



**Canadian
Institute
of Actuaries**

**Institut
canadien
des actuaires**

Biais et équité dans la tarification et la souscription des risques d'assurances IARD

Date : Avril 2023

Numéro du document : 223056

Table des matières

Sommaire.....	3
Section 1 : Objet, champ d'application et renvois.....	4
1.1 Objet.....	4
1.2 Champ d'application.....	5
1.3 Renvois.....	5
Section 2 : Enjeux historiques et évolution actuelle.....	5
Section 3 : Définitions.....	7
3.1 Biais.....	7
3.2 Discrimination directe et indirecte.....	9
3.3 Équité.....	9
3.4 Contraste entre biais et équité.....	10
3.5 Éthique.....	10
Section 4 : Analyse et mesure des enjeux.....	11
4.1 Aperçu.....	11
4.2 Mesures.....	11
4.2.1 Mesure des biais.....	11
4.2.2 But de la mesure.....	12
4.2.3 Autres considérations pertinentes.....	13
4.3 Cadre éthique.....	13
Section 5 : Prochaines étapes pour le ou la spécialiste.....	14
5.1 Documentation relative à l'évaluation des biais et de l'équité.....	15
5.1.1 Cadre.....	15
5.1.2 Considérations relatives à la documentation.....	15
5.2 Rapports dans le cadre de la gouvernance du modèle.....	17
Section 6 : Autres recherches.....	18
Section 7 : Exemple illustratif.....	18
Annexe 1 : Diverses définitions sur lesquelles nous nous appuyons.....	22
Annexe 2 : Exemple de cadres éthiques.....	23
Annexe 3 : Définitions générales du biais.....	26
Références citées.....	27

Sommaire

Le présent document vise à mieux faire comprendre aux spécialistes les biais et l'équité, les considérations qui les accompagnent et les outils qui permettent de les détecter, d'en faire l'évaluation et d'atténuer les biais éventuels dans les modèles actuariels d'évaluation des risques. Il vise à fournir des conseils aux spécialistes qui effectuent des travaux d'actuariat liés à la souscription de risques en assurances IARD, entre autres, la tarification et la modélisation. Dans l'ensemble du document, il est supposé que le lecteur ou la lectrice connaît très bien les méthodes de modélisation et de tarification des assurances IARD. Le public cible est composé des actuaires en tarification des assurances IARD. Toutefois, le contenu du document s'applique également à d'autres domaines de la pratique actuarielle.

Les questions historiques et les exemples se rapportant aux biais et à l'équité, comme l'évaluation des risques fondés sur le profilage racial par les services correctionnels, contextualisent les enjeux en cause.

Plusieurs définitions des biais sont possibles. Aux fins du présent document, le concept est défini comme une propriété mesurable des modèles prédictifs. L'équité quant à elle se rapporte à l'utilisation de ces modèles dans la pratique pour prendre des décisions ayant une incidence sur les personnes, par exemple en matière de sécurité financière et de protection contre les risques. Il n'existe pas de définition universelle de l'équité; comme pour les autres concepts liés à l'éthique, l'équité est constamment débattue et adaptée dans une démocratie. Comme l'écrit un érudit en matière d'intelligence artificielle dans un article sur l'équité en matière d'algorithmes et de pratique actuarielle, [traduction] « l'équité constitue un enjeu dynamique et social, mais non statistique »¹. Toutefois, cela ne signifie pas que l'équité ne peut être une dimension de l'évaluation des modèles actuariels. Tel que noté par la Financial Conduct Authority du Royaume-Uni, l'équité se mesure d'après les biais qu'elle peut engendrer.

L'évaluation des biais est la première étape pour régler tout problème de biais ou d'équité lié au cycle de vie de la souscription en assurances IARD. Les biais peuvent être attribuables à plusieurs facteurs : la collecte inappropriée de données, des modèles et des hypothèses inappropriés. Toutes ces sources doivent être prises en compte. De nombreuses méthodes d'évaluation sont possibles et elles peuvent servir à vérifier la présence de biais dans les résultats.

Deux types sont abordés dans le présent document : procédurale et distributive. De nombreuses techniques pour évaluer l'équité sont disponibles, incluant le Fairness Tree mis au point par le Center for Data Science and Public Policy de l'Université de Chicago. Le choix d'une mesure d'évaluation de l'équité est fait selon que l'objectif de l'évaluation est l'équité individuelle ou collective.

Il est recommandé que le ou la spécialiste établisse un cadre éthique pour déterminer ce qui constitue l'équité et les mesures pertinentes à prendre pour veiller à ce que les algorithmes et les modèles de tarification satisfassent à ces normes. Lorsque ce cadre éthique est établi, il est également recommandé que le ou la spécialiste documente les éléments suivants : l'identification des groupes vulnérables, le processus utilisé, l'échéancier d'évaluation, la méthode d'évaluation suivie, le choix des seuils, l'utilisation de procédures de validation, les interventions mises en place ou recommandées, les méthodes utilisées pour la surveillance de l'efficacité du modèle et la mise en œuvre de la tenue de dossiers. L'évaluation des biais pourrait également faire partie d'une gouvernance élargie du modèle interne.

Le présent document ne vise pas à servir de guide détaillé décrivant exactement comment régler tous les problèmes d'équité possibles que peut soulever un algorithme de tarification. Dans la pratique, il est important de tenir compte du contexte de chaque situation, et il n'existe pas de solution universelle pour assurer l'équité en tarification. Le document sert plutôt de point de départ afin que les spécialistes puissent mieux comprendre les concepts de biais et d'équité, sans tenter d'être exhaustif à ces égards.

¹ Joi Ito, « [Supposedly 'Fair' Algorithms Can Perpetuate Discrimination](#) », *Wired*, 5 février 2019.

Par conséquent, les recommandations ne sont ni normatives ni exécutoires; elles énoncent simplement des suggestions sur la façon d'assurer l'équité dans les algorithmes de tarification. Les recommandations soulignent certaines considérations importantes dans ce domaine, dans le but de sensibiliser davantage les spécialistes aux enjeux concernant les biais et l'équité dans le cadre du travail actuariel.

Les concepts énoncés dans le présent document sont étroitement liés à la section 1400 des Normes de pratique.

Remerciements

Le présent document a été préparé par les membres de la Sous-commission sur les biais et l'éthique de la Commission sur la tarification des assurances IARD soit : Qi An (co-présidente), Carl Lussier (co-président), Joshua Snow, Erik Christianson, Alexandre Monette-Pagny et Alina Rogozhnikova.

Commentaires

Prière d'adresser vos commentaires concernant le présent document à pratique@cia-ica.ca.

Section 1 : Objet, champ d'application et renvois

1.1 Objet

De nombreuses discussions tenues au cours des dernières années ont porté sur le risque de biais dans les algorithmes de tarification. Divers documents ont traité des biais et de l'équité dans les domaines de la tarification et de la modélisation prédictive, en particulier les méthodes utilisées par les scientifiques des données.

Nous rappelons au lecteur ou à la lectrice qu'une prime d'assurance est calculée en fonction d'un taux par unité d'exposition². Le taux est déterminé en fonction de plusieurs variables de tarification différentes, dont les usages sont régis par des lois fédérales et provinciales pour assurer la transparence et l'équité.

À mesure que le volume de données disponibles augmente et que les algorithmes de tarification deviennent de plus en plus complexes, le secteur des assurances IARD s'en remet davantage à des processus automatisés, à des modèles et à des techniques d'apprentissage automatique pour calculer les primes. Les modèles et les processus automatisés sont adaptés à plusieurs ensembles de données qui déterminent la prime en fonction d'un groupe de caractéristiques d'intrants. L'ajustement est ensuite affiné davantage en fonction de la précision de la prédiction des modèles. Ceci peut entraîner un risque de biais indésirable dans le produit du travail du ou de la spécialiste.

Bien que les algorithmes axés sur les données semblent s'appuyer sur des données objectives, ils reposent encore fortement, de façon générale, sur des décisions subjectives en ce qui concerne la façon dont ces données devraient être analysées (p. ex. les caractéristiques qui devraient être incluses et les catégories d'observations qui devraient être triées, etc.). De plus, la fréquence à laquelle différents groupes ou catégories apparaissent dans les données peut avoir des effets involontaires importants sur les résultats prévus. L'article intitulé « Understanding Bias in Algorithmic Design » par ASME Demand souligne ces enjeux à l'aide de différents exemples, dont un problème présent dans la recherche sur LinkedIn, où l'algorithme semble présenter une préférence pour les noms masculins par rapport aux noms féminins³. Bien que LinkedIn ait répondu que le biais apparent quant au sexe est attribuable au fait que la fonction de recherche est basée sur la fréquence des mots – puisque les hommes sont plus

² Claudine Modlin, Willis Towers Watson et Geoff Werner, *Basic Ratemaking*, 5^e édition, Casualty Actuarial Society, mai 2016.

³ A.R. Lange et Natasha Duarte, « [Understanding Bias in Algorithmic Design](#) », *ASME Demand*, printemps 2017.

susceptibles d'avoir des noms communs que les femmes – il en découle néanmoins un avantage professionnel potentiel involontaire pour les hommes.

Le présent document a pour but de donner aux spécialistes une meilleure compréhension des biais et de l'équité et de leur fournir des outils pour les évaluer dans le contexte de la tarification actuarielle et de la modélisation.

Nous reconnaissons qu'un traitement intégral de l'équité dépasse largement la portée du présent document, car l'équité est un concept dynamique qui est relatif au contexte pour lequel il est envisagé. Pour cette raison, le document met davantage l'accent sur les biais que sur l'équité en limitant la portée aux applications des concepts connexes dans la pratique par le biais de mesures quantitatives.

