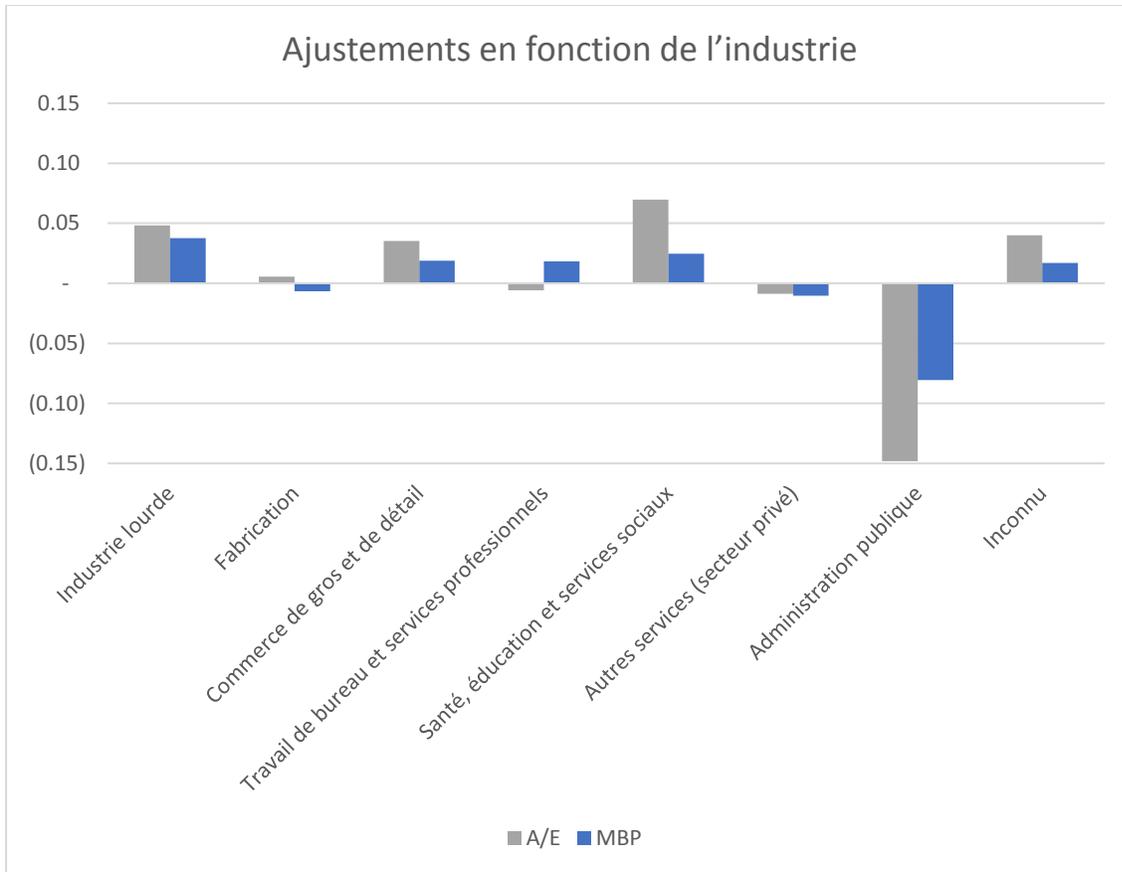
La valeur du modèle selon la PBM comparée à l'analyse R/P unidirectionnelle traditionnelle s'explique par le fait que les facteurs de la PBM ont éliminé l'effet de toutes les autres variables prédictives. Par exemple, dans la table 7 ci-dessous, l'analyse unidirectionnelle révèle que les taux de terminaison dans l'industrie de la santé, de l'éducation et des services sociaux sont 8 % plus élevés que ceux du secteur Travail de bureau et services professionnels ($1,07 - 0,99 = 0,08$). Le modèle selon le PBM n'affiche aucun écart (le facteur d'ajustement de ce modèle est de 1,02 pour les deux industries). L'écart apparent est en fait attribuable à d'autres facteurs (le diagnostic peut-être).

