

Recherche quantitative originale

Validation de Canproj pour les projections de données sur l'incidence du cancer au Canada

Alain Demers, Ph. D. (1); Zhenguo Qiu, Ph. D. (2); Ron Dewar, M. Sc. (3); Amanda Shaw, M. Sc. (1)

Cet article a fait l'objet d'une évaluation par les pairs.

 Diffuser cet article sur Twitter

Résumé

Introduction. Les projections relatives au cancer fournissent une information essentielle pouvant aider à établir l'ordre de priorité des stratégies de lutte contre le cancer, à répartir les ressources et à évaluer les interventions et les traitements actuels. Canproj est un outil de projection pour le cancer utilisant le progiciel R qui est inspiré de Nordpred et qui ajoute une sélection de modèles de projection. L'objectif de ce projet consistait à valider Canproj, programmé avec R, pour les projections à court terme relatives aux taux de cancer.

Méthodologie. Nous avons utilisé les données nationales sur l'incidence du cancer tirées du Système national de déclaration des cas de cancer et du Registre canadien du cancer pour 1986 à 2014. Nous avons utilisé la validation croisée pour estimer l'exactitude des projections générées par Canproj, avec le biais relatif (BR) comme mesure de validation. Nous avons également évalué l'arbre décisionnel de Canproj pour la sélection automatique du modèle.

Résultats. Cinq des six modèles ont présenté un BR moyen se situant entre 5 % et 10 % et un BR médian d'environ 5 %. Dans le cas de certains sièges de cancer pour lesquels il était plus difficile d'établir des projections, l'utilisation d'une période plus courte a permis d'améliorer la fiabilité. Le modèle Nordpred a été sélectionné dans 79 % des cas par la fonction de sélection automatique de modèle de Canproj, même s'il offrait le BR le plus faible seulement 24 % du temps.

Conclusion. Le progiciel Canproj a été en mesure de fournir des projections très semblables aux données réelles pour la plupart des sièges de cancer.

Mots-clés : néoplasmes, prévisions, études de validation

Introduction

Depuis 30 ans, la Société canadienne du cancer et le gouvernement du Canada (Agence de la santé publique et Statistique Canada) publient un rapport annuel exhaustif, intitulé *Statistiques canadiennes sur le cancer* (SSC). Ce rapport présente une série de données sur le nombre de nouveaux cas de cancer et sur la mortalité par cancer au sein de la population et les taux projetés qui comblent l'écart entre la dernière année pour laquelle des données

étaient disponibles et l'année de publication du rapport. Ces projections constituent une ressource qui permet d'assurer la planification et d'établir les priorités pour les intervenants et elles permettent également de tenir la population canadienne informée du fardeau considérable que constitue le cancer.

Divers modèles de projection ont été utilisés au fil des ans pour produire la publication SCC. La régression de Poisson¹ utilisée de 2003 à 2012 a été remplacée par

Points saillants

- L'éventail de modèles offerts par Canproj permet d'effectuer des prévisions fiables pour la plupart des sièges de cancer.
- Lorsque des variations ont été observées en ce qui concerne les taux d'incidence, on a pu utiliser une période plus courte comme base de projection afin d'améliorer l'exactitude des taux d'incidence projetés.
- En ce qui concerne l'ensemble de données nationales, le modèle Nordpred était celui qui était le plus souvent sélectionné par l'arbre décisionnel Canproj.
- Le modèle Nordpred a été sélectionné par l'arbre décisionnel Canproj dans 79 % des cas, même s'il présentait le biais relatif (BR) le plus faible seulement 24 % du temps.

Nordpred en 2011-2012. Nordpred, un progiciel R développé en Norvège, offre un seul modèle de projection, le modèle âge-période-cohorte (APC) avec un terme de dérive². Ce progiciel a fait l'objet d'études approfondies et permet d'améliorer la fiabilité des projections relatives au cancer³⁻⁷.

Dans le but d'améliorer l'établissement de projections relatives au cancer, Qiu et ses collaborateurs ont mis au point Canproj, lui aussi un progiciel R⁸. Canproj comporte trois avantages importants par rapport à Nordpred : 1) remplacement de la distribution de Poisson par la distribution

Rattachement des auteurs :

1. Agence de la santé publique du Canada, Ottawa (Ontario), Canada
2. Clinical Workforce Planning, Health Professions Strategy & Practice (HPSP), Alberta Health Services, Edmonton (Alberta), Canada
3. Registries and Analytics, Nova Scotia Health Authority Cancer Care Program, Halifax (Nouvelle-Écosse), Canada

Correspondance : Alain Demers, Agence de la santé publique du Canada, 785, avenue Carling (603A1), Ottawa (Ontario) K1A 0K9; tél. : 613-797-2129; courriel : alain.demers@canada.ca

binomiale négative en présence d'une surdispersion; 2) inclusion d'un modèle âge-cohorte et 3) une série de modèles hybrides qui combinent les forces du modèle de Poisson ou de la régression binomiale négative, la méthode de régression par segmentation⁹ et une méthode de moyenne pour les projections en fonction du nombre de cas selon l'âge. Certaines des fonctions de Canproj ont été utilisées pour les SCC de 2017¹⁰ et le progiciel complet a été utilisé pour les SCC de 2019.

Canproj est un outil relativement nouveau de projection sur le cancer, qui n'a été ni largement utilisé ni validé^{11,12}. L'objectif de notre projet était de valider les projections sur l'incidence du cancer au Canada à court terme (jusqu'à 5 ans) générées par le progiciel Canproj à l'aide des données canadiennes. Plus particulièrement, nous avons comparé les résultats des modèles de projection Canproj aux données réelles en utilisant la méthode de validation croisée d'échantillon-test (méthode *holdout*)¹³ et la représentation graphique. Nous avons également évalué les fonctions de sélection automatique du modèle offertes par Canproj (arbres décisionnels) pour déterminer si elles étaient efficaces pour choisir le meilleur modèle.

Méthodologie

Données

Pour notre analyse, nous avons utilisé les données sur l'incidence du cancer tirées du Système national de déclaration des cas de cancer et du Registre canadien du cancer (RCC) pour 1986 à 2014¹⁴. Les données du Québec n'ont pas été incluses, car cette province n'a pas soumis de données au RCC depuis 2010. Le fichier de données est fondé sur le document International Rules for Multiple Primary Cancers du Centre international de recherche sur le cancer¹⁵. Les résultats ont été compilés pour tous les cancers combinés ainsi que selon le siège de cancer (les mêmes sièges de cancer que ceux qui figurent dans la publication annuelle SCC) et le sexe¹⁰. Un ensemble de données a été créé pour chaque combinaison de sexe (2) et de type de cancer (19 types communs aux hommes et aux femmes et 5 types spécifiques à un sexe : col de l'utérus, ovaire, utérus, prostate et testicule). L'information contenue dans les ensembles de données est présentée par année (1986 à 2014) et par groupes d'âge (18 groupes, par tranche de 5 ans, soit 0 à 4 ans, 5 à 9 ans, ...,

85 ans et plus). Les estimations annuelles de population en fonction de la zone géographique, de l'âge et du sexe ont été fournies par Statistique Canada avec des estimations de population postcensitaires reposant sur le Recensement canadien de 2016¹⁶. Les estimations intercensitaires et postcensitaires ont été ajustées par Statistique Canada pour tenir compte du sous-dénombrement net. Les taux ont été normalisés selon l'âge à l'aide de la méthode directe et des effectifs de la population canadienne de 2011¹⁷.

Canproj

Le progiciel R de Canproj contient plusieurs modèles destinés à effectuer les projections de données sur l'incidence du cancer ou sur la mortalité par cancer : le modèle Nordpred, qui intègre les effets de l'âge, de la dérive, de la période et de la cohorte; le modèle âge-cohorte; trois modèles hybrides qui intègrent les effets de l'âge et éventuellement les effets de période (selon l'âge ou pour tous les âges) et le modèle fondé sur le calcul de la moyenne sur cinq ans (tableau 1)¹⁸.

Le progiciel Canproj utilise deux arbres décisionnels pour sélectionner le modèle le plus approprié selon l'importance des variables. On peut également sélectionner les modèles individuellement. Dans un premier temps, Canproj tient compte de quatre variables, soit l'âge, la période, la cohorte et un paramètre de dérive; il s'agit du modèle le plus complexe, et ce sont les variables que Nordpred utilise. Canproj détermine d'abord si la variable de cohorte est significative. Si elle l'est, Canproj détermine si le paramètre de dérive est

significatif. Si la variable de cohorte et le paramètre de dérive sont tous deux significatifs, Canproj sélectionne le modèle Nordpred pour faire les projections. Si la variable de cohorte est significative mais que le paramètre de dérive ne l'est pas, Canproj sélectionne le modèle âge-cohorte.

Si l'effet de cohorte n'est pas significatif, Canproj sélectionne l'un des modèles hybrides. Si le nombre de cas est trop faible pour exécuter un modèle de régression, une moyenne sur cinq ans est calculée. Si le nombre de cas est suffisamment important, Canproj utilisera deux modèles : un modèle « tendance commune pour l'âge » et un modèle « tendance spécifique par âge ». Si le modèle « tendance spécifique par âge » est mieux adapté, c'est ce modèle qui sera sélectionné. Sinon, le modèle « tendance commune pour l'âge » sera sélectionné. La pente de la variable de tendance commune est ensuite testée pour déterminer si elle diffère de zéro. Si elle ne diffère pas de zéro, seule la variable de l'âge est utilisée dans le modèle, alors que, si elle diffère de zéro, c'est le modèle « âge + tendance commune » qui est utilisé.