1.2 Champ d'application

Nous visons à fournir aux actuaires des outils, des conseils et des considérations auxquels ils ou elles doivent réfléchir lorsqu'ils ou lorsqu'elles offrent divers services, notamment dans le cadre des tâches suivantes :

1. la préparation de segments ou de niveaux de risque;
2. l'évaluation des écarts de prix, des rabais et des surprimes;
3. l'analyse prédictive d'autres types pour déterminer le niveau des coûts périodiques ou le potentiel de croissance;
4. tout autre modèle pour lequel l'actuaire juge ces concepts applicables.

Le présent document ne s'applique pas à la détermination sociétale des notions d'équité.

Il doit être envisagé dans son ensemble. Bien que chaque section soit importante, les concepts abordés doivent être considérés de façon globale.

1.3 Renvois

Lorsque le présent document renvoie aux dispositions d'autres documents, le renvoi comprend les documents cités en référence dans la mesure où ils pourraient être modifiés ou reformulés à l'avenir, ainsi que tout document qui leur succéderait, quel que soit leur titre. Si un document modifié ou reformulé diffère sensiblement du document de référence initial, l'actuaire devrait tenir compte des conseils énoncés dans le présent document dans la mesure où ils demeurent applicables et pertinents.

Section 2 : Enjeux historiques et évolution actuelle

L'équité peut être définie comme un concept social dont la manifestation distincte peut être évaluée au moyen de diverses mesures quantitatives. Ce qui est équitable ou inéquitable évolue au fil du temps et dans divers contextes sociétaux, de sorte que les spécialistes devront adapter leur interprétation du concept aux circonstances de son application.

Dans les travaux d'assurance traditionnels, les enjeux entourant l'équité dans la société et l'assurance sont présents sur le spectre du potentiel de segmentation des risques, à partir de l'absence de segmentation liée au risque jusqu'à une vision extrêmement détaillée de la segmentation. Cela s'explique par le fait que le processus de segmentation en assurance tente de reconnaître les caractéristiques de risque qui visent à différencier le niveau de risque, ou la probabilité du résultat, entre les personnes, puis à fixer un prix pour la couverture.

Les processus de collecte de données et d'élaboration de modèles peuvent entraîner des biais dans le processus décisionnel automatisé et même perpétuer des résultats injustes. L'exemple suivant a été documenté dans le journal *The Globe and Mail* (TGAM) comme étant intrinsèquement biaisé dans l'application des modèles de notation pour prédire les résultats potentiels des détenus et ainsi

discriminatoire pour un segment de la population. Il est cité ici pour illustrer une application hors du domaine de l'assurance qui pourrait donner lieu à un résultat biaisé intrinsèquement injuste.

À la fin de 2020, TGAM a publié une série d'articles intitulés « [Bias Behind Bars : A Globe Investigation Finds a Prison System Stacked Against Black and Indigenous Inmates](#) »⁴ qui faisaient référence à l'utilisation par le Service correctionnel du Canada d'une évaluation des risques pour établir la cote de sécurité des détenus dans les prisons fédérales, évaluer leur potentiel de réinsertion sociale et déterminer quels programmes leur seraient offerts. L'évaluation des risques comportait plusieurs éléments, qui débouchaient prétendument sur une note actuarielle se traduisant par un résultat attendu.

TGAM a constaté un biais systémique dans le calcul de la note, ce qui a entraîné des résultats plus défavorables pour une partie de la population carcérale. En particulier, les détenus autochtones et noirs étaient beaucoup plus susceptibles d'obtenir une cote de sécurité « maximale » que les détenus de race blanche, une classification qui a eu une incidence sur leur accès aux programmes de traitement et entraîné des notes plus faibles au chapitre de la réinsertion sociale et ainsi, a augmenté la probabilité d'incidence négative sur les décisions concernant leur libération conditionnelle.

TGAM a ajouté que l'évaluation du risque entraînait non seulement des peines carcérales plus défavorables pour les personnes racialisées mais semblait être basée sur des données inexactes :

[traduction] Au cours de la période de sept ans pour laquelle TGAM a obtenu des données, les hommes autochtones et noirs sont moins susceptibles que les hommes blancs de commettre un nouveau délit, et notre analyse suggère que ces notes surestiment souvent la probabilité qu'ils se retrouveront de nouveau en prison. Selon les experts, ce résultat crée une boucle de rétroaction criminelle, avec des notes négatives qui haussent les probabilités de récidive à long terme de la part d'un détenu, ce qui accroît en fin de compte la surreprésentation flagrante des Autochtones et des Noirs en détention dans les établissements fédéraux.

L'exemple ci-dessus illustre clairement comment la boucle, à partir des données jusqu'aux résultats, perpétue la discrimination systémique. On utilisait (et au moment d'écrire ces lignes), on utilise encore un modèle qui prédit un résultat non étayé par les données des sept dernières années. Le rapport de TGAM appuyait une série antérieure de conclusions du Sénat du Canada quelques années plus tôt⁵.

Le prochain exemple a trait à l'assurance. Si nous adoptons une approche essentiellement technique, nous pourrions soutenir que les habitants des communautés plus pauvres devraient payer beaucoup plus cher pour l'achat d'assurance, compte tenu de la nature du risque mesuré d'un point de vue purement mathématique basé sur le marché⁶. Une pratique discriminatoire selon laquelle les résidents et les résidentes des quartiers « indésirables », souvent définis ou coïncidant avec leur composition raciale ou ethnique, se voient refuser des services financiers, incluant la couverture d'assurance, était basée sur cette interprétation purement mathématique de l'équité et de l'exactitude. À l'instar des cotes de récidive de l'exemple précédent qui créaient une prophétie de récidive accrue, la boucle de rétroaction créée par les pratiques discriminatoires formelles et informelles aux É.-U. et au Canada a conduit à un état de délabrement dans ces communautés et mené à un renforcement de la justification technique.

⁴ Tom Cardoso, « [Bias Behind Bars: A Globe Investigation Finds a Prison System Stacked Against Black and Indigenous Inmates](#) », *The Globe and Mail*, octobre 2020.

⁵ Sénat du Canada, *Étude concernant les droits de la personne des prisonniers dans le système correctionnel fédéral*, Rapport provisoire du Comité sénatorial permanent des droits de la personne, février 2019, https://sencanada.ca/content/sen/committee/421/RIDR/reports/RIDR_Report_Prisoners_f.pdf.

⁶ Ibid, note 1 ci-dessus.

Un deuxième exemple lié à l'assurance concerne l'utilisation du genre dans les algorithmes de tarification. Bien que la différenciation des prix selon le sexe soit permise au Canada en vertu de la Charte, ce n'est pas le cas dans toutes les lois provinciales, et elle était source de litige jusqu'au jugement de la Cour suprême du Canada (CSC) à cet égard⁷. Même si la CSC a jugé que les assureurs peuvent poursuivre l'utilisation du genre en assurance à titre de facteur de risque dans certains contextes, cette décision n'était pas unanime, deux juges étant en désaccord. Il convient de noter que de l'autre côté de l'Atlantique, l'Union européenne a interdit l'utilisation du genre en 2012, du fait qu'il était intrinsèquement injuste.

Nous estimons que la section suivante du jugement de la CSC dans *Zurich Insurance Co. c. Ontario* (Commission des droits de la personne) est pertinente sur le plan contextuel, même quelques décennies après sa rédaction initiale :

Les droits de la personne ne peuvent être écartés pour des raisons uniquement commerciales. Autoriser une pratique discriminatoire "statistiquement défendable" aurait pour effet de porter atteinte au but des lois sur les droits de la personne, qui tentent de protéger les personnes contre la faute collective. Une telle pratique ne ferait que perpétuer les stéréotypes traditionnels avec tous leurs préjugés insidieux. Il faut déterminer s'il existe une autre solution pratique applicable dans toutes les circonstances.

Ces exemples sont fournis à titre d'illustration de cas où une notion d'équité s'appliquait et produisait parfois des conséquences imprévues. Le dernier exemple est particulièrement pertinent, car la CSC a fourni une justification de son jugement et des considérations (mais avec dissidence) en ce qui concerne ce à quoi pourrait ressembler l'équité dans le contexte de la tarification d'assurance. Tel que mentionné précédemment, le concept d'équité est, de par sa nature, vaste et dynamique et donne lieu à différentes interprétations et points de vue. Il s'agit d'un concept qui évolue dans la société dans laquelle il est employé.

Des discussions récentes des biais et de l'équité en tarification de l'assurance aux États-Unis comprennent :

[Study Points to Rate Bias in U.S. Auto Insurance Industry](#)

[NAIC – Innovation and Technology \(EX\) Task Force](#)

[NAIC – Race and Insurance](#)

Section 3 : Définitions

La présente section renferme les principales définitions utilisées dans le document, à savoir « biais », « équité » et « éthique ».

Le lecteur ou la lectrice doit garder à l'esprit que les définitions ci-dessous peuvent être considérées comme étant plus normatives que celles que l'on retrouve habituellement dans la littérature. Le but est de mettre en contraste les diverses considérations présentes lorsque ces sujets sont discutés.