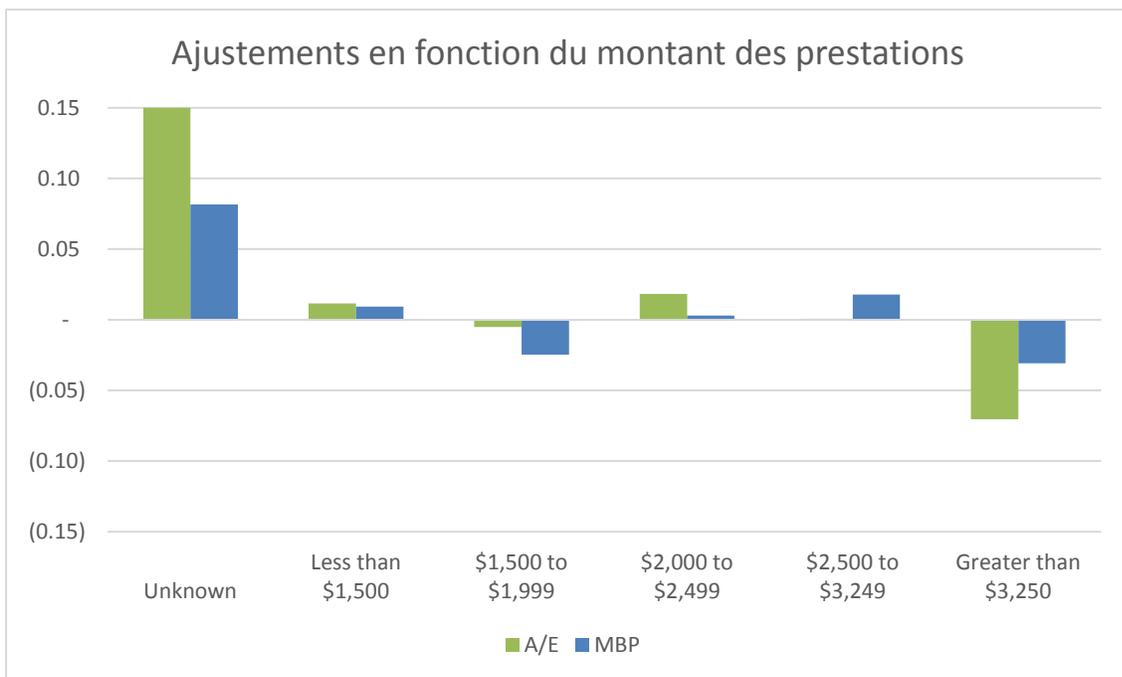
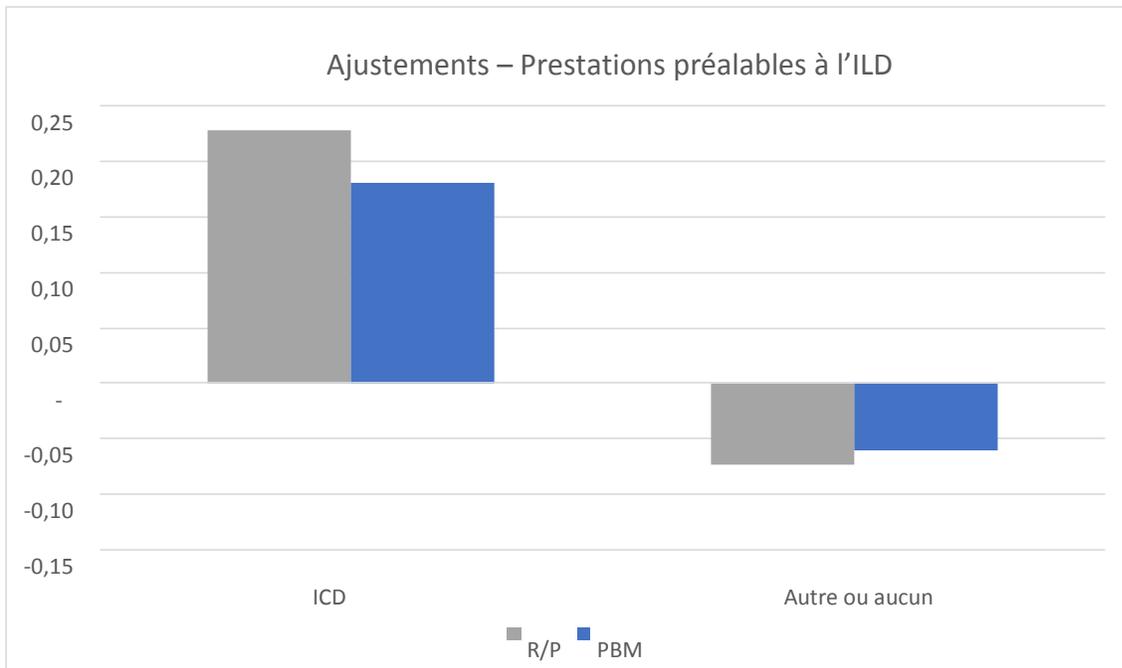
Table 7 : Exemple – Facteur PBM par rapport au facteur R/P traditionnel