Validation

La validation croisée a été utilisée pour estimer l'exactitude des projections générées par Canproj au moyen d'un sous-ensemble de données (données d'entraînement) et pour valider les résultats de l'autre sous-ensemble (données d'essai indépendantes). Cette étude a fait appel à la méthode de l'échantillon-test (méthode *holdout*)¹³ pour créer les données d'entraînement et les ensembles de

TABLEAU 1
Modèles offerts dans le progiciel R de Canproj et variables incluses dans les modèles

Modèle ^a	Variables du modèle			
	Âge	Période	Cohorte	Dérive
Nordpred	✓	✓	✓	✓
Âge-cohorte	✓	–	✓	–
Hybride, tendance spécifique par âge ^b	✓	✓	–	–
Hybride, tendance commune pour l'âge ^c	✓	✓	–	–
Hybride, âge seulement (moyenne) ^d	✓	–	–	–
Moyenne sur 5 ans ^e	✓	–	–	–

^a La distribution de Poisson ou distribution binomiale négative peut être sélectionnée pour le modèle Nordpred, le modèle âge-cohorte et le modèle hybride selon l'âge.

^b La tendance relative à la période est calculée selon le groupe d'âge.

^c La tendance relative à la période est commune à tous les groupes d'âge.

^d Le taux moyen fondé sur un certain nombre d'années est déterminé par l'importance du taux normalisé selon l'âge.

^e Taux moyen fondé sur les 5 années de données les plus récentes.

données d'essai indépendantes. Les données pour 1986-2010 (soit 5 périodes de 5 ans) ont été utilisées comme données d'entraînement, et les données pour 2011-2014 (quatre dernières années de données) comme données d'essai indépendantes. Les prédictions issues du modèle des données d'entraînement et les données réelles des quatre dernières années ont été comparées pour évaluer l'exactitude des modèles de projection.

La mesure de validation que nous avons utilisée, soit le biais relatif (BR), permet de comparer la valeur attendue générée par les modèles de projection aux valeurs observées dans l'ensemble de données d'essai pour les années de diagnostic 2011 à 2014. Le BR mesure la différence relative en pourcentage entre la valeur prévue (E) (ou projetée) et la valeur observée (O).

$$BR_t = \frac{|E_t - O_t|}{O_t} \times 100, \text{ où } t = 2011 \text{ à } 2014$$

Dans le cas présent, la « valeur » étudiée est le taux normalisé selon l'âge.

Les BR ont été rassemblés en fonction du modèle de projection, du type de cancer et du sexe.

Nous avons comparé les BR moyens et médians en fonction du modèle, du type de cancer et du sexe pour la période (d'essai) de projection de 4 ans. Le BR médian indique la performance habituelle d'un modèle, tandis que le BR moyen (en raison de sa sensibilité aux valeurs extrêmes) aide à mettre en évidence les modèles dont les résultats sont généralement exacts, mais qui sont susceptibles de se révéler occasionnellement très inexacts.

Analyses Joinpoint

Nous avons utilisé le logiciel d'analyse des tendances Joinpoint, version 4.5.0.1 (National Cancer Institute, Bethesda, Maryland, États-Unis)¹⁹, pour calculer les tendances relatives à l'incidence du cancer au Canada selon le type de cancer et le sexe pour 1986-2010. Les estimations tirées de modèles Joinpoint ont été utilisées pour calculer les BR de 1986 à 2010. Cette mesure, qui permet d'obtenir une estimation de la variabilité des données d'entraînement, a été comparée au BR mesuré dans les données projetées. Le nombre maximum de points de jonction a été fixé à 4, le nombre minimum d'observations entre un point de jonction et le

début ou la fin des données a été fixé à 3 et le nombre minimum d'observations entre deux points de jonction a été fixé à 4. Sinon, ce sont les paramètres par défaut relatifs aux points de jonction qui ont été utilisés. On a calculé les taux normalisés selon l'âge avec transformation logarithmique et les erreurs-types connexes entrés dans le modèle à points de jonction à l'aide du progiciel statistique SAS version 9.3 (SAS Institute Inc., Cary, Caroline du Nord, États-Unis). Canproj a été exécuté à l'aide du logiciel R version 3.5.1 (R Foundation for Statistical Computing, Vienne, Autriche) et du logiciel RStudio version 1.1.453 (RStudio Inc., Boston, Massachusetts, États-Unis).

Indicateurs de rendement

Deux indicateurs ont été utilisés pour mettre en évidence les modèles susceptibles de générer des projections moins fiables. Le premier indicateur est l'identification d'un point de jonction au cours de la période de 10 ans la plus récente utilisée pour former les modèles de projection (2001 à 2010). De récents changements dans les tendances sont susceptibles d'indiquer qu'il pourrait être plus difficile pour les modèles d'effectuer des projections fiables. Nous avons divisé les points de jonction en deux groupes : ceux situés entre 10 et 6 ans avant la dernière année où des données d'entraînement étaient disponibles et ceux situés entre 5 et 3 ans avant la dernière année de données. Les points de jonction situés entre 0 et 3 ans ne sont pas permis. Dans le tableau 2, les cellules des points de jonction situés entre 2001 et 2005 sont en jaune et celles situées entre 2006 et 2008 sont en orange.

Pour ce qui est du second indicateur, nous avons utilisé le rapport entre les BR, c'est-à-dire le ratio entre le BR pour 1986-2010 et celui pour 2011-2014. Nous avons estimé que le biais dans les taux projetés devrait au moins être égal ou supérieur au biais dans les taux utilisés pour réaliser les modèles de projection. Pour obtenir le BR pour 1986-2010, nous avons utilisé les résultats de l'analyse des points de jonction. Dans le tableau 2, les cellules sont en jaune lorsque le BR pour 2011-2014 était 2 à 5 fois plus élevé que le BR pour 1986-2010, et elles sont en orange lorsque le BR pour 2011-2014 était au moins 5 fois plus élevé que le BR pour 1986-2010. Ces valeurs seuils ont été établies de façon arbitraire à partir de l'examen de la distribution des résultats.

Résultats

Modèles Canproj

Cinq des six modèles (le modèle Nordpred, le modèle âge-cohorte, le modèle hybride fondé sur la tendance commune pour l'âge, le modèle hybride fondé sur la tendance spécifique par âge et le modèle fondé sur le calcul de la moyenne sur cinq ans) ont présenté un BR moyen situé entre 5 % et 10 % et un BR médian d'environ 5 % (figure 1). Une variation plus importante a été observée dans le BR moyen et le BR médian lors de la comparaison de l'exactitude des modèles de projection par siège de cancer (figure 2). Aucun des modèles n'a obtenu de bons résultats pour les prédictions relatives au cancer de la prostate, et une plus grande variabilité des prédictions a été observée pour les cancers de la glande thyroïde, du larynx, de la vessie, du foie et de l'encéphale/du système nerveux central (SNC).

Un portrait plus détaillé et légèrement différent se dégage lorsque les modèles graphiques sont comparés selon le siège de cancer et le sexe (figure 3, tableau 2). La performance de tous les modèles de projection s'est révélée mauvaise dans les cas suivants : tous les cancers confondus chez les hommes, cancer colorectal chez les hommes et les femmes, cancer de la prostate, cancer de la vessie chez les hommes, cancer de l'encéphale/du SNC chez les hommes et les femmes, lymphome de Hodgkin chez les femmes et myélome chez les hommes. La plus grande variation observée à la figure 2 en ce qui concerne les cancers du foie, du larynx et de la glande thyroïde semble attribuable à l'incapacité de certains modèles à prédire les taux.