3.1 Biais

Il est important de noter qu'il existe plusieurs méthodes d'évaluation et de définition du biais. Par conséquent, le ou la spécialiste devrait tenir dûment compte de la définition à appliquer. Toutefois, dans

⁷ *Zurich Insurance Co. c. Ontario* (Commission des droits de la personne), [1992] 2 RCS 321. Disponible à <https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/895/index.do>.

un souci de clarté, le groupe de travail considère que la définition présentée dans le projet de loi C-27 *Loi sur l'intelligence artificielle et les données* constitue un bon fondement sur lequel débiter une discussion :

Résultat biaisé Contenu généré, prédiction ou recommandation faite ou décision prise par un système d'intelligence artificielle qui défavorisent, directement ou indirectement et sans justification, un individu sur le fondement d'un ou plusieurs motifs de distinction illicite prévus à l'article 3 de la Loi canadienne sur les droits de la personne, ou de leur effet combiné⁸.

Bien que le projet de loi C-27 vise les responsables des systèmes d'intelligence artificielle (IA) – avec ses obligations plus étendues réservées aux systèmes IA à incidence élevée –, sa définition de biais met en lumière les exigences générales que doit respecter tout modèle prédictif qui classe les personnes dans des catégories, comme des catégories de risque. Autrement dit, si un modèle produit des résultats qui varient selon le groupe, il faut pouvoir justifier ce traitement inégal en se référant à une certaine différence marquante entre les groupes. En vertu de la *Loi canadienne sur les droits de la personne*, le projet de loi interdit l'utilisation de caractéristiques, par exemple la race ou la situation d'incapacité, comme éléments de modèles prédictifs, sauf dans le cas où une telle utilisation permettrait de prévenir ou de supprimer le désavantage fondé sur un motif de protection ou lié à celui-ci.

Donc, aux fins du présent document, nous définissons le biais comme suit :

S'entend de biais dans la tarification des assurances IARD toute situation où les résultats des modèles de tarification sont systématiquement moins favorables pour les personnes d'un groupe particulier et où il n'y a pas de différence pertinente entre les groupes justifiant un tel écart de primes ou de taux.

En d'autres termes, un résultat est biaisé lorsqu'il entraîne pour des personnes ou des groupes des primes plus élevées ou plus basses qui ne sont pas justifiées par les écarts de coût de l'assurance. Dans la tarification des assurances IARD, cette justification prend généralement la forme d'une corrélation statistique entre une variable utilisée pour la classification des risques, comme un signal territorial, et le risque sous-jacent. Plus rarement, il est possible de comprendre le mécanisme causal sous-jacent expliquant comment une caractéristique déclenche un processus qui mène à une augmentation des risques et des coûts, mais ce niveau plus élevé est souvent difficile à atteindre dans le domaine des assurances IARD. Prenons la tarification territoriale en exemple. Il suffit habituellement de montrer qu'une variable de tarification est prédictive de sinistres ou de dépenses, et les spécialistes corrèleront généralement plusieurs sources d'information (p. ex., données de recensement, données routières, statistiques sur la criminalité, densité de la circulation, conditions météorologiques, mesures de proximité, statistiques de l'environnement et expérience de l'industrie) pour obtenir un signal corrélé du coût des sinistres pour l'unité géographique⁹. Dans ce cas, bien qu'il soit difficile de comprendre exactement pourquoi un point géographique particulier est plus risqué que tout autre point sur une carte, le résultat modélisé devrait néanmoins suivre la sinistralité ou une justification devrait être fournie expliquant pourquoi l'écart est approprié. Le raisonnement peut être étendu aux dimensions non géographiques de la tarification et de la souscription.

Il est important de faire la distinction entre la ou les définitions du biais ci-dessus et la définition statistique de biais, cette dernière faisant référence à une situation où la valeur attendue d'un estimateur diffère de

⁸ Projet de loi C-27, [Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois](#), 44^e législature, 1^{re} session, 2022.

⁹ Pinnacle Actuarial Services, [Rapport d'évaluation de la tarification territoriale de Pinnacle](#), rapport spécial pour l'Autorité ontarienne de réglementation des services financiers, octobre 2022, 10 (en anglais seulement).

la valeur sous-jacente réelle. Dans le présent document, le terme « biais » ne renvoie pas à la définition statistique, lequel devrait être considéré comme un concept non lié.

Le ou la spécialiste peut tenir compte des deux définitions dans son travail, mais il ou elle devrait vraisemblablement choisir la définition qui convient le mieux au travail qu'il ou qu'elle a entrepris.

3.2 Discrimination directe et indirecte

À l'heure actuelle, est-il besoin de rappeler que les assureurs sont assujettis à la législation sur les droits de la personne et ne peuvent pas utiliser certaines variables (p. ex., race, situation d'incapacité, orientation sexuelle et autres caractéristiques associées à la discrimination systémique ou à la marginalisation) pour la classification des risques. Toutefois, un problème plus grave surgit, c'est-à-dire la possibilité que des données apparemment neutres se substituent à l'appartenance à un groupe protégé, intentionnellement ou non, en saisissant cet état de façon détournée. L'évolution et l'adoption accélérées de l'IA soulignent l'urgence du défi posé par la discrimination indirecte ou par procuration dans la tarification, et un corpus croissant de recherches examine les conséquences éthiques et réglementaires à ce sujet. Le présent document ne se prononce pas sur ces discussions importantes et ne formule pas de recommandations à l'intention des organismes de réglementation, des décideurs ou des entreprises qui cherchent à élaborer des politiques internes. Il offre plutôt des outils et des conseils pratiques aux spécialistes des assurances IARD qui souhaitent adopter une approche réfléchie et appropriée sur le plan actuariel à l'égard de ces questions dans le cadre de leur travail quotidien.

Pour clore la discussion sur le biais, nous considérons donc les concepts de discrimination suivants comme étant pertinents pour les spécialistes :

- Un modèle de tarification évite la discrimination directe si aucune caractéristique discriminatoire (c.-à-d. caractéristiques protégées par la législation en matière des droits de la personne comme la *Loi canadienne sur les droits de la personne*) n'est utilisée comme facteur de tarification.
- Un modèle de tarification évite la discrimination indirecte s'il évite la discrimination directe et, de plus, les caractéristiques non-discriminatoires sont utilisées de façon qu'elles ne permettent pas l'inférence implicite des caractéristiques discriminatoires à partir de ces caractéristiques non-discriminatoires¹⁰.

Il convient de noter que le groupe de travail a constaté que les exigences d'inférence implicite ci-dessus peuvent être satisfaites tout en obtenant des résultats différents entre les groupes.

3.3 Équité

Il n'existe pas de définition ou de mesure universelle de l'équité. Tel que mentionné ci-dessus, le ou la spécialiste doit comprendre que « l'équité est un enjeu dynamique et social, et non statistique »¹¹.

Lorsqu'ils examinent l'équité, les spécialistes devraient tenir compte des questions suivantes¹²:

- Qui est lésé par le biais de tarification ou le risque de biais?
- Dans quelle mesure les personnes sont-elles lésées?
- Quelle est l'ampleur du bassin de personnes lésées?
- Le produit ou le service est-il essentiel?
- La société considère-t-elle la discrimination des prix comme flagrante/socialement injuste?

¹⁰ M. Lindholm, R. Richman, A. Tsanakas et M.V. Wüthrich, « [Discrimination-Free Insurance Pricing](#) », *ASTIN Bulletin: The Journal of the IAA* 52, v. 1 (2021): 55-89.

¹¹ Ibid, note 1 ci-dessus.

¹² Adapté de Mary Starks et coll., « [Price Discrimination in Financial Services](#), » Financial Conduct Authority, juillet 2018.

Dans les scénarios où un biais pourrait causer une injustice et un préjudice important, compte tenu des considérations susmentionnées, les spécialistes devraient faire preuve de prudence accrue, plus particulièrement lorsqu'il y a risque de perpétuer les biais existants ou de rendre les groupes protégés plus vulnérables aux inégalités sociétales existantes.

L'équité peut habituellement être divisée en deux catégories différentes, soit procédurale et distributive. Aux fins de la tarification; les spécialistes devraient essayer de les identifier et de les examiner séparément. La réglementation actuelle en matière d'assurance tend à mettre davantage l'accent sur l'équité procédurale que sur l'équité distributive.

- L'équité procédurale désigne la façon dont les assurés et les assurées sont traité(e)s pendant tout le processus de tarification. Les décisions relatives à la façon de traiter les valeurs manquantes dans les données ou les variables à inclure dans un modèle de tarification entreraient dans la catégorie de l'équité procédurale.
- L'équité distributive a trait à la répartition des résultats de la tarification entre divers assurés et assurées. Elle n'est pas liée au processus de tarification, mais plutôt aux résultats du processus et à son incidence sur les assurés et les assurées.

3.4 Contraste entre biais et équité

Il est important de noter que biais et équité ne sont pas des synonymes. Dans le contexte de la tarification des assurances IARD, le biais n'implique pas nécessairement une injustice; de même, un manque d'équité n'implique pas nécessairement un biais.

Le biais en tarification découle de facteurs comme les données utilisées, les paramètres du modèle, le type de modèle choisi et les hypothèses du ou de la spécialiste. À l'inverse, l'équité dépend à la fois du résultat d'un modèle et du contexte dans lequel ce résultat est appliqué. Cela comprend de nombreux facteurs qui sont externes au résultat du modèle et l'interaction de ces facteurs avec ledit résultat.

Dans ce contexte, le biais sera un concept statique. Un processus, un modèle ou un résultat qui est biaisé aujourd'hui le sera également au fil du temps (à moins d'être corrigé). Toutefois, en raison de la nature dynamique de l'équité, un résultat qui est juste aujourd'hui peut ne pas l'être à l'avenir.

