

N° 11-633-X au catalogue — N° 007
ISSN 2371-3437
ISBN 978-0-660-08654-5

Études analytiques : méthodes et références

Base de données longitudinales sur l'immigration (BDIM) Rapport technique, 2014

par Rose Evra et Elena Prokopenko

Date de diffusion : le 16 juin 2017

 Statistique
Canada Statistics
Canada

Canada 

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-514-283-9350

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^p provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- ^E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2017

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Base de données longitudinales sur l'immigration(BDIM) Rapport technique, 2014

par

Rose Evra et Elena Prokopenko

Division de la statistique sociale et autochtone

11-633-X — N° 007

ISSN 2371-3437

ISBN 978-0-660-08654-5

Juin 2017

Direction des études analytiques : méthodes et références

Les documents de cette série traitent des méthodes utilisées pour produire des données qui seront employées pour effectuer des études analytiques à Statistique Canada sur l'économie, la santé et la société. Ils ont pour but de renseigner les lecteurs sur les méthodes statistiques, les normes et les définitions utilisées pour élaborer des bases de données à des fins de recherche. Tous les documents de la série ont fait l'objet d'un examen par les pairs et d'une révision institutionnelle, afin de veiller à ce qu'ils soient conformes au mandat de Statistique Canada et qu'ils respectent les normes généralement reconnues régissant les bonnes pratiques professionnelles.

Les documents peuvent être téléchargés gratuitement de www.statcan.gc.ca.

Table des matières

Remerciements	7
Résumé	8
Glossaire	9
1 Introduction	10
2 Sources des données	11
2.1 Données sur l'immigration.....	11
2.1.1 Fichier des immigrants reçus (FIR)	11
2.1.2 Fichier des résidents non permanents (FRN).....	11
2.1.3 Fichier intégré des résidents permanents et non permanents (FRPN).....	11
2.1.4 Fichier des résidents non permanents, permis (FRN, permis).....	12
2.2 Fichier sur la famille T1	12
2.3 Fichiers auxiliaires.....	12
3 Concepts et variables	13
3.1 Statut d'immigrant au Canada.....	13
3.1.1 Aperçu de l'immigration au Canada	13
3.2 Population cible et période de couverture	14
3.3 Variables relatives à l'admission	14
3.3.1 Catégorie d'admission.....	14
3.3.2 Type de demandeur	16
3.3.3 Changements au fil du temps.....	17
3.3.4 Variables de catégorie d'admission du FRPN	17
3.4 Variables d'intérêt	18
3.4.1 Variables géographiques	19
3.4.2 Variables temporelles.....	19
3.4.3 Variables relatives à la scolarité.....	20
3.4.4 Variables de profession envisagée	21
3.4.5 Autres variables de la BDIM	21
4 Couplage d'enregistrements	22
5 Traitement des données	23
5.1 Traitement	23
5.2 Couplage du Fichier des résidents non permanents (FRN).....	24
5.3 Variables dérivées incluses dans le FFT1	25
5.4 Variables dérivées incluses dans le FRPN.....	25
5.5 Détection des valeurs aberrantes.....	26

6	Diffusion	27
6.1	Produits analytiques	27
6.2	Demande de fichiers analytiques	28
6.3	Autres programmes statistiques utilisant les données de la BDIM	28
6.4	Confidentialité.....	28
7	Évaluation des données et indicateurs de qualité	30
7.1	Sources d'erreurs	30
7.1.1	Erreurs de couplage d'enregistrements.....	30
7.1.2	Erreurs de mesure	30
7.1.3	Erreurs de couverture	30
7.2	Exactitude des données	31
7.2.1	BDIM de 2014 : taux de couplage.....	31
7.2.2	Disponibilité de la date de décès	33
7.2.3	Déclarants avant l'admission par rapport aux enregistrements du Fichier des résidents non permanents (FRN).....	34
7.2.4	Indicateur de conjoint	35
7.3	Imputation des variables relatives à la scolarité	36
7.4	Couverture	37
7.4.1	Couverture du Fichier intégré des résidents permanents et non permanents (FRPN)	37
7.4.2	Taille du Fichier sur la famille T1 et couverture par année.....	38
7.5	Évaluation de la qualité du Fichier intégré des résidents permanents et non permanents (FRPN).....	40
8	Comparabilité	43
8.1	Variation historique de la couverture	43
8.2	Changements méthodologiques	43
8.3	Changements du contenu de la base de données historiques	43
8.4	Comparabilité avec d'autres sources des données sur l'immigration	44
8.4.1	Banque de données administratives longitudinales (DAL)	44
8.4.2	Recensement.....	45
8.4.3	Enquête longitudinale auprès des immigrants du Canada (ELIC)	46
9	Nouvelles analyses possibles grâce à la BDIM	48
9.1	Possibilités analytiques découlant des données relatives aux résidents non permanents	48
9.2	Possibilités analytiques grâce aux données relatives aux décès	48
10	Résumé	49
Annexe	50
A)	Liens vers les principaux documents et pages Web relatifs à la BDIM.....	50
B)	Couverture	50
C)	Analyses précédentes	52

D)	Pratiques exemplaires et conseils pour les analystes	52
D.1	Conseils de programmation	52
D.2	Création d'une cohorte	54
D.3	Calcul des taux de rétention	55
D.4	Calcul des trajectoires de revenu au cours du temps	58
D.5	Arrondissement des données	59
D.6	Détection de valeurs aberrantes	59
D.7	Ajustement du revenu selon l'Indice des prix à la consommation (IPC)	61
D.8	Calcul de mesures de revenu clés	61
	Bibliographie	63

Remerciements

Nous souhaitons souligner la contribution particulière des personnes suivantes : **Laetitia Martin** de la Division de la statistique sociale et autochtone (DSSEA), qui a rédigé les sections 3.3.1 à 3.3.3 de ce rapport, **Alexandr Diaz-Papkovich** de la Division des méthodes d'enquêtes sociales (DMES), qui a vérifié les sections du rapport relatives à la méthodologie sous-tendant la production de la BDIM, ainsi que **Tristan Cayn**, **Derek Cho**, **Scott McLeish**, **Ian Marrs** et **Trevor Smith**, membres de l'équipe des données administratives, qui produisent la BDIM et ont contribué au contenu de plusieurs sections du rapport.