Les sièges de cancer ayant montré des changements récents sur le plan des tendances et pour lesquels les projections pourraient potentiellement être améliorées en modifiant la période visée étaient les suivants : tous les cancers confondus chez les hommes, cancer colorectal chez les hommes et les femmes et cancer de la prostate. Nous avons exécuté des modèles hybrides distincts pour ces sièges de cancer en utilisant les sept dernières années de données seulement. Nous avons été en mesure d'améliorer considérablement l'exactitude des projections pour ces quatre sièges de cancer. Dans le cas du cancer colorectal, nous avons pu faire passer le

TABEAU 2
Biais relatif médian (%) selon le modèle et le type de cancer pour 2011-2014

Sexe	Siège de cancer	Modèle ^a					Diagnostic			
		Nordpred	Âge-cohorte	Hybride			Moyenne sur 5 ans	PJ ^b	Ratio BR ^c	BR (%)
				Tendance spécifique par âge	Tendance commune pour l'âge	Âge seulement				
Hommes	Tous les cancers	11,0	10,1	5,9	5,6	7,8	7,5	2007	5,7	1,0
	Bouche	8,8	13,4	14,8	12,7	1,0	6,3	2003	0,7	1,3
	Œsophage	2,6	2,2	3,0	2,7	5,9	2,2	2005	0,8	2,8
	Estomac	3,1	3,8	3,9	3,9	25,2	8,5	1986	1,6	1,9
	Colorectal	7,4	7,7	6,4	8,1	10,2	7,8	2008	16,0	0,4
	Foie	4,4	4,2	4,8	3,9	26,2	10,0	1986	1,0	4,0
	Pancréas	3,0	6,9	6,5	5,4	3,1	3,5	1997	2,3	1,3
	Larynx	1,4	3,6	2,4	2,0	44,0	18,2	1986	0,6	2,4
	Poumon et bronches	3,1	1,8	1,9	1,7	30,2	11,1	1986	1,3	1,3
	Mélanome	1,4	6,1	1,5	4,6	17,6	8,3	1986	0,7	2,0
	Sein	4,9	4,3	6,3	6,0	6,6	6,7	1986	0,6	6,9
	Prostate	48,4	90,1	41,4	44,4	33,8	33,1	2001	11,2	2,9
	Testicule	1,4	1,4	1,3	1,3	13,9	7,2	1986	0,3	4,4
	Vessie	10,1	15,9	12,4	14,7	9,8	9,3	1990	3,5	2,7
	Rein et bassinnet du rein	5,0	2,8	2,0	2,0	9,1	4,8	1998	1,0	1,9
	Encéphale/SNC	4,1	2,9	4,3	3,4	7,7	5,6	1986	1,5	2,0
	Glande thyroïde	3,7	17,0	13,0	13,0	48,9	27,4	1997	0,8	4,6
	Lymphome de Hodgkin	1,4	1,4	2,1	1,6	3,4	1,3	1986	0,5	2,6
	Lymphome non hodgkinien	7,7	7,4	6,6	7,3	8,5	7,8	2007	4,4	1,5
	Myélome	5,2	4,7	5,1	4,6	10,0	6,8	1986	1,3	3,6
Leucémie	6,2	3,8	6,1	5,2	0,8	3,5	1994	0,4	2,0	
Tous les autres cancers	4,0	4,1	2,8	2,8	4,5	3,6	2003	2,1	1,3	
Femmes	Tous les cancers	0,9	0,8	0,8	0,8	3,3	0,9	1986	1,0	0,8
	Bouche	3,1	4,2	4,2	4,2	1,7	2,8	1986	0,6	2,9
	Œsophage	1,1	1,1	1,5	0,9	6,8	1,3	1986	0,3	3,4
	Estomac	1,3	2,7	7,2	4,4	20,5	3,5	1992	0,6	2,0
	Colorectal	4,0	3,7	2,7	4,2	8,3	4,9	2000	5,5	0,5
	Foie	5,3	4,8	5,1	4,4	21,2	8,9	1986	0,7	6,4
	Pancréas	4,3	5,1	5,5	4,9	3,5	4,2	1986	1,4	2,5
	Larynx	10,0	13,6	17,3	15,1	64,1	28,5	1986	1,8	5,5
	Poumon et bronches	1,3	1,3	9,0	4,9	3,9	1,9	2006	2,1	0,6
	Mélanome	2,8	8,4	3,5	3,4	16,3	9,3	1992	1,4	2,0
	Sein	2,3	3,2	0,7	1,3	1,9	1,6	1991	0,4	1,9
	Col de l'utérus	3,4	4,7	1,4	1,4	22,3	8,6	2006	0,7	2,1
	Utérus	3,1	8,0	3,4	2,6	10,0	10,0	2005	1,8	1,5
	Ovaire	1,1	1,1	1,2	1,7	9,9	4,7	1986	0,6	1,7
	Vessie	14,1	14,2	13,1	14,5	10,1	9,7	1986	3,1	3,1
	Rein et bassinnet du rein	12,0	4,3	4,6	4,3	6,1	4,2	1986	1,8	2,4
	Encéphale/SNC	5,6	5,3	5,7	5,5	8,8	6,2	1986	2,2	2,4
	Glande thyroïde	4,5	6,3	5,4	5,9	50,1	21,7	2005	2,7	1,7
	Lymphome de Hodgkin	10,3	10,9	12,2	11,6	8,0	10,4	1986	2,4	3,4
	Lymphome non hodgkinien	5,3	4,9	4,4	4,9	5,8	6,0	1997	3,9	1,1
Myélome	3,8	4,1	3,9	3,9	6,8	4,2	1986	1,0	3,7	
Leucémie	14,3	5,0	4,7	6,8	2,6	4,8	2001	1,5	1,7	
Tous les autres cancers	3,1	4,2	3,6	4,1	3,0	3,1	2004	3,5	0,8	

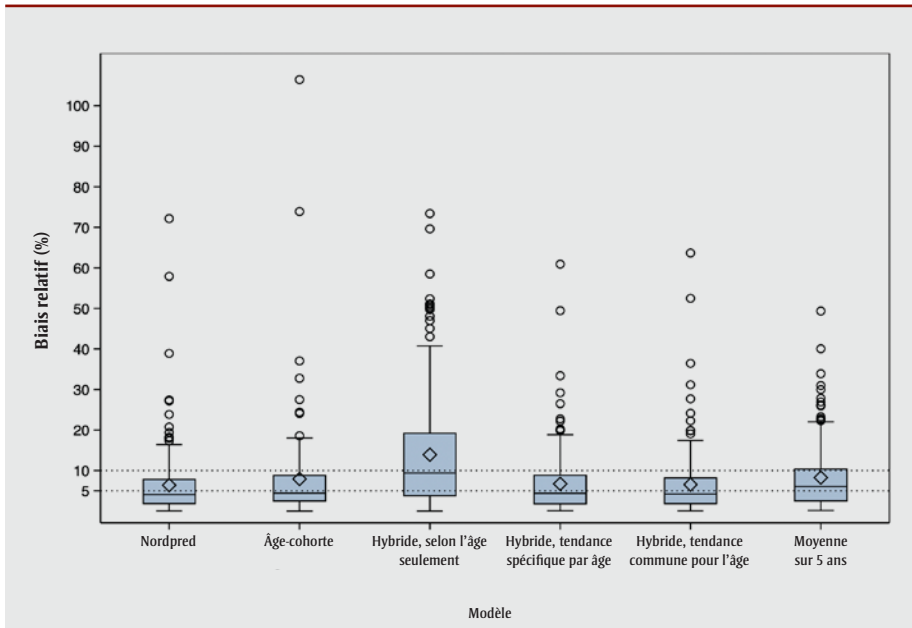
Abréviations : BR, biais relatif; PJ, points de jonction; SNC, système nerveux central.

^a Les modèles affichant le BR médian le plus faible pour 2011-2014 sont en vert pâle.

^b Année du point de jonction le plus récent pour les tendances relatives aux taux. Les points de jonction situés entre 2001 et 2005 sont en jaune, tandis que les points de jonction situés entre 2006 et 2008 sont en orange.

^c Le ratio BR correspond au ratio du BR médian pour 2011-2014 comparé au BR médian pour 1986-2010. Afin de mettre en évidence les sièges de cancer qui étaient plus difficiles à modéliser, les ratios BR continus ont été regroupés comme suit : cellules en jaune pour un BR médian pour 2011-2014 de 2 à 5 fois supérieur au BR médian pour 1986-2010; cellules en orange pour un BR médian pour 2010-2014 5 fois supérieur au BR médian pour 1986-2010.

FIGURE 1
Biais relatif selon le modèle de projection pour tous les sièges de cancer



Remarque : Voir le lien suivant pour une explication sur les diagrammes de quartiles : http://onlinestatbook.com/2/graphing_distributions/boxplots.html [en anglais seulement].

ratio de BR de 16,0 à 2,6 chez les hommes et de 5,5 à 1,9 chez les femmes. Nous avons également été en mesure de faire passer le ratio de BR de 5,7 à 5,4 pour tous les cancers confondus chez les hommes et de 11,2 à 7,7 pour le cancer de la prostate.

Arbres décisionnels de Canproj utilisés pour la sélection d'un modèle

Comme on le voit dans le tableau 3, l'effet de cohorte et le paramètre de dérive étaient significatifs dans 79 % des cas (soit pour 34 modèles sur 43), de sorte

que Nordpred a été le modèle le plus souvent sélectionné par Canproj. Cependant, Nordpred était le modèle présentant le BR le plus faible seulement 24 % du temps. Il n'en reste pas moins que le BR moyen se situait entre 0 et 5 % pour au moins l'un des six modèles dans 76 % des cas, et entre 6 % et 10 % pour au moins l'un des modèles dans 20 % des cas.