3.5 Éthique

Le préambule des Règles de déontologie de l'ICA indique que les règles énoncées « précisent les normes professionnelles et d'éthique auxquelles les membres doivent se conformer et ainsi servir l'intérêt public. ». Le Code ajoute qu'« en plus des règles, les membres sont assujettis à la loi applicable et aux règles de déontologie ou aux normes d'éthique promulguées par un organisme actuariel reconnu dans les juridictions où ils rendent des services professionnels. »

Le respect de la loi

Les spécialistes ont une obligation non équivoque d'observer la loi, ce qui comprend les lois en matière de groupes protégés qui définissent les caractéristiques dont l'utilisation est interdite comme variables pour la classification (p. ex., le motif de distinction illicite à l'article 15 de la *Charte canadienne des droits et libertés*). Le ou la spécialiste doit tenir compte des répercussions juridiques et des divers cadres, équitables ou juridiques, qui sont en place dans les multiples administrations pour s'assurer que le travail est effectué dans le cadre applicable.

Par exemple, la discrimination fondée sur les conditions sociales est interdite au Québec, celle reposant sur l'âge est interdite au Nouveau-Brunswick, et la discrimination imputable à la cote de crédit et à la richesse est interdite en Ontario pour la tarification ou la souscription de l'assurance automobile.

Toutefois, le même élément d'information pourrait être considéré comme légal pour une autre province ou pour une série de produits différente.

Le respect et l'intention de la loi incombent aux spécialistes.

Nous recommandons aux spécialistes de se familiariser avec les Règles de déontologie de l'ICA et toutes les lois applicables. Les principes éthiques dont il est question à la section 4 ci-dessous servent à mieux comprendre comment appliquer ces lois et ces normes éthiques, mais ils ne remplacent pas les exigences juridiques et professionnelles existantes.

Section 4 : Analyse et mesure des enjeux

4.1 Aperçu

Le biais peut être introduit, identifié et mesuré à chaque étape du processus de tarification, de la phase de collecte des données, à la préparation des données, à la modélisation, aux prédictions, etc. La section 4.2 ci-dessous décrit les outils que les spécialistes peuvent utiliser pour détecter l'existence d'un biais potentiel et déterminer si ce biais peut être jugé « injuste » selon les normes juridiques, éthiques et professionnelles tel que discuté précédemment. Nous recommandons aux spécialistes de vérifier le biais à chaque point de contrôle, en utilisant des mesures pertinentes, souvent multiples, qui s'appliquent au travail exécuté.

Lorsque le ou la spécialiste a déterminé qu'il existe un biais potentiel (tel que défini à la section 3.1), le ou la spécialiste doit tenir compte du compromis nécessaire entre l'exactitude et l'équité (tel que défini à la section 3.2) lorsqu'il ou lorsqu'elle décide d'un plan d'action pour le corriger. Les deux dernières sections (« Cadre éthique » et « Procédures générales ») visent à fournir au ou à la spécialiste les outils nécessaires pour évaluer la notion d'équité et les coûts sociaux possibles.

4.2 Mesures

4.2.1 Mesure des biais

Il existe de multiples sources de biais dans la tarification. En voici des exemples :

- Les processus de production de données, y compris l'information contextuelle concernant les circonstances ou les conditions sociales de laquelle proviennent les données (c.-à-d. les données peuvent refléter des inégalités historiques).
 - Un exemple, décrit plus amplement à la section 2, concerne les assureurs qui hésitent à offrir des garanties d'assurance aux résidents et aux résidentes de certains quartiers. Ceci entraîne un manque d'entretien des propriétés qui est ensuite intégré à l'approche technique utilisée par les actuaires pour déterminer le coût des sinistres, ce qui perpétue la discrimination.
- Des données incomplètes ou non représentatives (p. ex. lorsque les données ne couvrent pas l'espace du domaine).
 - Un exemple serait la collecte de données uniquement auprès des personnes qui se portent volontaires pour participer.
- Des modèles ou algorithmes inappropriés.
 - Un exemple serait l'utilisation de modèles linéaires pour tenter d'expliquer la relation non linéaire entre les caractéristiques et les observations.
- Des hypothèses inappropriées.
 - Un exemple serait une hypothèse inappropriée au sujet de l'ordre des variables des catégories.

Les spécialistes devraient tenir compte de toutes les sources de biais susmentionnées ainsi que de toute autre source dont ils sont au courant.

Lorsqu'ils évaluent un modèle de tarification pour détecter un biais, les spécialistes doivent examiner l'ensemble des extrants produits par le modèle de tarification pour vérifier les résultats anormaux. Les mesures suivantes peuvent être envisagées pour déterminer si le modèle de tarification comporte un biais :

- les différences entre les valeurs moyennes prévues d'une catégorie à l'autre;
- les ratios des valeurs moyennes prévues entre les catégories;
- la différence d'exactitude, c'est-à-dire le nombre de prédictions corrigées par rapport au nombre total de prédictions, entre les catégories;
- l'écart du taux d'erreur, soit le nombre de prédictions non corrigées par rapport au nombre total de prédictions, entre les catégories;
- l'impact disparate, qui correspond au ratio des valeurs prédictives pour l'ensemble des catégories.

Dans un document de 2022 de la Casualty Actuarial Society intitulé « Methods for Quantifying Discriminatory Effects on Protected Classes in Insurance », Roosevelt Mosley et Radost Wenman classent différentes mesures d'équité en trois catégories : indépendance, séparation et suffisance¹³. L'indépendance pose la plus forte exigence et elle mesure les similitudes des prédictions entre les catégories. La séparation pose une exigence plus faible en nous permettant de conditionner les mêmes valeurs observées avant de comparer les valeurs prévues entre les catégories. Enfin, la suffisance nous permet de conditionner les mêmes valeurs prédictives avant de comparer les valeurs observées entre les catégories. Le document cite des exemples de mesures dans chaque catégorie et les illustre également dans le contexte de la tarification.

4.2.2 But de la mesure

Pour déterminer la méthodologie à utiliser pour mesurer le biais, le ou la spécialiste peut décider du type d'équité qu'il ou qu'elle tente d'atteindre.

- L'équité individuelle fait en sorte que des personnes semblables soient traitées de la même façon.
- L'équité collective signifie s'assurer que certaines mesures statistiques sont égales d'un groupe à l'autre (c.-à-d. les catégories).

Si l'objectif consiste à atteindre l'**équité individuelle**, les mesures comprennent divers types de différences ou ratios de moyennes entre des échantillons particuliers de différentes catégories.

Si l'objectif consiste à atteindre l'**équité collective**, les mesures comprennent les différences, ou les ratios, des moyennes entre les groupes (c.-à-d. les catégories) proprement dits.

Comme le biais est pris en compte pour s'assurer que nous ne renforçons pas les inégalités pour certains groupes, les spécialistes sont encouragés à atteindre l'équité collective. Il convient de noter que le point de vue du ou de la spécialiste sur la véritable relation entre les groupes peut avoir une incidence sur la mesure pertinente du biais. Par exemple, si le ou la spécialiste estime que tous les groupes ont des capacités semblables en ce qui concerne la tâche sur laquelle ils sont évalués, même si cela ne peut être observé, des mesures de parité démographique au niveau du groupe devraient être utilisées. Si le ou la spécialiste estime que ce qu'il ou qu'elle observe reflète la capacité réelle de chaque groupe en ce qui

¹³ Roosevelt Mosley et Radost Wenman, « [Methods for Quantifying Discriminatory Effects on Protected Classes in Insurance](#) », Casualty Actuarial Society Research Paper Series on Race and Insurance Pricing, mai 2022.

concerne la tâche, alors il conviendrait d'utiliser des mesures au niveau individuel, comme la moyenne des différences de prévision des risques avec les mêmes valeurs observées pour tous les groupes.

De nombreux outils d'évaluation de source ouverte comme le cadre d'IA Fairness 360 mis au point par IBM, peuvent être utilisés par les spécialistes et autres chercheurs et chercheuses pour partager et évaluer le biais lié aux données des algorithmes.

Une autre considération importante pour déterminer le type d'équité à atteindre consiste à savoir si nous cherchons à obtenir une représentation juste ou un résultat juste. Un outil de prise de décision bien connu a été le Fairness Tree mis au point par le Center for Data Science and Public Policy de l'Université de Chicago, qui fournit de l'aide en posant des questions et dirige les utilisateurs et les utilisatrices vers différentes branches de l'arbre en fonction de leurs réponses afin de déterminer les paramètres d'évaluation correspondants.

4.2.3 Autres considérations pertinentes

Le ou la spécialiste peut consulter les sous-sections suivantes des Normes de pratique de l'ICA (NP) et de l'American Academy of Actuaries (ASOP) pour une plus ample discussion sur les biais et l'équité :

- NP 1440 – Suffisance et fiabilité des données
- NP 1450 – Pertinence et limites du modèle
- NP 1460 – Processus d'assurance de la qualité
- ASOP 12 – Classification des risques
- ASOP 25 – Crédibilité

4.3 Cadre éthique

Examinons l'exemple suivant :

Les techniques de mesure énoncées à la section 4.2 ont permis au ou à la spécialiste de déceler un biais dans un modèle de tarification. Les primes produites par le modèle favorisent par inadvertance les personnes d'une race par rapport à une autre en raison de différences de prix fondées sur le facteur de tarification lié au code postal.

Le ou la spécialiste a conclu que la race n'était pas utilisée comme facteur de notation lorsque les taux ont été établis, et que les différences fondées sur l'emplacement géographique sont attribuables aux différences dans la fréquence et la gravité des sinistres – ce qui permet de saisir avec exactitude la nécessité de primes plus élevées ou plus faibles. Toutefois, même involontaire, la différenciation de la prime en fonction de la race est jugée injuste selon la définition de l'équité dont il est question dans le présent document.