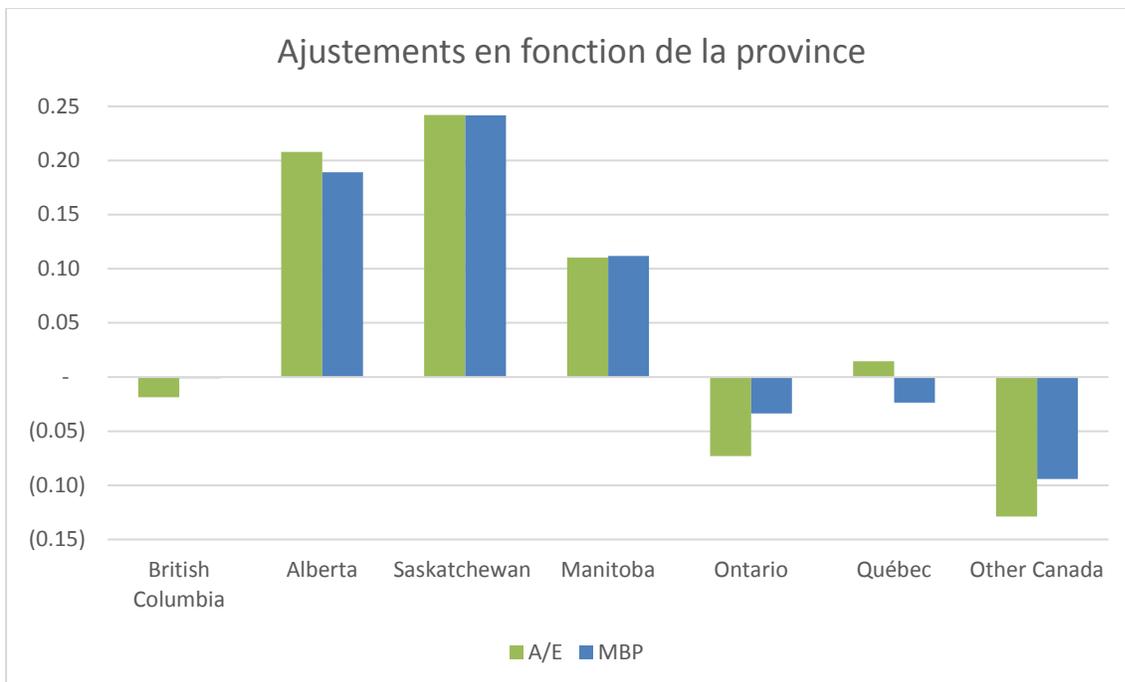
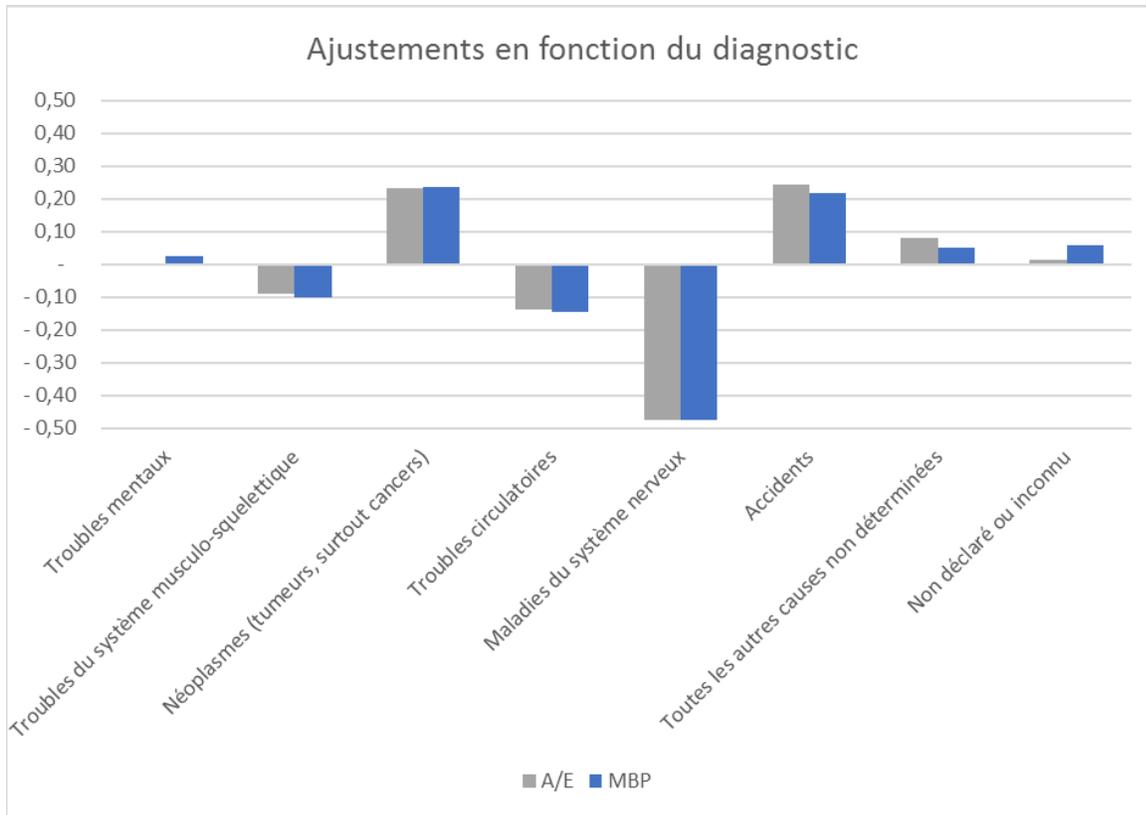
	Facteur R/P	Facteur PBM	Écart
Industrie lourde	1,05	1,04	-0,01
Fabrication	1,01	0,99	-0,01
Commerce de gros et de détail	1,04	1,02	-0,02
Travail de bureau et services professionnels	0,99	1,02	0,02
Santé, éducation, services sociaux	1,07	1,02	-0,04
Autres services (secteur privé)	0,99	0,99	-0,00
Administration publique	0,85	0,92	0,07
Inconnu	1,04	1,02	-0,02

Les graphiques qui suivent comparent les facteurs R/P calculés avec les facteurs PBM pour chaque variable. En règle générale, les facteurs PBM afficheront une variabilité semblable ou inférieure à celle de l'analyse R/P unidimensionnelle. Les facteurs seront semblables pour une variable lorsqu'il existe une très faible corrélation ou association avec d'autres variables.

[À noter : il n'a pas été possible de traduire une partie de l'axe x dans certains des graphiques qui suivent.]







Facteurs d'ajustement – Version 2

L'une des préoccupations entourant le facteur PBM est qu'il suppose implicitement que l'ajustement indiqué est constant pour toutes les durées. Il est peu probable qu'un facteur multiplicateur soit approprié pour une réclamation à toutes les durées, étant donné que le comportement lié à la terminaison est influencé par différents effets le long de la courbe de durée. Par exemple, les terminaisons d'assurance invalidité de courte durée (ICD) représentent surtout des recouvrements du fait que les prestataires retournent au travail. Après 36 mois, les taux de terminaison tendent principalement à refléter les décès.