Analyse

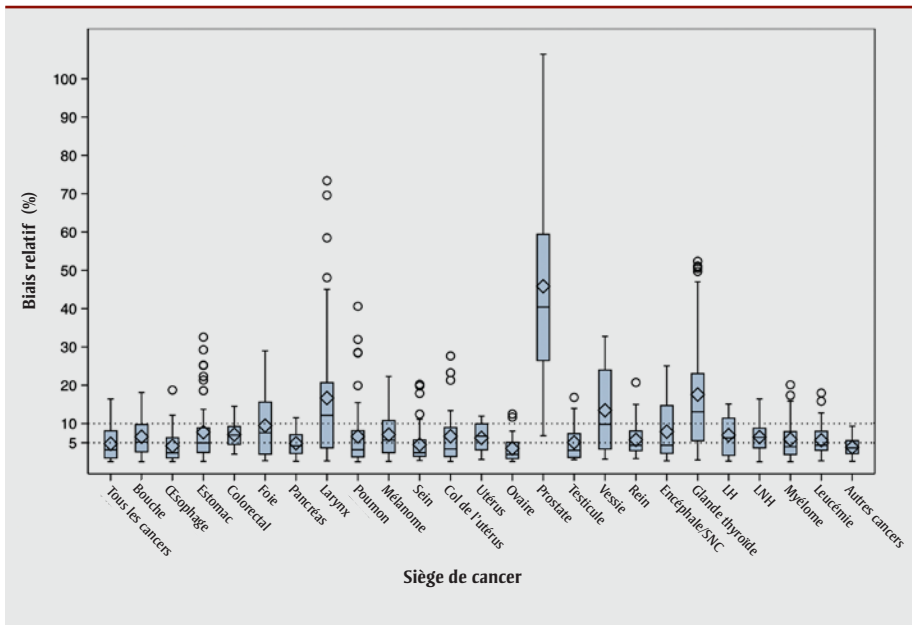
Notre objectif consistait à valider les projections à court terme générées par Canproj à l'aide des données sur l'incidence du cancer au Canada. Les résultats montrent que l'éventail de modèles offerts par Canproj valide l'établissement de projections fiables pour la plupart des sièges de cancer étudiés. Lorsque des variations ont été relevées au cours des 10 ans de données d'entraînement, on a utilisé une période plus courte comme base de projection afin d'améliorer l'exactitude des taux projetés.

La forte hausse des taux de cancer de la vessie en 2013-2014 découle de la modification des règles de déclaration en Ontario²⁰ : en effet, à partir de 2013, cette province a ajouté à son registre les cas de cancer in situ de la vessie dans la catégorie des tumeurs malignes de la vessie.

Le cancer de l'encéphale/du SNC, le cancer colorectal, le lymphome de Hodgkin chez les femmes, le cancer de la prostate et tous les cancers confondus chez les hommes connaissent un déclin plus rapide que ce que prévoient les modèles, alors que le myélome chez les hommes connaît une croissance plus rapide. La piètre performance en matière de prédiction de ces taux de cancer est liée aux changements récents et rapides dans des taux qui ne faisaient pas partie de l'ensemble de données d'entraînement ou qui sont survenus au cours des dernières années de l'ensemble de données d'entraînement.

Nous avons évalué les fonctions de sélection automatique du modèle offertes par Canproj (arbres décisionnels) pour déterminer si elles permettaient de choisir le meilleur modèle. En ce qui concerne l'ensemble de données nationales, le modèle Nordpred était celui qui a été le plus souvent sélectionné par l'arbre décisionnel Canproj, même s'il offrait le BR le plus faible dans seulement 24 % des cas. D'autres modèles peuvent surpasser le

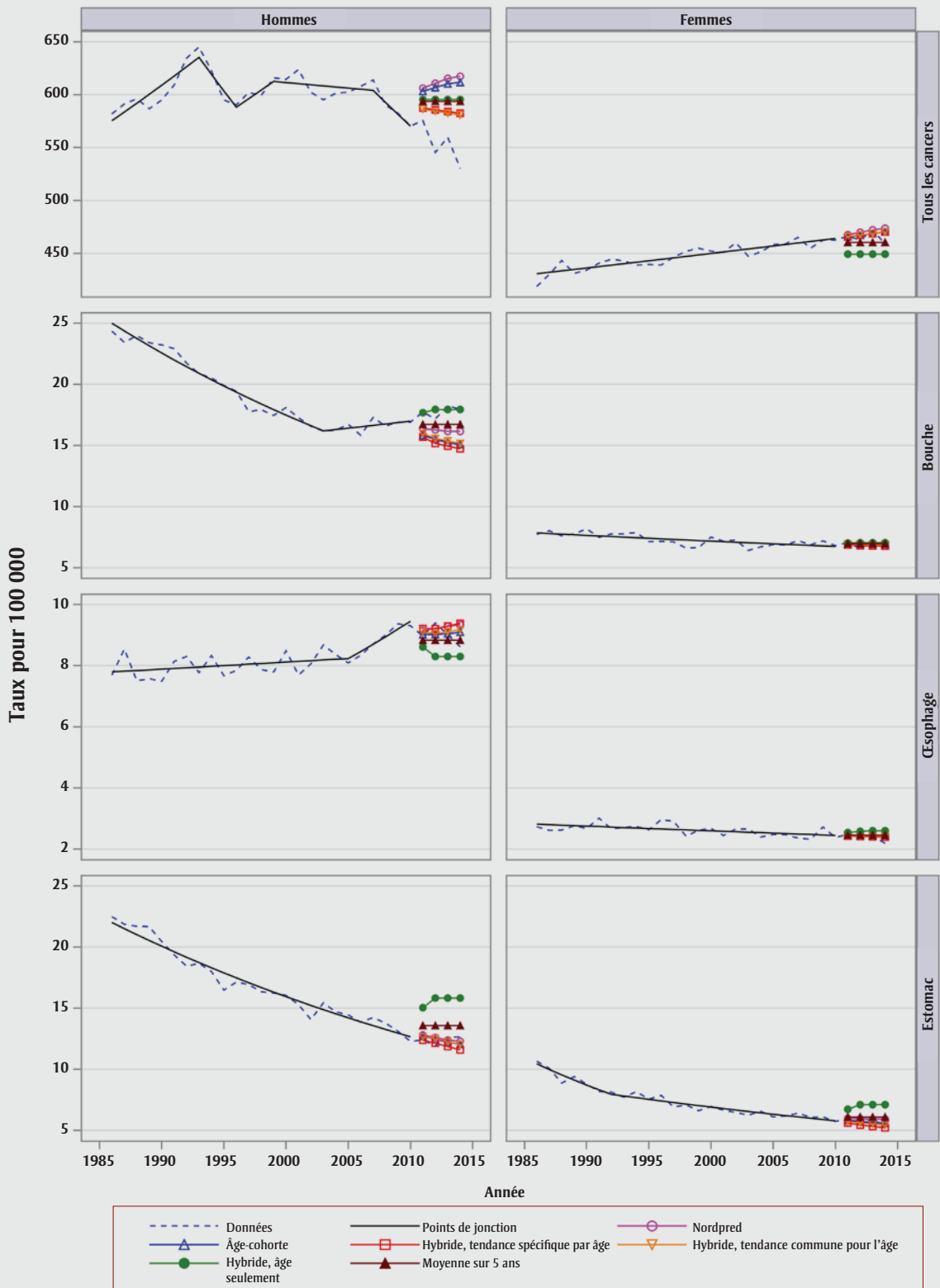
FIGURE 2
Biais relatif selon le siège de cancer pour tous les modèles de projection



Abréviations : LH, lymphome de Hodgkin; LNH, lymphome non hodgkinien; SNC, système nerveux central.

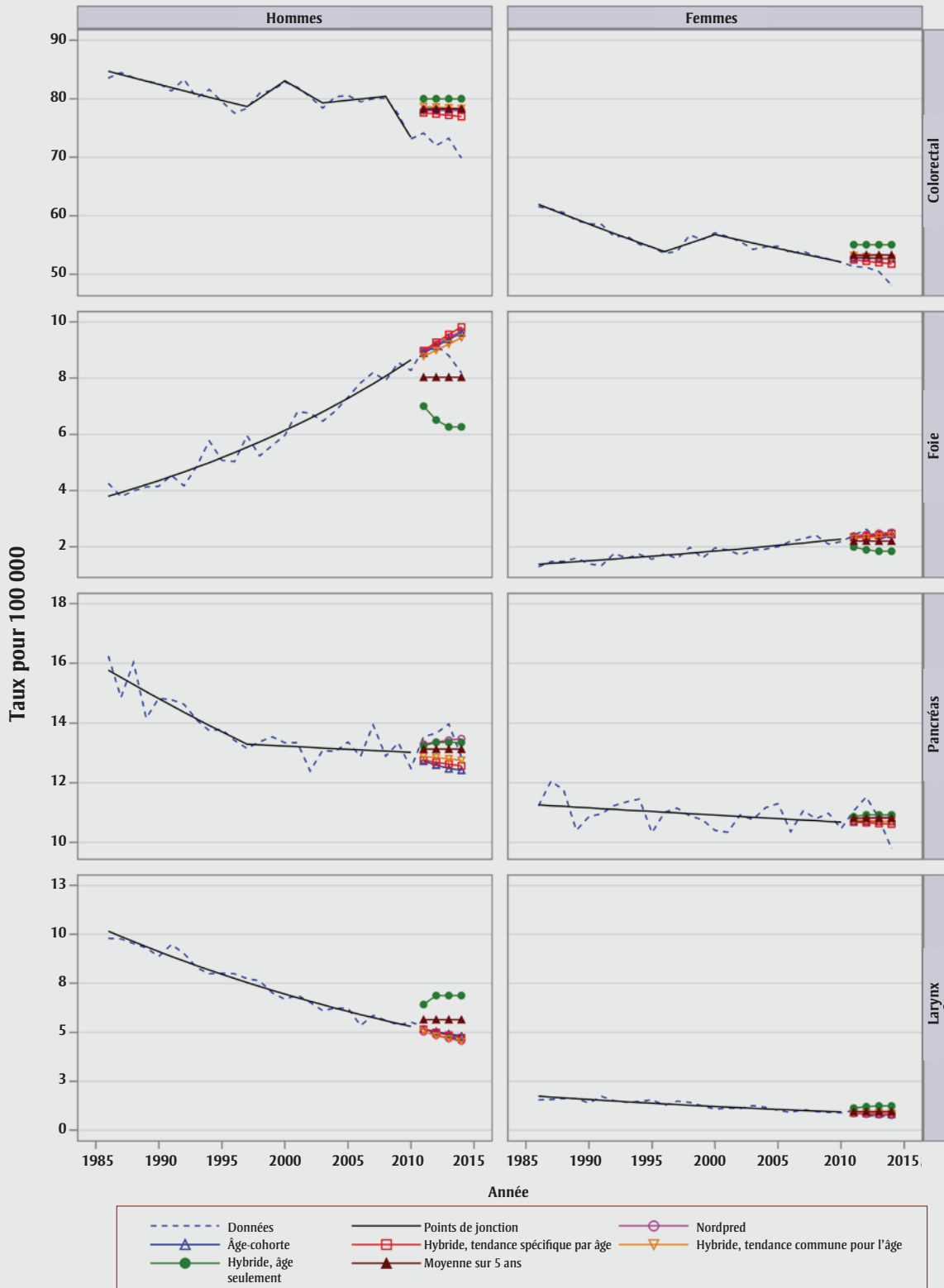
Remarque : Voir le lien suivant pour une explication sur les diagrammes de quartiles : http://onlinestatbook.com/2/graphing_distributions/boxplots.html [en anglais seulement].