Le ou la spécialiste dispose de plusieurs plans d'action, qu'il s'agisse de corriger immédiatement le biais de tarification ou de ne rien faire.

Dans de telles situations, nous recommandons au ou à la spécialiste d'établir un cadre éthique pour évaluer les notions d'équité et les répercussions possibles de manière cohérente, vérifiable et justifiable.

Le cadre éthique vise à permettre au ou à la spécialiste de mieux comprendre comment appliquer les principes éthiques pour déterminer ce qui constitue l'équité et l'objectivité lorsqu'il ou lorsqu'elle examine les résultats produits par les modèles ou les algorithmes de tarification. L'annexe 2 fournit des exemples de théories éthiques clés, de cadres éthiques correspondants et de la façon dont ils varient lorsqu'ils sont appliqués.

Nous constatons que les exemples de l'annexe 2 ne sont pas exhaustifs et que le ou la spécialiste doit déterminer l'approche qui convient le mieux pour régler les problèmes propres à son travail.

Un processus qui peut être suivi pour appliquer un cadre éthique est décrit ci-dessous.

1. Définir l'enjeu éthique.

Formuler clairement la question éthique à l'étude. Déterminer les aspects du processus ou du résultat qui pourraient être perçus comme injustes et ceux qui nécessitent un jugement éthique.

2. Comprendre le contexte dans lequel la décision est prise.

Être conscient de la diversité inhérente au milieu social et des affaires du Canada, ainsi que des inégalités historiques qui continuent d'avoir une incidence sur les groupes marginalisés sur la base de leur race ou ethnicité, genre ou l'expression de genre, l'orientation sexuelle, la situation d'incapacité, la classe socio-économique ou autre identifiant social.

Comprendre ce que signifie l'exercice des activités à titre d'actuaire spécialiste au Canada, y compris les exigences du Code de déontologie de l'ICA et les autres attentes professionnelles réglementaires et éthiques.

3. Tenir compte des parties en cause.

Tenir compte des parties qui seront touchées par la décision et comment elles le seront. Préciser si des groupes doivent être ciblés et évalués plus en détail.

4. Rassembler l'information.

Recueillir tous les renseignements disponibles et pertinents en évaluant la fiabilité des sources. Utiliser cette information pour cerner les lacunes potentielles dans les connaissances qui rendent certaines approches éthiques impossibles.

5. Formuler des mesures et envisager des solutions de rechange.

Formuler les plans d'action possibles et en tenir compte (p. ex. coûts, avantages, temps requis, ressources nécessaires, etc.). Déterminer si d'autres mesures pourraient être prises et préciser leurs répercussions et les résultats connexes.

6. Examiner.

Tenir compte de la décision prise avant de prendre des mesures :

- Déterminer si la décision est conforme au Code de déontologie de l'ICA.
- Tenir compte de son propre point de vue sur la décision.

Tenir compte de la décision prise après l'application des mesures :

- Tenir compte des conséquences des mesures – intentionnelles et non intentionnelles.
- Prendre en considération l'évaluation du résultat compte tenu des conséquences actuelles ou déterminer si l'action aurait pu être différente.

Section 5 : Prochaines étapes pour le ou la spécialiste

Le ou la spécialiste doit envisager les étapes suivantes une fois le biais mesuré ou détecté. Les mesures prises dépendront de l'importance relative du biais dans le contexte de l'équité. Ces considérations peuvent être notées dans la documentation sur la gouvernance des modèles internes en référence aux NP applicables. Des documents précis de mesure et d'évaluation des biais peuvent être préparés pour

faire le suivi des biais identifiés, des mesures correctives adoptées ou des considérations ou recommandations futures.

Il se peut qu'aucune solution ne soit disponible dans l'immédiat pour corriger un biais détecté et réputé sensiblement injuste. Dans ce scénario, le ou la spécialiste est invité(e) à élaborer un plan de redressement ou d'atténuation. Ce plan pourrait prévoir la collecte d'éléments de données supplémentaires ou de données provenant de nouvelles sources. Par ailleurs, il pourrait nécessiter des changements plus importants dont la mise en œuvre pourrait être longue.

À titre de point de départ, les NP particulièrement pertinentes pour les prochaines étapes sont les sous-sections 1440 Données, 1450 Modèles, 1460 Assurance de la qualité et 1490 Documentation. Cette liste n'est pas exhaustive et d'autres sections des NP sont également susceptibles d'être pertinentes.

La prochaine section décrit les mesures que les spécialistes pourraient prendre pour mesurer, documenter et corriger les biais et l'injustice qui en découle dans le cadre d'un processus de tarification.

5.1 Documentation relative à l'évaluation des biais et de l'équité

5.1.1 Cadre

Le ou la spécialiste peut établir et documenter un cadre d'évaluation des biais afin d'obtenir des conseils objectifs sur la façon dont les biais peuvent être identifiés, mesurés et traités. Il est recommandé que le cadre soit cohérent entre les divers modèles de tarification pourvu qu'il demeure applicable. Cela aidera à promouvoir l'objectivité et la comparabilité dans le processus d'évaluation des biais. Ce cadre pourrait faire partie d'un modèle référentiel d'assurance conforme aux sous-sections 1490 Documentation et 1460 Assurance de la qualité des NP. Il est recommandé de documenter le cadre et de le valider au moyen de processus d'examen par les pairs et d'approbation interne avant de l'utiliser pour évaluer le biais d'un modèle.

5.1.2 Considérations relatives à la documentation

Le cadre d'évaluation des biais pourrait comprendre les sections suivantes, entre autres. Les spécialistes pourraient également élaborer leurs cadres personnalisés pour traiter des sujets de préoccupation propres à leur domaine de tarification.

1. Identification des groupes vulnérables

Le ou la spécialiste pourrait déterminer les groupes vulnérables ou protégés susceptibles d'être touchés par un biais potentiel menant à une injustice dans son processus de tarification. Les groupes vulnérables ou protégés peuvent souvent être définis en fonction des lois canadiennes. Si la définition du groupe vulnérable n'est pas conventionnelle (par exemple, d'autres vulnérabilités applicables à l'entreprise, mais non précisées dans les lois canadiennes), il est recommandé de documenter la justification de la désignation de ce groupe comme vulnérable. Il convient de noter qu'a priori les groupes vulnérables pour lesquels un résultat de tarification est biaisé peuvent ne pas être évidents ou apparents. Par conséquent, le processus de mesure du biais pourrait être plus large que ces groupes identifiés.

Nous conseillons au ou à la spécialiste de consulter un groupe diversifié d'intervenants et d'intervenantes pour que l'approche soit suffisamment multidimensionnelle et englobe divers points de vue.

2. Identification des processus

Le ou la spécialiste pourrait tenir compte des aspects du processus de tarification ou de modélisation qui devraient faire l'objet d'un examen des biais en se reportant à la définition d'un biais et aux sources potentielles de biais décrites ci-dessus. La prise en compte d'un biais causant une injustice importante

dans le résultat d'un modèle pourrait être utilisée pour concentrer cette analyse sur les domaines jugés les plus fondamentaux des biais menant à ces résultats injustes.

L'utilisation d'une méthode et d'un cadre de gouvernance transparents pour les algorithmes de tarification fondés sur l'apprentissage automatique ou sur l'IA favorisera la confiance, en veillant à ce que la détection des biais soit effectuée aux premières étapes pour les atténuer ou les éviter. Cette démarche aidera à réduire ou à prévenir les résultats négatifs potentiels et leur impact.

3. Moment propice à l'évaluation

Le ou la spécialiste peut déterminer à quel moment le modèle ou le processus doit être testé pour détecter les biais, compte tenu des données, des hypothèses et de la méthodologie retenues. Il est recommandé d'évaluer le biais lorsqu'un nouveau modèle est élaboré ou amélioré et d'effectuer des examens périodiques pour s'assurer qu'aucun biais n'est introduit par des changements dans les intrants ou l'étalonnage du modèle existant.

4. Mesure

Le ou la spécialiste devrait tenir compte de la méthode la plus favorable pour mesurer la présence de biais dans le modèle ou le processus de tarification. De nombreuses techniques de mesure sont proposées dans la section 4.2.

La justification du choix de la méthode de mesure devrait être documentée.

5. Seuil

Pour maintenir l'exactitude des taux, il pourrait ne pas toujours être prudent ou acceptable d'éliminer tous les biais présents dans le modèle. Le ou la spécialiste devrait tenir compte du seuil pour lequel le biais est considéré comme sensiblement injuste et ainsi, des mesures correctives pourraient être prises.

Nous recommandons de consulter le cadre éthique dans les sections qui précèdent pour déterminer le seuil pour lequel un traitement ou la tarification différentiel pourrait être considéré comme injuste. Comme il est indiqué, nous recommandons de maintenir le cadre éthique utilisé pour déterminer un seuil cohérent entre les modèles et de documenter clairement la justification pour son choix.

Enfin, le seuil devrait faire l'objet d'examens périodiques pour s'assurer qu'il demeure acceptable dans le contexte de la réglementation, des lois, des normes de l'ICA et des opinions sociétales sur l'équité.

6. Validation

Nous recommandons au ou à la spécialiste de soumettre ses constatations à un pair examinateur. Un pair examinateur peut évaluer le processus, la méthodologie et la conclusion ainsi que les jugements éthiques en cause.

Nous recommandons de demander un examen ou une approbation supplémentaire d'un ou d'une cadre supérieur(e) (CGR).