Pour tenir compte de cette situation, nous avons élargi le modèle de la version 1 en créant deux autres ensembles de facteurs en nous fondant sur l'expérience relative :

- À la durée des réclamations de un à 36 mois;
- À la durée des réclamations au-delà de 36 mois.

Cette méthode produit des résultats plus précis en générant différents facteurs pour modifier les taux de terminaison à différentes durées.

En théorie, ce processus pourrait être élargi afin d'apporter encore plus de précision en divisant les données en intervalles plus courts. La recommandation relative à deux intervalles repose sur des considérations liées à la fiabilité des données et à la convivialité pour l'utilisateur final.

Les facteurs de la version 2 pour toutes les variables sont présentés à la table 8 ci-après. L'annexe 5 résume les deux versions du modèle et présente des données d'exposition complètes.

Table 8 : Facteurs liés au modèle d'analytique prédictive selon la PBM – Version 2 (selon la durée)

Industrie	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
Industrie lourde	1,033	1,105
Fabrication	0,997	0,941
Commerce de gros et de détail	1,022	0,994
Travail de bureau et services professionnels	1,025	0,950
Santé, éducation, services sociaux	1,024	1,018
Autres services (secteur privé)	0,989	0,968
Administration publique	0,906	1,083
Inconnu	1,025	0,928

DÉLAI DE CARENCE	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
0 à 3 mois	0,945	0,961
4 mois	1,021	0,984
5 à 6 mois	1,011	1,008
Plus de 6 mois	0,954	1,099

PRESTATIONS PRÉALABLES À L'ILD	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
Notre régime d'ICD	1,193	0,901
Autre ou aucune	0,933	1,019

MONTANT DES PRESTATIONS	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
Inconnu	1,080	1,080
Moins de 1 499 \$	1,003	1,087
1 500 à 1 999 \$	0,974	0,991
2 000 à 2 499 \$	1,002	1,012
2 500 à 3 249 \$	1,017	1,009
Plus de 3 250 \$	0,976	0,893

DIAGNOSTIC	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
Troubles mentaux	1,036	0,872
Troubles du système musculo-squelettique	0,906	0,822
Néoplasmes (tumeurs, surtout cancers)	1,181	2,656
Troubles circulatoires	0,854	0,877
Maladies du système nerveux	0,506	0,661
Accidents	1,227	1,036
Toutes les autres causes non déterminées	1,038	1,170
Non déclaré ou inconnu	1,086	0,811

PROVINCE	FACTEUR PBM (mois)	
	1 à 36 mois	Plus de 36 mois
Colombie-Britannique	1,002	0,980
Alberta	1,192	1,145
Saskatchewan	1,245	1,212
Manitoba	1,107	1,170
Ontario	0,963	1,009
Québec	0,976	0,966
Ailleurs au Canada	0,913	0,842

APPLICATION DU MODÈLE

Il est recommandé d'utiliser les facteurs des versions 1 et 2 dans les activités de tarification et d'évaluation. La version 2 assure une plus grande précision, mais cela n'est peut-être pas nécessaire pour toutes les utilisations. Voici des utilisations possibles :

- Évaluer les réserves pour réclamations en cours concernant les cas où un remboursement est prévu.
- Effectuer une analyse de rentabilité segmentée lorsque des segments sont définis selon l'une ou l'autre des dimensions utilisées dans le modèle (p. ex. province ou industrie).
- Évaluer les réserves pour réclamations prévues par la loi, en particulier en ce qui concerne les assureurs qui ont un petit bloc de polices ou lorsqu'un assureur estime que son portefeuille est faussé par rapport au marché global.

La procédure suivante est recommandée pour l'application du modèle construit :

1. Pour une réclamation donnée, déterminer un taux de terminaison à partir des tables désignées en fonction de l'âge au moment de l'invalidité, du sexe, de la région et de la durée de la réclamation.
2. Multiplier le taux calculé ci-dessus par chacun des facteurs pour les six variables du modèle. Le résultat est un taux de terminaison ajusté.

Voir l'annexe 6 pour des exemples de ce processus.

Si l'utilisateur ne souhaite pas utiliser l'une des variables ou n'est pas en mesure de le faire (en raison peut-être du manque de données disponibles), la variable peut être omise (en fixant toutes les valeurs de cette variable à 1,00). Naturellement, cela réduira la capacité prédictive du modèle.

ÉVALUATION

Dans cette section, nous discutons des implications pratiques des constatations présentées ci-dessus.

Le modèle permet l'ajustement des six variables, en plus de l'âge, du sexe et de la durée figurant dans les tables désignées. L'analyse effectuée ci-haut indique clairement qu'il est possible d'améliorer l'exactitude si les ajustements des variables Diagnostic et Province sont intégrés aux calculs de l'évaluation.

La variable Industrie pose cependant problème. À l'exception de l'administration publique, la plupart des catégories sont près de 1,00 pour toutes les durées. Pour des considérations pratiques, il serait utile d'intégrer cette variable dans Administration publique et Toutes les autres causes.

Pour les variables Montant des prestations et Délai de carence, l'amélioration de l'exactitude semble minime pour les réclamations individuelles et probablement inexistante pour un portefeuille de réclamations. Il est peu probable que l'ajout de ces variables permette de justifier la complexité accrue du processus de traitement et de validation des données. Il est évidemment utile de savoir que ces variables sont relativement peu importantes.