FIGURE 3
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



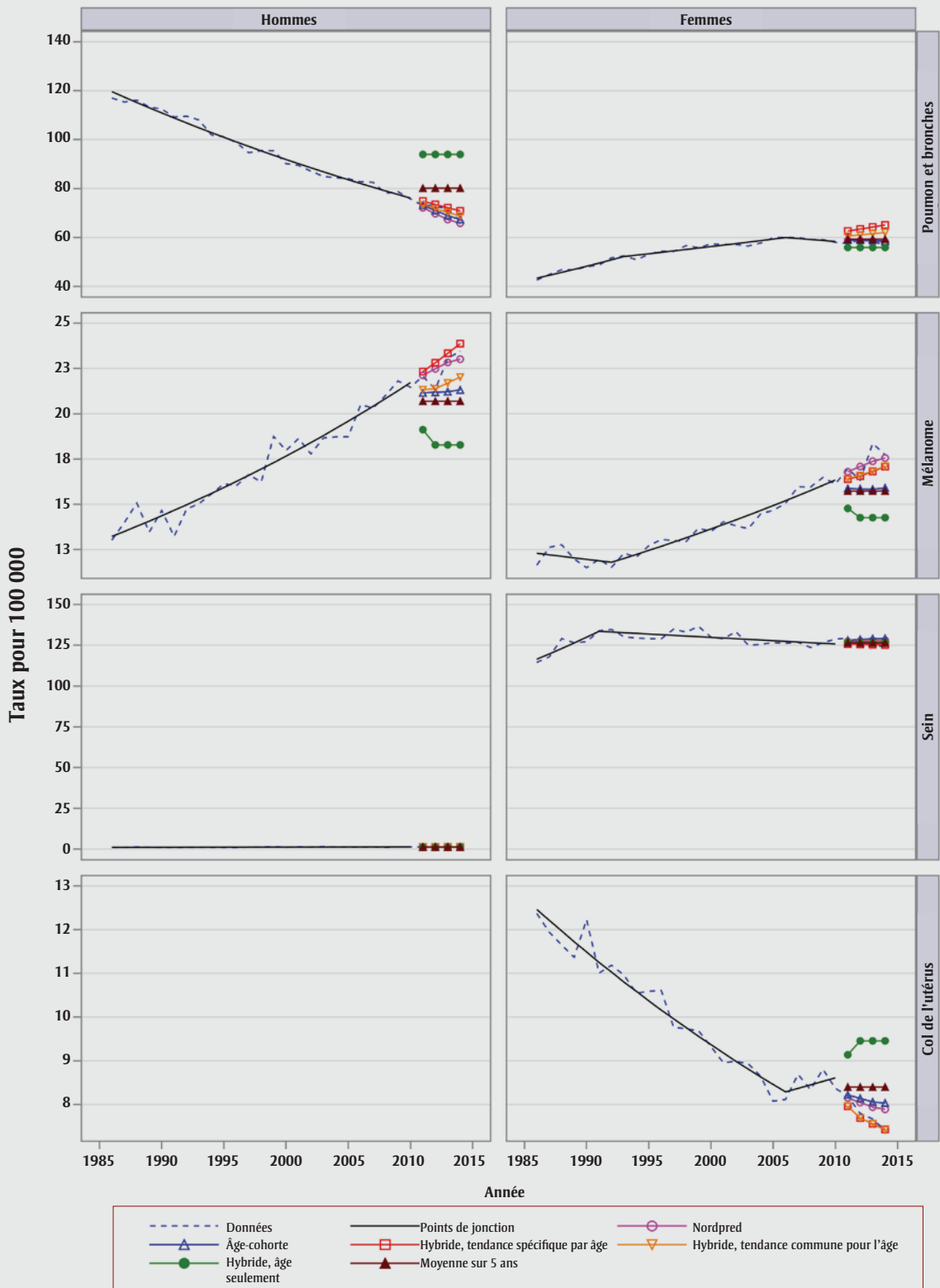
Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
 à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



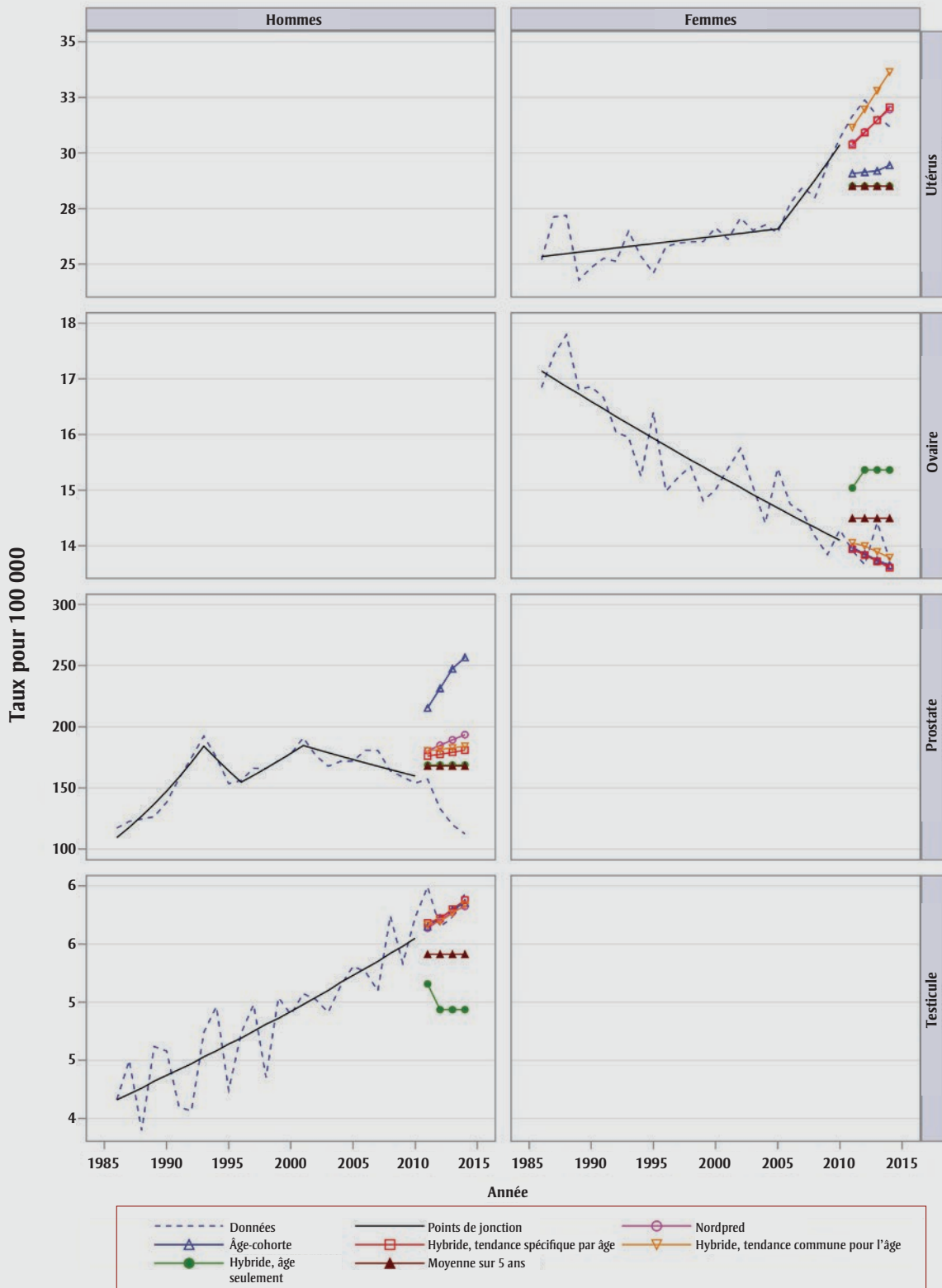
Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014) à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