7. Interventions

Nous recommandons de déterminer les mesures correctives qui sont possibles dans le cadre d'évaluation des biais et de préciser le type et le niveau (élevé, moyen ou faible) de mesures correctives des biais qui sont nécessaires. Par exemple, si un biais est décelé, mais qu'il est inférieur à un seuil interne, il pourrait être considéré comme peu important ou négligeable et sans aucune mesure corrective nécessaire. Toutefois, si le biais dépasse le seuil interne et est injuste, le ou la spécialiste doit déterminer les mesures qu'il ou qu'elle peut prendre pour le réduire ou l'éliminer. Nous recommandons de déterminer dans quelle mesure le biais est important et de préciser le domaine dans lequel il se trouve pour déclencher diverses mesures correctives. Au titre des mesures potentielles, mentionnons :

- corriger les biais au sein de sa propre fonction opérationnelle, s'ils sont mineurs;
- signaler les biais à un membre de rang supérieur de la haute direction;
- aviser l'organisme de réglementation;
- aviser le groupe de clients touchés.

Pour améliorer le processus d'intervention, nous recommandons de documenter les mesures proposées et celles mises en œuvre lorsque l'intervention a eu lieu.

8. Surveillance

Il est recommandé de surveiller les mesures correctives pour s'assurer qu'elles ont l'effet souhaité. Un examen périodique et des rapports réguliers sur les biais permettront de veiller à ce que des mesures correctives puissent être prises pour maintenir les biais sous des seuils acceptables.

9. Tenue de documents

Le processus d'évaluation du biais doit être documenté pour veiller à ce qu'un dossier soit disponible si une partie intéressée le demande, comme un membre de la haute direction interne, les organismes de réglementation ou d'autres spécialistes de l'actuariat. La documentation peut comprendre une référence à l'évaluation périodique continue du processus ou du modèle même si un biais n'a pas été détecté.

Veillez consulter le [tableur](#) ci-joint pour obtenir un exemple de liste de vérification périodique de l'évaluation des biais qui peut être incluse dans la documentation de modélisation comme preuve d'évaluation des biais.

5.2 Rapports dans le cadre de la gouvernance du modèle

L'évaluation des biais pourrait faire partie de la gouvernance élargie des modèles internes. Bien que les documents portant sur l'évaluation des biais puissent être conservés comme dossiers à l'appui, il est recommandé au ou à la spécialiste d'inclure des renseignements généraux dans les sections suivantes d'un rapport de gouvernance des modèles.

- **Biais de modèle** : Si un biais (tel que défini dans le présent document) a été identifié dans les données, les hypothèses ou les résultats, le ou la spécialiste devrait l'indiquer dans la section sur le biais de la documentation sur la gouvernance des modèles et indiquer l'importance du biais.
- **Faiblesses et limites** : En cas de biais, les distorsions des résultats du modèle peuvent être incluses dans la section des faiblesses et des limites de la documentation sur la gouvernance des modèles. Si le biais est jugé important, le ou la spécialiste peut inclure le plan d'élaboration pour réduire ou éliminer le biais au moyen du calendrier de correction adéquat.
- **Vérifications de la pertinence** : L'examen des biais connus et de nouveaux biais potentiels peut faire partie des vérifications de la pertinence d'un modèle. Le ou la spécialiste peut inclure des documents d'évaluation du biais comme preuve que la vérification a été effectuée.
- **Validation du modèle** : Les biais devraient être vérifiés dans le cadre d'un examen global par les pairs et d'une validation indépendante du modèle. Le ou la spécialiste peut inclure la documentation d'évaluation du biais comme preuve d'un examen par les pairs ou d'un examen de validation indépendant. Si le ou la spécialiste est un validateur indépendant, nous recommandons de demander le résultat de l'évaluation du biais tel que défini dans le présent document.
- **Changements apportés au modèle** : Si des changements ou des améliorations sont apportés à un modèle, le ou la spécialiste devrait effectuer une évaluation des biais pour confirmer que les changements n'ont pas introduit ou exacerbé de biais dans le modèle.

Section 6 : Autres recherches

Les professionnels de l'actuariat sont régulièrement confrontés à des biais sociaux et économiques, et l'évaluation de l'équité de ces biais devrait figurer au premier plan des travaux de l'actuaire. Le présent document donne un aperçu des concepts de biais, d'équité et d'éthique et il sert de point de départ pour cerner et atténuer les biais sociaux dans la tarification des assurances IARD. Bien que le document soit intentionnellement de grande portée, nous reconnaissons que l'application de considérations relatives aux biais et à l'équité serait primordiale dans le domaine de l'apprentissage automatique. Nous recommandons d'autres recherches axées sur l'apprentissage automatique et la modélisation prédictive, notamment :

- Développer les sources de biais, particulièrement dans les résultats d'apprentissage automatique.
- Développer des outils pour déterminer l'équité dans les modèles prédictifs complexes.
- Identifier des approches pour utiliser l'audit de l'apprentissage automatique afin de cerner les biais sociaux et l'injustice.
- Explorer l'application de facteurs de pénalité à l'injustice dans les techniques d'apprentissage automatique.
- Examiner des approches éthiques en matière d'équité et des techniques d'atténuation des biais.

D'autres recherches peuvent être appliquées à d'autres domaines de la tarification en actuariat, comme l'assurance vie, les régimes de retraite et la santé.

Section 7 : Exemple illustratif

Le ou la spécialiste doit vérifier les algorithmes de tarification des primes de son portefeuille de biens immobiliers pour y déceler un éventuel biais de tarification à l'égard de tout groupe vulnérable. La vérification vise à déterminer si un biais dans l'algorithme de tarification peut être perçu comme injuste et si des mesures correctives doivent être prises.

Identification des groupes vulnérables

Le ou la spécialiste tient compte des catégories suivantes de groupes vulnérables qui peuvent être présents au sein de la population assurée :

- Personnes âgées
- Personnes handicapées
- Populations historiquement marginalisées sur le plan racial et ethnique, y compris les peuples autochtones
- Nouveaux arrivants au Canada
- Personnes à faible revenu

Identification des biais

Le ou la spécialiste tient compte de la structure de tarification des biens immobiliers et des facteurs de tarification qui pourraient varier selon le risque de police dans le portefeuille de polices d'assurance de biens. Même si les facteurs de tarification ont été majoritairement associés à la présence de dispositifs d'atténuation des risques, de descriptions de biens ou de choix des titulaires, le ou la spécialiste relève les facteurs de tarification suivants qui pourraient entraîner un biais :

- « Âge des assurés » – incidence sur les titulaires en fonction de leur âge.

- « Type de bien immobilier (maisons, condominiums, propriétés de vacances) » – peut avoir une incidence sur les nouveaux arrivants et les nouvelles arrivantes et les personnes à faible revenu qui pourraient être plus susceptibles de vivre en copropriété.
- « Code postal » – peut avoir une incidence sur les minorités raciales et ethniques qui pourraient être davantage susceptibles de vivre dans certains quartiers.
- « Nombre de personnes occupant le bien immobilier » – peut avoir une incidence sur les nouveaux arrivants et les nouvelles arrivantes et les personnes à faible revenu qui pourraient être plus susceptibles de vivre avec des parents ou des amis.
- « Montant d'assurance (immeuble et contenu) » – peut avoir une incidence sur les personnes à faible revenu qui sont davantage susceptibles d'avoir un faible montant d'assurance; peut également avoir une incidence sur les personnes en situation de handicap, car la faible limite d'accès à la propriété et la limite élevée de contenu pourrait indiquer l'assurance d'améliorations apportées pour une personne atteinte d'invalidité ou de problèmes de mobilité.

Le ou la spécialiste examine les facteurs de tarification identifiés pour déterminer la probabilité de biais.

Mesure du biais

Le ou la spécialiste applique la mesure du biais désignée « différences de moyennes ». Si une société d'assurances impose aux titulaires des taux de prime proportionnels au risque qu'ils représentent, les ratios moyens de sinistres devraient être constants d'une cohorte à l'autre. Le ou la spécialiste calcule le ratio de sinistres moyen pour chaque groupe de risques et pour chaque variable prise en compte, et il ou elle note que les ratios de sinistres demeuraient modérément uniformes (entre 0 % et 5 %) pour toutes les variables de risque, sauf le « code postal ».

Pour le « code postal », le ratio de sinistres moyen a produit des ratios de sinistres instables variant de 0 % à 15 % selon la région.

Seuil d'importance relative

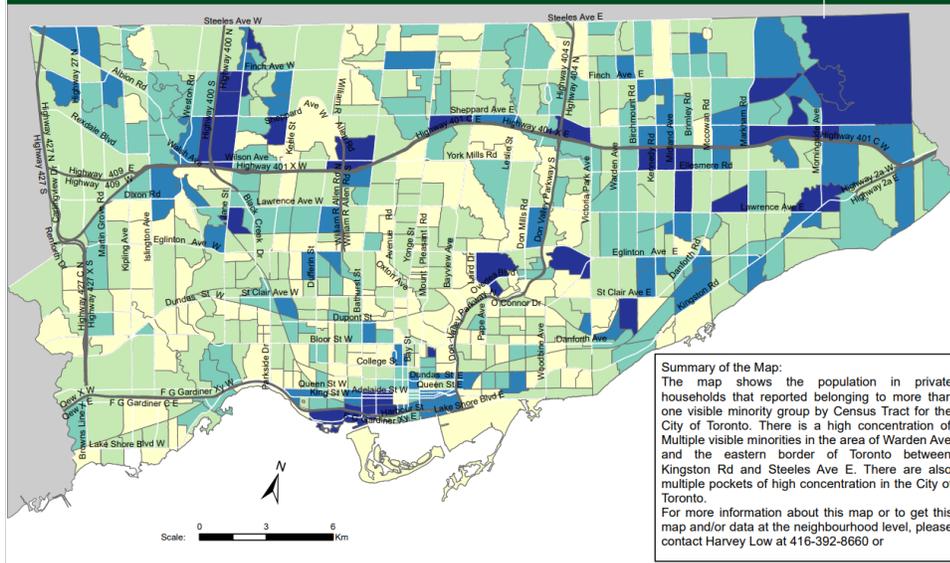
Le ou la spécialiste consulte les lignes directrices internes sur l'importance relative et il ou elle détermine qu'une erreur équivalant à 7 % de la prime souscrite brute est réputée d'importance relative. Il ou elle décide d'utiliser le même seuil d'importance relative pour déterminer si les différences de rapport de sinistres sont « d'importance relative » et méritent d'être examinées plus en détail.