La variable Prestations préalables à l'ILD donne à penser que les réserves pour les réclamations seraient de 25 % inférieures dans le cas des prestataires qui reçoivent également des prestations d'ICD du même assureur. L'explication suggérée est que les assureurs des réclamations dans cette catégorie sont avantagés par un accès plus tôt aux données et la possibilité de gérer plus tôt les retours au travail.

Étant donné que cette hypothèse dépend des procédures internes de chaque assureur, il serait prudent que celui-ci reproduise cette constatation sur son propre bloc de polices avant d'adopter cet ajustement.

MISES EN GARDE

Les utilisateurs de l'étude devraient prendre connaissance des commentaires suivants :

1. Les tables désignées et les facteurs d'ajustement recommandés dans le présent rapport sont basés sur le nombre de vies et non sur le montant des prestations.
2. Les travaux présentés ici sont nouveaux dans le contexte de l'expérience de terminaison d'ILD au Canada et il n'existe pas de corpus de recherches publiées appuyant l'utilisation de la PBM pour établir les taux de terminaison prévus d'ILD en assurance collective.
3. Les facteurs recommandés ici sont conçus pour être utilisés conjointement avec les tables désignées et ils ont été validés en fonction de l'expérience globale de l'industrie. Ils peuvent ne pas être entièrement applicables si les tables désignées sont modifiées pour refléter un cas ou un portefeuille particulier.
4. Les facteurs recommandés ici sont conçus pour être utilisés avec les taux de terminaison globaux dans les tables désignées et n'ont pas été validés pour être utilisés seulement avec les taux de mortalité compris dans les tables désignées.
5. Le modèle recommandé utilise le résultat de plusieurs facteurs d'ajustement. Même si la PBM utilisée pour créer le modèle recommandé permet de minimiser le biais associé à la corrélation entre les variables prédictives, elle n'élimine pas nécessairement tous les biais.
6. Les facteurs du modèle (p. ex. la province ou l'industrie) ne doivent pas être utilisés directement comme facteurs d'ajustement dans un modèle de tarification sans tenir dûment compte de la corrélation possible entre les taux d'incidence et les taux de terminaison.

CONCLUSION ET RECOMMANDATIONS

1. Les actuaires en évaluation devraient envisager d'utiliser les modèles proposés conjointement avec les tables désignées récemment publiées.
2. Bien qu'une application intégrale de la version 2 du modèle procure le maximum d'avantages potentiels, il est recommandé d'appliquer au départ une version moins étoffée. La version 1 du modèle peut être utilisée pour simplifier les exigences administratives. En outre, toutes les variables de l'une ou l'autre version peuvent être exclues.
3. Lorsque l'ICA envisagera de mettre à jour les tables de terminaison pour la période de 2008 à 2015 récemment publiées, nous recommandons que les techniques d'analytique prédictive soient considérées comme une solution de rechange ou un complément à la méthode traditionnelle. Quelle que soit la méthode utilisée, nous recommandons d'intégrer, à tout le moins, les variables suivantes dans la table ou le modèle qui en résulte :
 - a. Âge;
 - b. Sexe;
 - c. Province (au lieu du Québec ou du Reste du Canada);
 - d. Diagnostic;
 - e. Industrie.

ANNEXE 1 – REVUE DE LA LITTÉRATURE

Voici un résumé de la revue de la littérature pertinente effectuée à ce jour :

Tables des taux de terminaison de l'industrie

Le taux de terminaison pour une durée particulière de réclamation depuis l'invalidité est calculé en fonction de facteurs (p. ex. âge et sexe) qui correspondent au profil de chaque réclamation. La méthode d'évaluation actuarielle de l'exposition fondée sur un vaste échantillon est utilisée pour estimer le taux de terminaison pour chaque durée d'un segment du profil des réclamations. Pour certaines tables, les taux de mortalité et de rétablissement sont séparés et un processus de graduation est utilisé.

- É.-U. : Commissioner's Disability Tables (*Tables du commissaire sur l'invalidité*, 64CDT); Commissioner's Individual Disability Table A (*Table A du commissaire sur l'invalidité individuelle*, 85CIDA); Society of Actuaries (SOA) 1987 Basic Group LTD Table (*Table sur l'ILD collective de base de la SOA de 1987*, 87GLTD); SOA Group Term Life Waiver Table (*Table sur l'exonération temporaire collective de la SOA*, 2005GLTD, 2008GLTD); 2004–2012 GLTD Database (*Base de données sur l'ILD collective 2004-2012*);
- Canada : Étude de l'expérience canadienne de cessation de l'ILD en assurance collective 1988-1994 de l'ICA, Étude de l'expérience canadienne de cessation d'ILD en assurance collective 1988-1997 de l'ICA;
- Royaume-Uni : Continuous Mortality Investigation Bureau (Table CMIR12);
- Australie : Institute of Actuaries of Australia (Table pour la période de 1989-1993 de l'IAD).

Analyse statistique

Un grand nombre d'études ont été menées à l'aide de méthodes d'analyse de la survie ou de la durée pour estimer la probabilité de la terminaison (décès ou rétablissement). Les méthodes vont des modèles non paramétriques (Kaplan-Meier), semi-paramétriques (régression à risque proportionnel de Cox) aux modèles paramétriques (régression accélérée de la durée de vie). Voici quelques références à souligner :

D. J. Doudna. « Effect of the Economy on Group Long Term Disability Claims », *Journal of Risk and Insurance*, 1977, 44(2), p. 223-235.