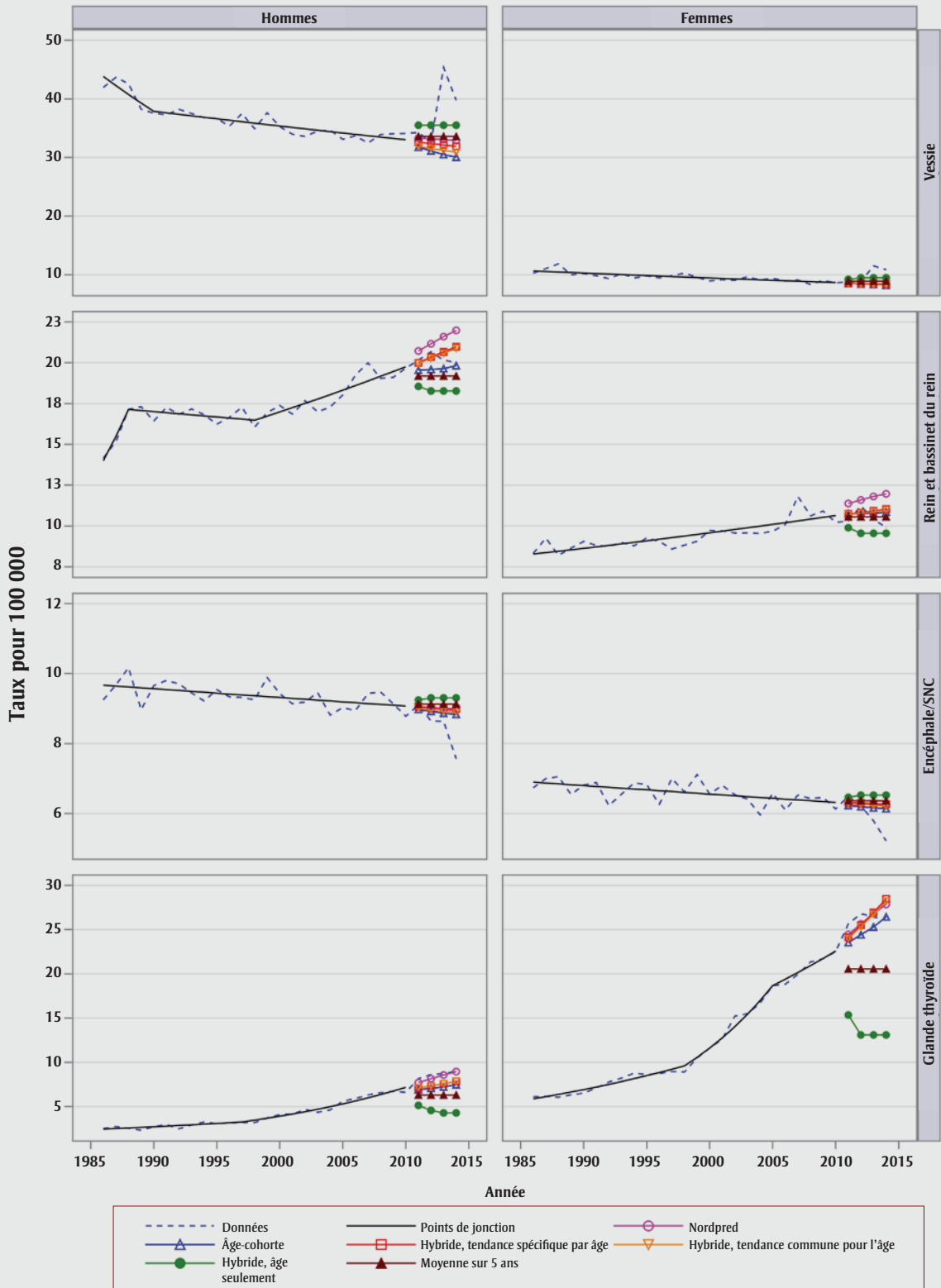
Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
 à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



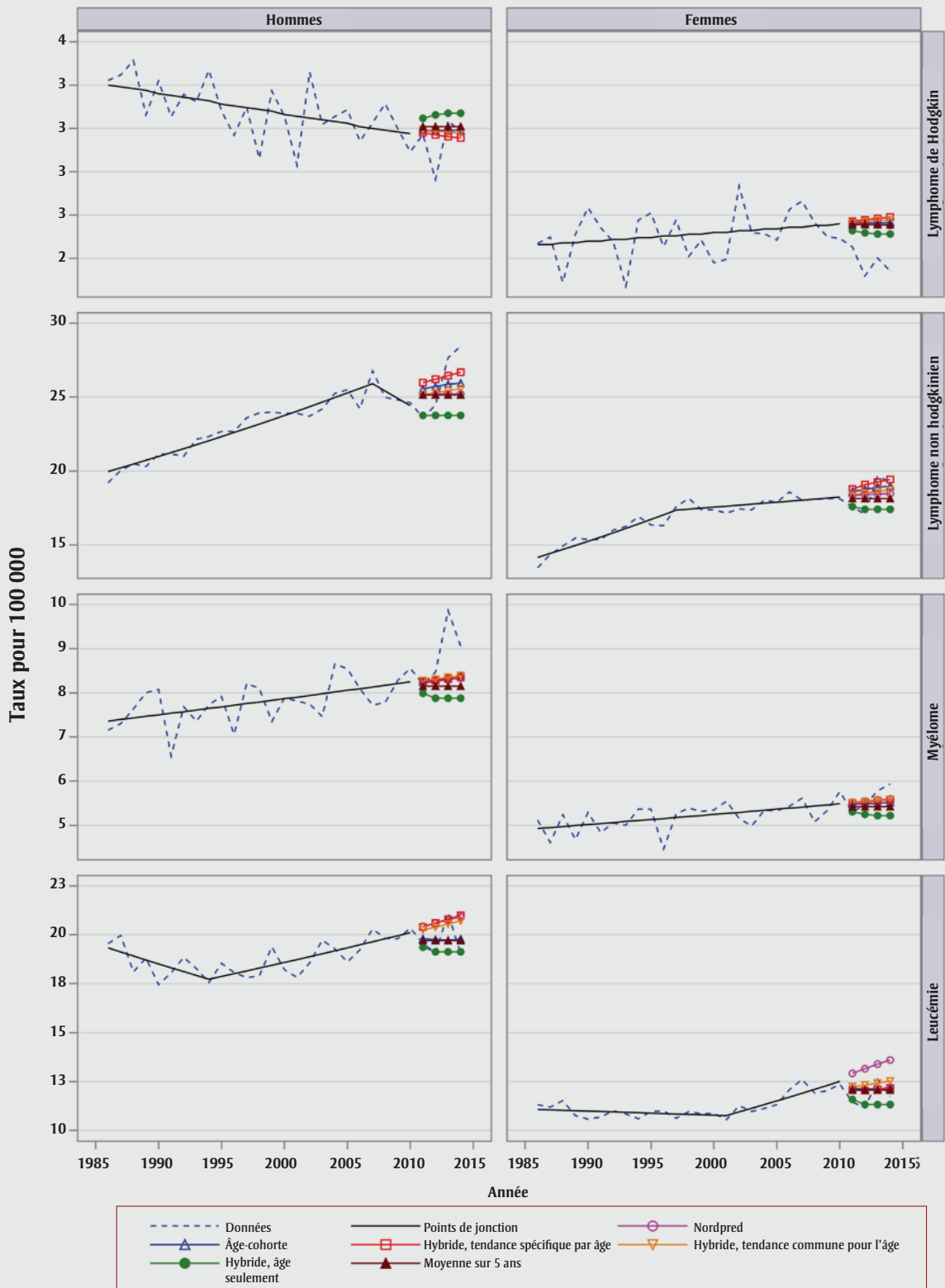
Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