Lors de l'examen des variables liées au code postal, le ou la spécialiste remarque que, pour certains codes postaux de la région de Toronto et du Grand Toronto (RGT), la prime facturée aux titulaires est sensiblement plus élevée que celle suggérée en fonction de l'expérience de risque dans ces régions.

Le ou la spécialiste conclut qu'il existe un biais important dans la variable code postal.

Existe-t-il un problème d'équité?

Bien que l'on ait supposé que le code postal est utilisé pour représenter les quartiers des minorités raciales et ethniques, le ou la spécialiste veut déterminer si les groupes vulnérables sont effectivement touchés. À cette fin, il ou elle examine d'autres sources de données pour déterminer la répartition des minorités dans Toronto et la RGT et il ou elle a trouvé la carte ci-dessous.



Source: Ville de Toronto, [Recensement 2016 de Statistique Canada](#).

La superposition des profils de codes postaux sur la carte ci-dessus par le ou la spécialiste permet de constater que les membres des minorités visibles paient des primes anormalement élevées par rapport aux sinistres réglés. Le ou la spécialiste conclut que ce traitement des minorités vulnérables pourrait être jugé injuste par le public, les organismes de réglementation ou l'ICA.

Mesures correctives

Le ou la spécialiste examine la façon dont les facteurs liés au code postal sont générés et il ou elle établit que ces facteurs n'ont pas été mis à jour depuis 2010. Les profils de sinistralité et d'exposition ont radicalement changé au cours des 12 dernières années dans certaines régions.

Le ou la spécialiste soulève cette question auprès des hautes instances du Ministère, il ou elle prépare un article et il ou elle demande un examen par les pairs pour valider les constatations.

Le ou la spécialiste élabore et propose les correctifs pertinents qui suivent :

Mesures à prendre	Avantages	Inconvénients
Mettre à jour tous les facteurs liés au code postal pour qu'ils reposent sur les données les plus récentes sur la sinistralité et l'exposition.	<ul style="list-style-type: none"> Des primes plus précises et plus équitables seront facturées à tous les titulaires. Dans certains cas, les primes facturées aux titulaires étaient trop faibles. La correction permet à l'assureur de percevoir des revenus de primes qui correspondent davantage à son exposition au risque. 	<ul style="list-style-type: none"> La collecte et l'évaluation des données sont coûteuses et longues. D'après l'analyse du ratio sinistres-primes moyen, il semble qu'en moyenne, pour la plupart des titulaires, les facteurs liés au code postal demeurent pertinents.
Mettre à jour uniquement les facteurs liés au code postal	<ul style="list-style-type: none"> Des primes plus précises et plus équitables seront 	<ul style="list-style-type: none"> Les facteurs liés au code postal demeureront désuets

associés aux groupes vulnérables.	facturées aux groupes vulnérables. <ul style="list-style-type: none"> • La collecte et l'analyse de données plus ciblées sont moins coûteuses et moins longues. 	pour la majorité des titulaires, ce qui donne lieu à une perte marginale de rentabilité pour l'assureur.
Ne rien faire	<ul style="list-style-type: none"> • Puisque les primes facturées à 95 % des titulaires varieront peu ou pas du tout, la prime acquittée par la plupart des clients et des clientes est juste. • Aucun coût n'est engagé pour corriger les facteurs liés au code postal. 	<ul style="list-style-type: none"> • La mauvaise tarification pour les groupes vulnérables, si elle est connue publiquement, est susceptible de porter atteinte à la réputation de l'assureur. • Le fait de ne pas mettre à jour les facteurs liés au code postal pourrait entraîner une disparité croissante et coûteuse des primes au fil du temps.

L'employeur du ou de la spécialiste a pour mandat opérationnel de se concentrer sur le client ou la cliente et de traiter tous les clients et les clientes de façon équitable et intègre. Le service de tarification du ou de la spécialiste juge qu'il est éthique de s'assurer qu'une prime équitable est facturée à tous les clients.

Après des discussions internes au sein du service de tarification du ou de la spécialiste, il est conclu que les facteurs liés au code postal associés aux groupes vulnérables seront corrigés à la main, et ce, sans tarder. La mise à jour complète de tous les facteurs liés au code postal devrait être effectuée et appliquée en temps opportun.

Surveillance

Une fois les changements mis en œuvre et les facteurs liés au code postal mis à jour, les modifications apportées au code postal et à d'autres facteurs de tarification font l'objet d'un suivi périodique pour s'assurer que les facteurs utilisés sont fondés sur les résultats de sinistralité et d'exposition les plus récents et les données les plus crédibles.

Tenue de documents

L'analyse effectuée par le ou la spécialiste et les mesures correctives mises en œuvre sont documentées et conservées par le service de tarification du spécialiste.

Annexe 1 : Diverses définitions sur lesquelles nous nous appuyons

Mesurer le biais

Certains facteurs de tarification, comme l'âge, le sexe et l'état matrimonial, sont interdits dans certaines provinces du Canada et ne peuvent pas être utilisés directement pour la tarification. De même, d'autres facteurs socioéconomiques comme l'ethnicité, la race et la religion sont interdits partout, non pas seulement dans certaines provinces.

Si nous prenons le ratio sinistres-primés d'un groupe de titulaires et le ratio sinistres-primés divisé par un facteur de tarification « indésirable », le ratio sinistres-primés devrait être très semblable en l'absence de biais dans les données. Par ailleurs, il pourrait indiquer un biais introduit par un substitut involontaire.

L'importance statistique des écarts de sinistres-primés pourrait être testée au moyen de la valeur p .

Éthique de la science des données

L'Institute and Faculty of Actuaries (R.-U.) et la Royal Statistical Society ont produit un document intitulé *A Guide for Ethical Data Science*¹⁴. Les cinq thèmes du document sont les suivants :

1. Chercher à accroître la valeur de la science des données pour la société.
2. Éviter les préjugés.
3. Appliquer et maintenir les compétences professionnelles.
4. Chercher à maintenir ou à accroître la fiabilité.
5. Maintenir la responsabilisation et la supervision.

Les concepts contenus dans le présent document peuvent également être appliqués aux considérations éthiques que les actuaires devraient connaître lorsqu'ils effectuent des travaux de tarification.

¹⁴ « [A Guide for Ethical Data Science](#) » Royal Statistical Society et Institute and Faculty of Actuaries, 2019.

Annexe 2 : Exemple de cadres éthiques

Les principales théories éthiques normatives sont décrites ci-dessous.

Utilitarisme

L'approche utilitaire porte sur les conséquences éthiques d'une action particulière. Le principe de base de l'utilitarisme est que « le choix le plus éthique est celui qui produira le plus grand bien pour le plus grand nombre »¹⁵.

L'approche utilitaire est la plus courante pour prendre des décisions éthiques ayant des conséquences qui visent un grand nombre de personnes. Le ou la spécialiste doit soupeser toutes les actions potentielles et mettre de l'avant celles qui procurent le plus d'avantages et qui causent le moins de préjudices à la majorité.

Déontologie

L'approche déontologique met l'accent sur les principes qui guident la personne qui prend la décision ou qui applique une mesure particulière. Selon cette approche, [traduction] « les gens doivent suivre les règles et faire leur devoir »¹⁶.

Selon cette approche, les décisions prises par le ou la spécialiste doivent être conformes aux obligations contractuelles de son employeur, au Code de déontologie de l'ICA, aux exigences réglementaires et aux lois.

Vertu

L'approche fondée sur la vertu met l'accent sur la position morale de la personne plutôt que sur la moralité d'une seule action. L'approche fondée sur la vertu suppose qu'[traduction]« en s'exerçant à être honnête, courageuse, juste, généreuse, etc., une personne développe un caractère honorable et moral »¹⁷.

Cette approche suppose que les mesures prises sont éthiques si elles concernent une personne ayant un niveau de moralité élevé. Les spécialistes ayant une bonne réputation professionnelle et qui pratiquent pour un employeur appliquant des valeurs et des principes axés sur le client doivent prendre une décision moralement correcte en cas d'incertitude éthique.

Trois cadres éthiques

Les théories normatives discutées ci-dessus peuvent servir à établir un cadre distinct pour la prise de décisions éthiques. Chacun des trois cadres permet de prendre des décisions, mais il comporte des limites connexes que le ou la spécialiste doit connaître. Ces cadres sont résumés dans le tableau qui suit :

¹⁵ « [Utilitarianism](#) », *Ethics Unwrapped*, McCombs School of Business, University of Texas at Austin.

¹⁶ « [Deontology](#) », *Ethics Unwrapped*.

¹⁷ « [Virtue Ethics](#) », *Ethics Unwrapped*.