S. M. Mulla, S. Makosso-Kallyth, N. St-Hilaire, K. Munsch, P. B. Gove, D. Heels-Ansdell, G. H. Guyatt, et J. W. Busse. « Factors Associated with the Duration of Disability Benefits Claims among Canadian Workers: A Retrospective Cohort Study », 2017, *CMAJ Open*, 5(1), p. 109-115.

D. Pitt. « Modelling the Claim Duration of Income Protection Insurance Policyholders Using Parametric Mixture Models », *Annals of Actuarial Science*, 2007, 2(1), p. 1-24.

Modélisation prédictive

Diverses méthodes statistiques d'apprentissage et d'exploration des données ont été proposées afin de faire un usage plus judicieux des renseignements contenus dans le système d'information des assureurs, p. ex. *Predictive Modeling : A Modeler's Introspection* (SOA, juin 2015), *Predictive Analytics in Life and Health Insurance* (blogue de la SOA, février 2018) et, plus précisément, afin de prédire les taux de terminaison de l'invalidité, mais il n'y a que quelques études disponibles :

Mervyn Kopinsky. « Predicting Group Long Term Disability Recovery and Mortality Rates using Tree Models », Society of Actuaries, 2015. Disponible à <https://www.soa.org/experience-studies/2017/2017-gld-recovery-mortality-tree/>.

- Cette étude applique la méthode par arbre de régression à la base de données sur l'ILD collective de 2004 à 2012 de la SOA.

Q. Liu, D. Pitt, et X. Wu. « On the Prediction of Claim Duration for Income Protection Insurance Policyholders », *Annals of Actuarial Science*, 2014, 8(1), p. 42-62.

- On compare l'exactitude des prévisions selon plusieurs méthodes statistiques d'apprentissage, comme la régression linéaire, les analyses linéaires et quadratiques discriminantes, les régressions logistiques et la méthode des k plus proches voisins.

ANNEXE 2 – RECOURS À LA RÉGRESSION LOGISTIQUE POUR PRODUIRE LES TABLES DÉSIGNÉES

Les modèles de RL prédisent une variable dépendante binaire. Les variables indépendantes peuvent être continues ou catégoriques.

Bien qu'elle soit hors de la portée du présent projet, nous avons examiné l'utilisation de modèles de RL pour créer des tables de taux de terminaison de base.

La variable dépendante a été définie en fonction d'une réclamation ayant pris fin ou non (0 = réclamation en cours, 1 = réclamation ayant pris fin). Les variables indépendantes étaient les suivantes :

- Âge de début de l'invalidité;
- Sexe;
- Durée de la réclamation;
- Hors-Québec/Québec.

La durée d'une réclamation peut être incluse soit comme continue ou catégorique. Nous avons traité la durée des réclamations comme une variable catégorique pour deux raisons :

1. Par souci de cohérence avec l'étude des tables, qui présente les taux de terminaison annuels plutôt que mensuels après 60 mois.
2. Pour éviter un « lissage excessif », car les taux de terminaison toujours décroissants peuvent ne pas être appropriés.

Les conclusions suivantes sont tirées de l'analyse de la régression logistique :

- L'un des avantages de cette approche est que le modèle lisse de façon inhérente les données brutes et élimine la nécessité d'autres approches (p. ex. les moyennes mobiles) qui sont fastidieuses, requièrent le jugement de l'actuaire et ne sont pas de nature statistique.
- Les résultats sont statistiquement valables et fournissent des valeurs p et d'autres mesures de la qualité de l'ajustement.
- Les tables finales peuvent être construites à l'aide d'une seule équation, conjointement avec des variables qui varient selon les caractéristiques du prestataire.
- L'un des inconvénients de cette approche est qu'il est difficile de rendre compte de l'effet du changement dans la définition de l'invalidité pour une réclamation donnée. Nous croyons qu'il est possible de le faire en isolant des parties des données, mais nous n'avons pas examiné cette question puisqu'elle est hors de la portée de l'étude.

La régression logistique est une méthode viable de modélisation prédictive qui peut être examinée plus à fond dans de futures études.

ANNEXE 3 – CODAGE DES INDUSTRIES

Les sociétés participantes ont utilisé divers systèmes de codage.

L'étude des tables a utilisé le système de codage suivant, qui est fondé sur les deux premiers chiffres de la version canadienne du Système de classification des industries de l'Amérique du Nord (SCIAN) de 2008, avec les modifications indiquées.

Les codes 96, 97, 98 et 99 ne font pas partie du SCIAN.

- 96 – *Code non valide* signifie que le code soumis n'était pas valide dans le système de codage de l'assureur (probablement en raison d'une erreur de saisie des données).
- 97 – *Aucune donnée* signifie que les données soumises n'ont fourni aucune information sur l'industrie.
- 98 – *Correspondance impossible* signifie que les données soumises ont fourni des renseignements sur l'industrie, mais nous n'avons pas été en mesure d'établir de correspondance entre ces données et notre plan de codage.
- 99 – *Inconnu* signifie que les données soumises ont fourni des renseignements sur l'industrie, mais en indiquant que ce dernier était « inconnu ».

Pour chaque assureur, nous avons créé des tables de concordance entre les codes soumis et le plan de l'étude.