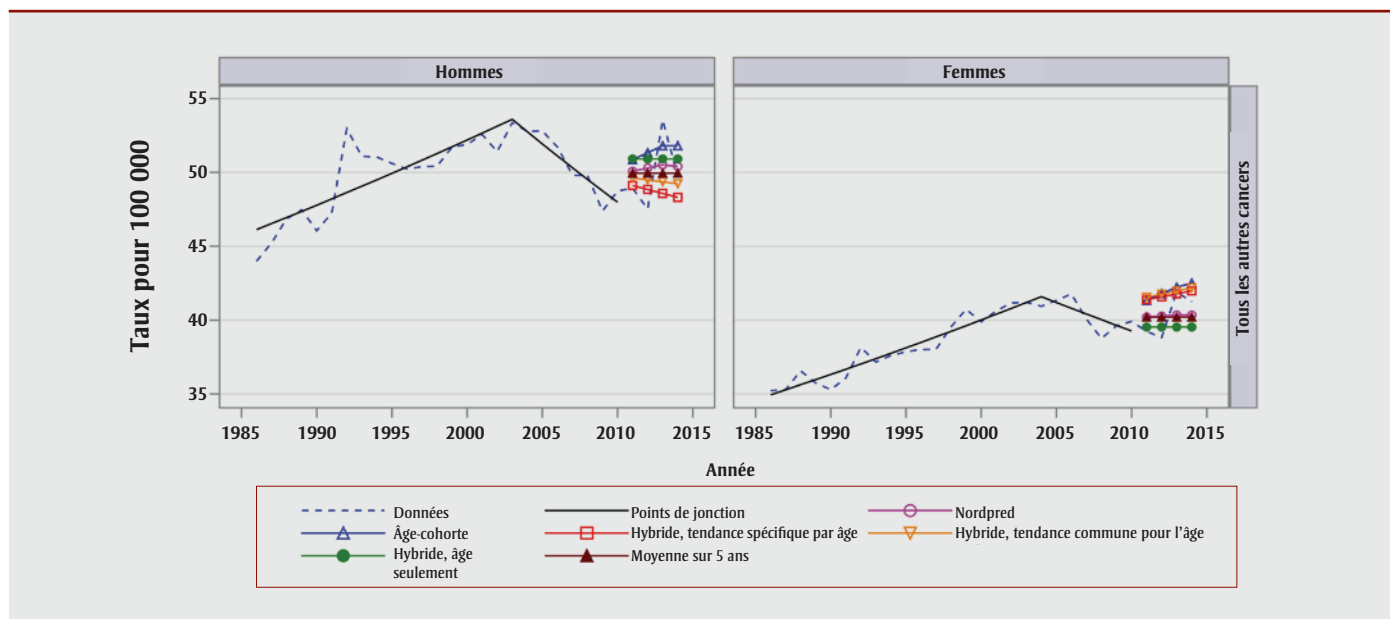
Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
 à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



Suite à la page suivante

FIGURE 3 (suite)
Taux d'incidence normalisés selon l'âge réels (1986 à 2010) et taux d'incidence normalisés selon l'âge projetés (2011 à 2014)
à l'aide des modèles de projection Canproj, en fonction du sexe et du siège de cancer, Canada



modèle Nordpred pour les analyses de données portant sur de plus petites populations²¹. D'après notre expérience personnelle et celle d'autres auteurs avec Canproj, il faudrait utiliser l'arbre décisionnel de sélection conjointement avec les résultats individuels de chaque modèle et les conseils d'experts pour choisir le meilleur modèle de projection²².

Les résultats de ce projet s'appuient sur plusieurs études antérieures menées au Canada, qui ont examiné différentes méthodes de projection du cancer. Lee et ses collaborateurs ont comparé, en 2011, l'exactitude de 16 modèles et variations de modèles de projection à court terme des taux de mortalité par cancer²³. Ils ont constaté qu'il n'existait pas de méthode unique permettant d'établir des prévisions exactes pour un large éventail de sièges de cancer et qu'il est préférable d'utiliser un éventail de modèles. Qiu et ses collaborateurs ont comparé, en 2010, le modèle Nordpred, le modèle additif généralisé et le modèle bahésien⁸. Ils ont conclu qu'en présence des effets d'âge, de dérive et de cohorte, la méthode Nordpred est l'approche à privilégier, qu'en présence des effets d'âge et de cohorte, c'est le modèle âge-cohorte qui représente la meilleure approche et qu'en l'absence d'effets de cohorte, c'est une méthode hybride qui devrait être utilisée. Qiu et ses collaborateurs ont également constaté que, dans le cas des sièges de cancer pour

lesquels les cas sont peu nombreux, il faut regrouper les données pour pouvoir appliquer la méthode hybride. En 2010, le Cancer Projections Network (C-Proj) du Canada a diffusé un rapport offrant une évaluation du modèle Nordpred, du modèle hybride, du modèle âge-cohorte et du modèle bayésien à l'aide des méthodes de projection de l'incidence du cancer de Monte-Carlo par chaînes de Markov au moyen des données du Registre du cancer de la Nouvelle-Écosse²¹. Selon le C-Proj, la méthode âge-cohorte devrait être utilisée pour les projections relatives au cancer dans les provinces où les effectifs de population sont faibles et stables.

Bien que l'on procède régulièrement à l'établissement de projections relatives à l'incidence du cancer, seules quelques études ont évalué les méthodes de remplacement et leurs recommandations dépendent de la population visée et de la période de projection. Stock et ses collaborateurs ont utilisé, en 2018, une approche bayésienne pour projeter les taux d'incidence du cancer jusqu'en 2030 à l'aide de données du registre du cancer de l'Allemagne²⁴. Ils ont constaté que cette méthode offrait des avantages sur le plan de la souplesse, de la facilité d'interprétation, de la transparence et du niveau de détail, mais n'ont pas recommandé son utilisation pour les données à court terme. Pesola et ses collaborateurs ont comparé, en 2017, divers modèles (nul, âge-dérive,

âge-période, âge-cohorte et APC) pour prédire l'incidence du cancer chez les enfants et les adolescents en Angleterre jusqu'en 2030²⁵. Les résultats de l'ajustement du modèle révèlent que le modèle âge-dérive était bien adapté aux données, tout comme les modèles plus complexes pour tous les cancers chez les enfants. Un modèle APC avec splines cubiques naturelles a été évalué pour prédire l'incidence du cancer et de la mortalité par cancer au Royaume-Uni jusqu'en 2035²⁶. Le modèle APC repose sur l'hypothèse d'une continuité des tendances antérieures pour le futur. Si l'on met au point des vaccins ou de nouveaux traitements qui ont une influence sur l'incidence du cancer ou sur la mortalité par cancer, le modèle ne sera pas en mesure d'anticiper ces changements, d'où l'importance d'utiliser des données récentes et de réaliser des projections à intervalles réguliers²⁷. Katanoda et ses collaborateurs ont examiné, en 2014, trois modèles de projection pour établir des projections de l'incidence du cancer à court terme au Japon : modèle linéaire généralisé utilisant l'âge et la période comme variables indépendantes (modèle linéaire A + P); modèle linéaire généralisé utilisant l'âge, la période et leurs interactions (modèle linéaire A * P); et modèle additif généralisé avec spline de lissage de l'âge, de la période et de leurs interactions (modèle avec spline A * P)²⁸. Ils ont utilisé Nordpred dans leur analyse préliminaire,

TABEAU 3
Arbre décisionnel Canproj : biais relatif moyen selon le modèle, le sexe et le siège de cancer

Sexe	Type de cancer	Modèle					Moyenne sur 5 ans
		Nordpred	Âge-cohorte	Hybride			
				Tendance spécifique par âge	Tendance commune pour l'âge	Âge seulement	
Hommes	Tous les cancers	10,9	10,1	5,9	5,7	7,8	7,5
	Bouche	8,6	13,2	14,8	12,6	1,6	5,9
	Œsophage	3,9	2,6	4,2	3,2	6,8	3,0
	Estomac	3,4	3,7	4,1	3,7	26,1	9,5
	Colorectal	7,9	8,2	6,9	8,6	10,7	8,3
	Foie	6,8	6,5	7,6	6,2	26,0	8,6
	Pancréas	3,2	7,0	6,2	5,0	3,2	3,8
	Larynx	2,1	3,6	2,5	2,2	39,4	16,2
	Poumon et bronches	3,2	1,7	2,6	1,6	32,4	12,9
	Mélanome	2,1	5,5	2,8	3,9	17,6	7,8
	Sein	6,2	6,8	8,6	8,3	6,4	6,2
	Prostate	45,9	86,4	38,9	41,8	31,3	30,6
	Testicule	2,4	2,4	2,2	2,1	14,3	7,1
	Vessie	12,3	17,3	13,7	15,9	11,2	11,6
	Rein et bassinnet du rein	5,5	3,0	2,4	2,3	9,4	5,3
	Encéphale/SNC	6,7	6,0	7,0	6,3	10,0	8,0
	Glande thyroïde	3,3	16,8	13,4	13,5	46,8	26,9
	Lymphome de Hodgkin	3,4	3,4	3,5	3,4	5,5	3,8
	Lymphome non hodgkinien	7,5	7,2	6,9	7,0	8,5	7,5
	Myélome	6,6	6,3	6,6	6,4	10,7	7,9
Leucémie	6,0	3,9	6,0	5,5	2,8	3,6	
Tous les autres cancers	3,8	4,8	3,8	3,6	4,6	3,6	
Femmes	Tous les cancers	1,3	1,1	1,0	1,1	3,4	1,1
	Bouche	3,4	4,1	4,4	3,9	2,9	3,3
	Œsophage	3,6	3,5	3,2	3,0	8,8	3,9
	Estomac	2,7	2,9	7,2	4,1	20,7	4,9
	Colorectal	5,2	4,9	3,7	5,4	9,5	6,1
	Foie	5,4	4,9	5,2	5,5	21,8	9,0
	Pancréas	4,9	5,1	5,2	5,0	4,9	4,7
	Larynx	10,2	14,5	16,7	15,3	53,4	24,0
	Poumon et bronches	1,6	1,6	10,3	5,8	3,5	2,4
	Mélanome	3,0	8,4	4,2	4,1	16,9	9,1
	Sein	2,3	2,9	1,2	1,5	1,8	1,7
	Col de l'utérus	3,4	4,6	1,5	1,5	20,9	8,3
	Utérus	2,8	7,9	3,0	3,6	10,1	10,1
	Ovaire	1,9	1,8	1,8	1,8	9,7	4,0
	Vessie	14,6	14,7	13,7	15,0	10,7	10,7
	Rein et bassinnet du rein	12,0	5,4	6,0	5,8	7,7	4,1
	Encéphale/SNC	7,8	7,1	8,2	7,7	10,8	9,1
	Glande thyroïde	4,3	5,7	5,5	5,3	47,9	21,6
	Lymphome de Hodgkin	10,1	10,5	11,5	11,0	7,9	10,0
	Lymphome non hodgkinien	5,6	5,3	5,4	5,3	5,8	5,7
Myélome	3,7	3,9	3,8	3,8	6,4	4,4	
Leucémie	13,2	5,6	5,2	6,4	4,1	5,1	
Tous les autres cancers	3,0	4,2	3,7	4,0	3,1	3,2	

Abréviations : BR, biais relatif; SNC, système nerveux central.