	Cadre utilitaire	Cadre déontologique	Cadre fondé sur la vertu
Point de mire	Effets futurs des mesures. Qui serait touché directement et indirectement.	Obligations actuelles et futures.	Traits de caractère qui motivent le processus décisionnel.
Mesures à prendre	Recherche des meilleures conséquences	Chercher à se conformer aux exigences	Chercher à développer son propre caractère
Considérations	Buts et objectifs Analyse coûts-avantages	Obligations contractuelles Lois et règlements Code de déontologie de l'ICA	Valeurs personnelles Réputation personnelle et professionnelle Le caractère de l'entreprise
Limites	Pourraient ne pas convenir lorsque la décision concerne des minorités. Souvent, ne respecte pas les décisions morales ordinaires ou ne se conforme pas à l'évaluation morale. Manque de ressources, de temps et d'information pour évaluer tous les plans d'action.	Aucune attention n'a été portée au résultat de la mesure. Les mesures multiples ne sont pas classées par ordre de mérite.	Suppose que le décideur a déjà un caractère solide – ce n'est peut-être pas le cas en pratique. La décision est subjective; elle peut varier selon la vertu des décideurs.

Source: Adapté de Sally Haslanger, « [Three Moral Theories](#) », (notes de cours), MIT Open Courseware, 2017, et Paul Lewis et Johann le Roux, « [The Ethics of Claims Assessment Practices in the South African Life Insurance Industry](#) » (présentation PowerPoint, 2019 Convention of the Actuarial Society of South Africa, 22-23 octobre 2019).

Pour revenir à l'exemple de la section 4.3, le ou la spécialiste peut appliquer les cadres éthiques susmentionnés dans son processus décisionnel comme suit :

1. Le ou la spécialiste pourrait choisir de ne rien faire :

- **Justification fondée sur l'approche utilitaire** : Pour la majorité des assurés, les prix étaient exacts, et des mesures correctives amélioreraient les taux pour la minorité touchée, mais se traduiraient par des taux plus élevés pour la plupart des autres groupes. Il y a également le risque que les primes soient trop élevées pour de meilleurs risques (peu importe la race). Ainsi, les répercussions négatives potentielles sur l'entreprise ne valent pas la peine de corriger la situation.
- **Justification fondée sur l'approche déontologique** : Le règlement stipule que les assureurs ne sont pas autorisés à utiliser la race comme facteur de tarification. Dans ce cas, le modèle de

tarification n'utilise pas ce facteur – donc, l'assureur se conforme à la primauté du droit et peut choisir de ne rien faire.

2. Le ou la spécialiste peut choisir d'introduire la mesure corrective immédiatement :

- **Justification fondée sur l'approche utilitaire** : Une enquête plus approfondie révèle que l'exclusion du facteur code postal et la tarification fondée plutôt sur la région réduirait l'exactitude, mais se traduirait par de meilleurs tarifs pour la catégorie privilégiée de titulaires (qui ont de meilleurs dossiers de conduite, peu importe la race). La correction se traduirait par un investissement de coût et de temps, mais elle améliorerait la rentabilité à long terme; elle est donc tout à fait justifiable.
- **Justification fondée sur l'approche déontologique** : Le règlement stipule clairement que la « race » ne peut être utilisée comme facteur de tarification, même par substitution. L'assureur doit apporter une correction pour se conformer aux exigences des lois et règlements.
- **Justification selon l'approche fondée sur la vertu** : Bien que l'assureur se conforme à la loi et que la correction des modèles de tarification soit coûteuse et réduise la précision, il est moralement faux de différencier la tarification en fonction de la race et l'assureur doit cesser de perpétuer les inégalités raciales au sein de la société.

Annexe 3 : Définitions générales du biais

Les définitions générales de base et détaillées du biais sont incluses dans la présente annexe à titre de référence. Les définitions 1 à 3 sont des définitions de base, tandis que les définitions 4 et 5 sont plus détaillées. La définition 4 envisage le biais dans le contexte de la science des données.

1. [traduction] Tendence (connue ou inconnue) à privilégier une chose par rapport à une autre, ce qui empêche l'objectivité et influence la compréhension ou les résultats d'une certaine façon¹⁸.
2. [traduction] L'action qui consiste à soutenir ou à s'opposer à une personne en particulier ou à quelque chose d'une manière injuste parce que l'on permet que des opinions personnelles influencent le jugement¹⁹.
3. [traduction] Tendence ou préjudice pour ou contre une personne ou un groupe, surtout d'une manière jugée injuste.
4. [traduction] Le biais comporte plusieurs définitions et son usage courant est résolument négatif. Nous utilisons généralement ce terme pour signifier le favoritisme systématique d'un groupe. De façon générale, le terme « biais » provient du grec ancien qui décrit une ligne oblique (c.-à-d. une déviation de l'horizontale). Dans la science des données, le biais est un écart par rapport aux attentes dans les données. De façon plus fondamentale, le biais désigne une erreur dans les données. Mais l'erreur est souvent subtile ou elle passe inaperçue²⁰.
5. « biais » Terme que nous définissons de façon générale par rapport aux résultats qui sont systématiquement moins favorables aux personnes d'un groupe particulier et lorsqu'il n'y a pas de différence pertinente entre les groupes qui justifie de tels préjudices²¹.

¹⁸ *Open Education Sociological Dictionary*, "[bias \(n.\)](#)."

¹⁹ *Cambridge Dictionary*, "[bias \(n.\)](#)."

²⁰ Will Goodrum, « [Statistical and Cognitive Biases in Data Science: What Is Bias?](#) » *Elder Research* (blogue), 6 octobre 2017.

²¹ Nicol Turner Lee, Paul Resnick et Genie Barton, « [Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms](#) », *Brookings*, 22 mai 2019.

Références citées

- Cardoso, Tom. « [Bias Behind Bars: A Globe Investigation Finds a Prison System Stacked Against Black and Indigenous Inmates](#) », *The Globe and Mail*, octobre 2020.
- « [Deontology](#) ». *Ethics Unwrapped*. McCombs School of Business, University of Texas at Austin.
- Goodrum, Will. « [Statistical and Cognitive Biases in Data Science: What Is Bias?](#) » *Elder Research* (blogue). 6 octobre 2017.
- Haslanger, Sally. « [Three Moral Theories](#) ». Notes de cours. MIT Open Courseware. 2017.
- Ito, J. « [Supposedly 'Fair' Algorithms Can Perpetuate Discrimination](#) ». *Wired*. 5 février 2019.
- Lange, A.R. et Natasha Duarte. « [Understanding Bias in Algorithmic Design](#) ». *ASME Demand*. Printemps 2017.
- Lewis, Paul et Johann le Roux. « [The Ethics of Claims Assessment Practices in the South African Life Insurance Industry](#) ». Présentation PowerPoint. 2019 Convention of the Actuarial Society of South Africa. 22-23 octobre 2019.
- Lindholm, M., R. Richman, A. Tsanakas et M.V. Wüthrich, « [Discrimination-Free Insurance Pricing](#) ». *ASTIN Bulletin: The Journal of the IAA* 52, v. 1 (2021).
- [Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois](#), 44^e législature, 1^{re} session, 2022.
- Modlin, Claudine, Willis Towers Watson et Geoff Werner. *Basic Ratemaking*, 5^e édition. Casualty Actuarial Society, mai 2016.
- Mosley, Roosevelt et Radost Wenman. « [Methods for Quantifying Discriminatory Effects on Protected Classes in Insurance](#) ». Casualty Actuarial Society Research Paper Series on Race and Insurance Pricing. Mai 2022.
- Pinnacle Actuarial Services. [Rapport d'évaluation de la tarification territoriale de Pinnacle](#), rapport spécial pour l'Autorité ontarienne de réglementation des services financiers, octobre 2022, 10 (en anglais seulement).
- Royal Statistical Society et Institute and Faculty of Actuaries, [A Guide for Ethical Data Science](#), 2019.
- Sénat du Canada, [Étude concernant les droits de la personne des prisonniers dans le système correctionnel fédéral](#), Rapport provisoire du Comité sénatorial permanent des droits de la personne, février 2019.
- Starks, Mary et coll., « [Price Discrimination in Financial Services](#) », Financial Conduct Authority, juillet 2018.
- Turner Lee, Nicol, Paul Resnick et Genie Barton. « [Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms](#) ». *Brookings*. May 22, 2019.
- “[Utilitarianism](#).” *Ethics Unwrapped*. McCombs School of Business, University of Texas at Austin.
- “[Virtue Ethics](#).” *Ethics Unwrapped*. McCombs School of Business, University of Texas at Austin.
- [Zurich Insurance Co. c. Ontario \(Commission des droits de la personne\)](#). [1992] 2 RCS 321.



© 2022 Institut canadien des actuaires

Institut canadien des actuaires

360, rue Albert, bureau 1740

Ottawa (Ontario) K1R 7X7

613-236-8196

head.office@cia-ica.ca

cia-ica.ca

<https://www.voiraudeladurisque.ca/>



L'Institut canadien des actuaires (ICA) est l'organisme qui représente la profession actuarielle au Canada. Nous élaborons et respectons des normes rigoureuses, nous partageons notre expertise en gestion des risques et nous faisons progresser la science actuarielle pour améliorer la vie des gens au Canada et à l'échelle du monde. Nos quelque 6 000 membres mettent à profit leurs connaissances en mathématiques, en statistiques, en analyse de données et en affaires pour offrir des services et des conseils de la plus haute qualité afin d'aider les Canadiens et les organisations canadiennes à envisager l'avenir avec confiance.