Table 9 : Codes des industries

Code	Description	Modification par rapport au SCIAN
11	Agriculture, foresterie, pêche et chasse	
21	Extraction minière, de pétrole et de gaz	
22	Services publics	
23	Construction	
31	Fabrication	Plus SCIAN 32, 33
41	Commerce de gros	
44	Commerce de détail	Plus SCIAN 45
48	Transport et entreposage	Plus SCIAN 49
51	Industrie de l'information et industrie culturelle	
52	Finance et assurances	
53	Services immobiliers et services de location et de location de bail	
54	Services professionnels, scientifiques et techniques	
55	Gestion de sociétés et d'entreprises	
56	Gestion des déchets	SCIAN 562 SEULEMENT, à l'exclusion de 561
61	Services d'enseignement	
62	Soins de santé	SCIAN 621, 622, 623
63	Services sociaux	SCIAN 624
71	Arts, spectacles et loisirs	
72	Services d'hébergement et services de restauration	

81	Autres services (sauf l'administration publique)	Plus SCIAN 561
91	Administration publique	
96	Code non valide	
97	Aucune donnée	
98	Correspondance impossible	
99	Inconnu	

Aux fins du modèle construit dans le cadre du projet, les catégories d'industrie ont été regroupées comme indiqué ci-dessous.

Table 10 : Correspondance des codes des industries

Catégorie d'industrie	Codes inclus
Industrie lourde	11, 21, 22, 23, 48, 56
Fabrication	31
Commerce de gros et de détail	41, 44
Travail de bureau et services professionnels	51, 52, 53, 54, 55
Santé, éducation, services sociaux	61, 62, 63
Autres services (secteur privé)	71, 72, 81
Administration publique	91
Inconnu	96, 97, 98, 99

ANNEXE 4 – CODAGE DES DIAGNOSTICS

La plupart des sociétés ont des systèmes de codage fondés sur la Classification internationale des maladies, 9^e révision (CIM-9). Trois sociétés utilisent la CIM-10 et quelques-unes ont des régimes exclusifs. Une société a fourni uniquement des descriptions en format texte libre et nous avons codé manuellement chaque enregistrement dans notre plan de codage.

Aux fins de la présente étude, nous avons créé le plan de codage suivant, qui est également fondé sur la CIM-9.

Table 11 : Codes des diagnostics

Code	Description
A	Maladies infectieuses et parasitaires
B	Tumeurs
C	Maladies endocriniennes, nutritionnelles et métaboliques et troubles immunitaires
D	Maladies du sang et des organes hématopoïétiques
E	Troubles mentaux
F	Maladies du système nerveux et des organes sensoriels
G	Maladies du système circulatoire
H	Maladies du système respiratoire
I	Maladies de l'appareil digestif
J	Maladies de l'appareil génito-urinaire
K	Complications liées à la grossesse, à l'accouchement et à la puerpéralité
L	Maladies de la peau et des tissus sous-cutanés
M	Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif
N	Anomalies congénitales
O	Certaines affections dont l'origine se situe dans la période périnatale
P	Symptômes, signes et états morbides mal définis
Q	Lésions traumatiques et empoisonnement
U	Inconnu
X	Aucune donnée
Y	Correspondance impossible

Comme on peut s'y attendre de dossiers de réclamations, on constate souvent une ambiguïté entourant la cause exacte de l'invalidité ou le code approprié à utiliser lorsque le prestataire souffre de plusieurs affections. Au besoin, nous avons exercé notre jugement. Nous n'avons pas demandé de données supplémentaires en plus de celles dans les dossiers de réclamations.

Table 12 : Correspondance des codes des diagnostics

Catégorie de diagnostic	Codes inclus
Troubles mentaux	E
Troubles du système musculo-squelettique	M
Néoplasmes (tumeurs, surtout cancers)	B
Troubles circulatoires	G
Maladies du système nerveux	F
Accidents	Q
Toutes les autres causes non déterminées	A, B, C, D, H, I, J, K, L, N, O, P
Non déclaré ou inconnu	U, X, Y, Z

ANNEXE 5 – RÉSUMÉ DES FACTEURS LIÉS AUX MODÈLES

Version 1

Toutes les durées

Industrie	Exposition	Facteurs
Industrie lourde	1 332 695	1,038
Fabrication	1 522 271	0,994
Commerce de gros et de détail	1 265 085	1,019
Travail de bureau et services professionnels	1 616 981	1,018
Santé, éducation, services sociaux	1 312 515	1,025
Autres services (secteur privé)	875 876	0,990
Administration publique	1 802 829	0,920
Inconnu	1 088 653	1,017
Total	10 816 905	

Version 2

Durée de moins et de plus de 36 mois

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
682 892	649 803	1,033	1,105
712 457	809 814	0,997	0,941
715 760	549 325	1,022	0,994
800 860	816 121	1,025	0,950
693 151	619 364	1,024	1,018
340 455	535 421	0,989	0,968
745 577	1 057 252	0,906	1,083
377 210	711 443	1,025	0,928
5 068 362	5 748 543		

Délai de carence	Exposition	Facteurs
0 à 3 mois	2 101 997	0,948
4 mois	4 772 036	1,018
5 à 6 mois	2 746 656	1,012
Plus de 6 mois	1 196 216	0,968
Total	10 816 905	

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
873 909	1 228 088	0,945	0,961
2 443 044	2 328 992	1,021	0,984
1 228 463	1 518 193	1,011	1,008
522 946	673 270	0,954	1,099
5 068 362	5 748 543		