Remarques : Les cellules en vert pâle correspondent aux modèles de projection sélectionnés par Canproj; les cellules lilas correspondent aux modèles présentant le BR le plus faible; les cellules vert foncé indiquent que le modèle sélectionné par Canproj est le modèle présentant le plus BR le plus faible.

et celui-ci n'a pas été en mesure de prédire que le cancer du foie atteindrait un sommet au milieu des années 1990.

Points forts et limites

Ce projet comporte plusieurs limites. Dans tous les modèles utilisés par Canproj, les variables de l'âge, de la période et de la cohorte tiennent compte de l'ensemble des améliorations et des changements relatifs aux facteurs de risque, aux caractéristiques et au profil ethnique de la population ainsi qu'à la prévention, à la détection précoce et au traitement. Le fait d'avoir plus de détails sur ces déterminants des taux de cancer améliorerait la capacité d'effectuer des projections plus fiables. Cependant, ce niveau d'information nécessaire serait difficile à obtenir dans certaines provinces et certains territoires et, pour la plupart des sièges de cancer étudiés dans le cadre du projet, les renseignements sur l'âge, la période et la cohorte se sont révélés suffisants pour effectuer des projections fiables.

Nous n'avons pas mené d'analyse exhaustive des données provinciales dans le cadre de cet exercice, mais nous prévoyons que d'autres modèles que Nordpred seraient le plus souvent sélectionnés par l'arbre décisionnel.

Les données du Québec étaient exclues des données utilisées dans les modèles, ce qui fait que les données ne sont pas représentatives de l'ensemble du pays.

Enfin, comme pour toutes les méthodes, les projections s'appuient sur l'hypothèse selon laquelle les tendances antérieures vont se maintenir dans le futur, ce qui ne sera pas forcément toujours le cas.

Conclusion

Les planificateurs de services de santé et les décideurs doivent disposer de données sur le fardeau à venir du cancer pour pouvoir établir des priorités en matière de stratégies de lutte contre le cancer, répartir les ressources et évaluer les traitements et les interventions. Le progiciel Canproj est apte à fournir des projections fiables sur le cancer afin de les appuyer dans ces tâches.

Conflits d'intérêts

Les auteurs déclarent n'avoir aucun conflit d'intérêts.

Contributions des auteurs et avis

AD, ZQ et AS ont participé à la conception et à la conceptualisation des travaux. AD et ZQ ont participé à l'analyse des données. AD a rédigé l'article. Tous les auteurs ont contribué à l'interprétation des résultats et à la révision de l'article.

Le contenu de l'article et les points de vue qui y sont exprimés n'engagent que les auteurs et ne correspondent pas nécessairement à ceux du gouvernement du Canada.

Références

1. Comité consultatif de la Société canadienne du cancer. Statistiques canadiennes sur le cancer 2011. Ottawa (Ont.) : Gouvernement du Canada; 2011.
2. Fekjær H, Møller B. Nordpred software package. Majorstuen (NO): Cancer Registry of Norway; [consultation le 20 juillet 2017]. En ligne à : <https://www.kreftregisteret.no/en/Research/Projects/Nordpred/Nordpred-software/>
3. Møller B, Fekjaer H, Hakulinen T, et al. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Stat Med.* 2003;22(17):2751-2766. doi:10.1002/sim.1481.
4. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med.* 1987;6(4):4694-81. doi:10.1002/sim.4780060406.
5. Weir HK, White MC. Cancer incidence and mortality through 2020. *Prev Chronic Dis.* 2016;13:E48. doi:10.5888/pcd13.160024.
6. Rapiti E, Guarnori S, Pastoors B, Miralbell R, Usel M. Planning for the future: cancer incidence projections in Switzerland up to 2019. *BMC Public Health.* 2014;14(1):102. doi:10.1186/1471-2458-14-102.
7. Nowatzki J, Moller B, Demers A. Projection du taux d'incidence du cancer et du nombre de nouveaux cas de cancer au Manitoba pour la période 2006-2025. *Maladies chroniques au Canada.* 2011;31(2):71-78.

8. Qiu Z, Jiang Z, Wang M, Hatcher J. Long-term projection methods: comparison of age-period-cohort model-based approaches. Technical report for Cancer Projections Network (C-Proj). Edmonton (AB): Alberta Health Services; 2010.
9. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2000;19:335-351. doi:10.1002/(sici)1097-0258(20000215)19:3 <335::aid-sim336 >3.0.co;2-z.
10. Comité consultatif de la Société canadienne du cancer. Statistiques canadiennes sur le cancer 2017. Ottawa (Ont.) : Gouvernement du Canada; 2017.
11. Fritschi L, Chan J, Hutchings SJ, Driscoll TR, Wong AY, Carey RN. The future excess fraction model for calculating burden of disease. *BMC Public Health.* 2016;16(1):386. doi:10.1186/s12889-016-3066-1.
12. Alberta Health Services, Surveillance and Reporting. The 2017 report on cancer statistics in Alberta. Edmonton (AB): Alberta Health Services; 2017.
13. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Statist Surv.* 2010;4:40-79. doi:10.1214/09-SS054.
14. Statistique Canada. Registre canadien du cancer (RCC). Information détaillée pour 2017. Ottawa (Ont.) : Statistique Canada; 2019 [consultation le 3 mai 2019]. En ligne à : https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3207
15. Working Group Report. International rules for multiple primary cancers (ICD-0 third edition). *Eur J Cancer Prev.* 2005;14(4):307-8. doi:10.1097/00008469-200508000-00002.
16. Statistique Canada. Produits de données, Recensement de 2016. Ottawa (Ont.) : Statistique Canada; 2019 [consultation le 6 mars 2019]. En ligne à : <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-fra.cfm>

17. Statistique Canada. Le Registre canadien du cancer : normalisation selon l'âge. Ottawa (Ont.) : Statistique Canada; 2019 [consultation le 5 mars 2019]. https://www.statcan.gc.ca/fra/programmes-statistiques/document/3207_D12_V3
18. Qiu Z, Hatcher J; Cancer Projection Analytical Network Working Team. Canproj: the R package of cancer projection methods based on generalized linear models for age, period, and/or cohort: Version I. 2013.
19. National Cancer Institute, Division of Cancer Control & Population Sciences, Surveillance Research Program. Joinpoint trends analysis software. Bethesda (MD): National Institutes of Health; 2019 [consultation le 6 mars 2019]. En ligne à : <https://surveillance.cancer.gov/joinpoint/>
20. Cancer Care Ontario. Ontario cancer statistics 2018. Toronto (Ont.) : Cancer Care Ontario; 2018.
21. Qiu Z, Wang H, Wang M, Dewar R, Hatcher J; Cancer Projection Analytical Network Working Team. Comparison of projection methods: validation analysis using Nova Scotia cancer registry database. Edmonton (AB): Alberta Health Services and Cancer Care Nova Scotia; 2011.
22. Poirier AE, Ruan Y, Walter SD, et al.; ComPARE Study Team. The future burden of cancer in Canada: Long-term cancer incidence projections 2013-2042. *Cancer Epidemiol.* 2019; 59:199-207. doi:10.1016/j.canep.2019.02.011.
23. Lee TC, Dean CB, Semenciw R. Short-term cancer mortality projections: a comparative study of prediction methods. *Stat Med.* 2011;30(29):3387-3402. doi:10.1002/sim.4373.
24. Stock C, Mons U, Brenner H. Projection of cancer incidence rates and case numbers until 2030: a probabilistic approach applied to German cancer registry data (1999-2013). *Cancer Epidemiol.* 2018;57:110-119. doi:10.1016/j.canep.2018.10.011.
25. Pesola F, Ferlay J, Sasieni P. Cancer incidence in English children, adolescents and young people: past trends and projections to 2030. *Br J Cancer.* 2017;117(12):1865-1873. doi:10.1038/bjc.2017.341.
26. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. *Br J Cancer.* 2016;115(9):1147-1155. doi:10.1038/bjc.2016.304.
27. De Souza Giusti AC, De Oliveira Salvador PT, Dos Santos J, et al. Trends and predictions for gastric cancer mortality in Brazil. *World J Gastroenterol.* 2016;22(28):6527-6538. doi:10.3748/wjg.v22.i28.6527.
28. Katanoda K, Kamo K, Saika K, et al. Short-term projection of cancer incidence in Japan using an age-period interaction model with spline smoothing. *Jpn J Clin Oncol.* 2014;44(1):36-41. doi:10.1093/jjco/hyt163.