Prestations préalables à l'ILD	Exposition	Facteurs
Notre régime d'ICD	1 936 321	1,181
Autre ou aucune	8 880 584	0,939
Total	10 816 905	

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
1 157 505	778 816	1,193	0,901
3 910 857	4 969 727	0,933	1,019
5 068 362	5 748 543		

Prestations mensuelles	Exposition	Facteurs
Inconnu	460 056	1,082
Moins de 1 499 \$	2 022 371	1,009
1 500 à 1 999 \$	2 102 488	0,975
2 000 à 2 499 \$	1 877 541	1,003
2 500 à 3 249 \$	2 111 827	1,018
Plus de 3 250 \$	2 242 622	0,969
Total	10 816 905	

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
280 099	179 957	1,080	1,080
752 597	1 269 774	1,003	1,087
917 803	1 184 685	0,974	0,991
926 853	950 688	1,002	1,012
991 775	1 120 052	1,017	1,009
1 199 235	1 043 387	0,976	0,893
5 068 362	5 748 543		

Diagnostic associé à la réclamation	Exposition	Facteurs
Troubles mentaux	3 072 438	1,026
Troubles du système musculo-squelettique	2 371 654	0,900
Néoplasmes (tumeurs, surtout cancers)	1 016 681	1,236
Troubles circulatoires	877 167	0,854
Maladies du système nerveux	1 200 137	0,526
Accidents	772 068	1,219
Toutes les autres causes non déterminées	1 336 964	1,049
Non déclaré ou inconnu	169 796	1,059
Total	10 816 905	

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
1 430 744	1 641 694	1,036	0,872
1 141 163	1 230 491	0,906	0,822
693 484	323 197	1,181	2,656
362 680	514 487	0,854	0,877
390 049	810 088	0,506	0,661
433 256	338 812	1,227	1,036
559 035	777 929	1,038	1,170
57 951	111 845	1,086	0,811
5 068 362	5 748 543		

Province	Exposition	Facteurs
Colombie-Britannique	1 286 554	0,999
Alberta	1 100 046	1,189
Saskatchewan	281 222	1,242
Manitoba	359 099	1,112
Ontario	4 299 717	0,966
Québec	2 352 523	0,976
Ailleurs au Canada	1 137 744	0,906
Total	10 816 905	

Exposition		Facteurs	
1 à 36 mois	Plus de 36 mois	1 à 36 mois	Plus de 36 mois
550 187	736 367	1,002	0,980
558 443	541 603	1,192	1,145
142 158	139 064	1,245	1,212
157 845	201 254	1,107	1,170
1 819 043	2 480 674	0,963	1,009
1 361 635	990 888	0,976	0,966
479 051	658 693	0,913	0,842
5 068 362	5 748 543		

ANNEXE 6 – EXEMPLES

Ces exemples sont également disponibles dans un [chiffrier Excel](#).

Exemple 1 – Taux de terminaison pour la durée de 18 mois

Renseignements sur le demandeur

De base

Résidence	Alberta
Sexe	Homme
Âge de début de l'invalidité	37

Supplémentaires

Industrie	Études
Délai de carence	4 mois
Prestations préalables à l'ILD	Aucune
Montant des prestations	2 200 \$
Diagnostic	Musculo-squelettique
Province	Alberta

Trouvez le taux de terminaison approprié dans les tables désignées : <http://www.cia-ica.ca/fr/publications/d%C3%A9tails-de-publication/219012A>

Tables – Reste du Canada,
hommes de 35 à 39 ans

Taux de terminaison non ajusté	0,04147
--------------------------------	---------

Trouvez les facteurs d'ajustement pertinents à l'annexe 5.

Version 2, moins de 36 mois

Industrie	1,024
Délai de carence	1,021
Prestations préalables à l'ILD	0,933
Montant des prestations	1,002
Diagnostic	0,906
Province	1,192

Multipliez tous les facteurs du modèle.

Facteur composite	1,056
-------------------	-------

Multipliez le taux de base par le résultat de la multiplication de tous les facteurs du modèle.

Taux de terminaison ajusté	0,043774
----------------------------	----------

Exemple 2 – Taux de terminaison pour la durée de 45 mois

Renseignements sur le demandeur

De base

Résidence	Québec
Sexe	Femme
Âge de début de l'invalidité	52

Supplémentaires

Industrie	Administration publique
Délai de carence	12 mois
Prestations préalables à l'ILD	Notre régime d'ICD
Montant des prestations	5 000 \$
Diagnostic	Système nerveux (maladie de Parkinson)
Province	Québec

Trouvez le taux de terminaison approprié dans les tables désignées :

<http://www.cia-ica.ca/fr/publications/d%C3%A9tails-de-publication/219012A>

Tables – Reste du Canada, hommes de 35 à 39 ans

Taux de terminaison non ajusté	0,00834
--------------------------------	---------

Trouvez les facteurs d'ajustement pertinents à l'annexe 5.

Version 2, plus de 36 mois

Industrie	1,083
Délai de carence	1,099
Prestations préalables à l'ILD	0,901
Montant des prestations	0,893
Diagnostic	0,661
Province	0,966

Multipliez tous les facteurs du modèle.

Facteur composite	0,611
-------------------	-------

Multipliez le taux de base par le résultat de la multiplication de tous les facteurs du modèle.

Taux de terminaison ajusté	0,0051
----------------------------	--------