De plus, nous remercions les personnes suivantes d'avoir relu le rapport avant sa publication : **Margareta Dovgal**, **Benoît St-Jean**, **Winnie Chan** et **Hélène Maheux** (Statistique Canada), **Yoko Yoshida** (Département de sociologie et d'anthropologie sociale, Université Dalhousie), **Michael Haan** (Chaire de recherche du Canada en migration et relations ethniques et Département de sociologie de l'Université de Western Ontario) et **Ian Clara** (Centre de données de recherche du Manitoba).

Résumé

La base de données longitudinales sur l'immigration (BDIM) est une source importante et complète de données contribuant à mieux comprendre le comportement économique des immigrants. Il s'agit du seul ensemble annuel de données canadiennes permettant d'étudier les caractéristiques des immigrants au Canada au moment de leur admission ainsi que leur situation économique et leur mobilité régionale (interprovinciale) sur plus de 30 ans.

La BDIM combine des fichiers de données administratives sur les admissions d'immigrants et sur les permis de résidence non permanente provenant d'Immigration, Réfugiés et Citoyenneté Canada (IRCC) à des fichiers de données fiscales provenant de l'Agence du revenu du Canada (ARC). Ces données sont disponibles pour des immigrants déclarants admis à partir de 1980. Les dossiers de déclarants sont disponibles pour les immigrants déclarants depuis 1982.

Ce rapport vise à discuter des sources de données, des concepts et des variables de la BDIM, du couplage d'enregistrements, du traitement des données, de la diffusion, de l'évaluation des données et des indicateurs de qualité, de la comparabilité avec d'autres ensembles de données relatifs à l'immigration ainsi que des analyses que permet la BDIM.

Mots clés : données administratives, immigration, BDIM, données longitudinales, résidents non permanents, déclarants

Glossaire

Ci-après figure la signification des sigles utilisés dans le présent rapport.

Acronyme français	Acronyme anglais	Définition
ARC	CRA	Agence du revenu du Canada
BDIM	IMDB	Base de données longitudinales sur l'immigration
BMA	AMDB	Base de mortalité amalgamée
DAL	LAD	Banque de données administratives longitudinales
ELIC	LSIC	Enquête longitudinale auprès des immigrants du Canada
ENM	NHS	Enquête nationale auprès des ménages
FCC	LCF	Fichier de contrôle des couplages
FFT1	T1FF	Fichier sur la famille T1
FIR	ILF	Fichier des immigrants reçus
FRN	NRF	Fichier des résidents non permanents
FRN, permis	NRF, permit	Fichier des résidents non permanents, permis
FRPN	PNRF	Fichier intégré des résidents permanents et non permanents
IPC	CPI	Indice des prix à la consommation
IRCC	IRCC	Immigration, Réfugiés et Citoyenneté Canada
RNP	NPR	Résident non permanent
RP	PR	Résident permanent

1 Introduction

La base de données longitudinales sur l'immigration (BDIM) est une source importante et complète de données contribuant à mieux comprendre le comportement économique des immigrants. Il s'agit du seul ensemble annuel de données canadiennes permettant d'étudier les caractéristiques des immigrants au Canada au moment de leur admission ainsi que leur situation économique et leur mobilité régionale (interprovinciale) sur plus de 30 ans.

La BDIM combine des fichiers de données administratives sur les admissions d'immigrants et sur les permis de résidence non permanente provenant d'Immigration, Réfugiés et Citoyenneté Canada¹ (IRCC) à des fichiers de données fiscales provenant de l'Agence du revenu du Canada (ARC). Ces données sont disponibles pour des immigrants déclarants admis à partir de 1980. Les dossiers de déclarants sont disponibles pour les immigrants déclarants depuis 1982.

La BDIM vise à fournir des données détaillées et fiables sur l'efficacité et les répercussions des programmes d'immigration. En tant que base de données sur les immigrants déclarants, la BDIM peut être utilisée pour répondre aussi bien à de vastes questions de recherche qu'à des questions plus précises. Cette base de données fournit également des renseignements sur l'expérience avant l'admission, comme les permis de travail ou d'études. Sa principale force est qu'elle permet d'analyser la situation socioéconomique sur une période suffisamment longue pour pouvoir évaluer les répercussions des caractéristiques de l'immigration à l'admission (y compris la catégorie d'admission, le niveau de scolarité et les connaissances en français ou en anglais) sur les résultats. De plus, les renseignements annuels sur le lieu de résidence permettent d'étudier la migration secondaire (les déménagements ultérieurs des immigrants au Canada).

Depuis son origine, la BDIM est constituée de plusieurs fichiers : un pour chaque année d'imposition depuis 1982, deux contenant les caractéristiques d'immigration au niveau personnel et un fichier sur les permis de résidence non permanente. La BDIM est mise à jour chaque année au moyen de techniques de couplage décrites dans le présent rapport. Chaque année, une nouvelle année d'imposition est ajoutée à la base de données ainsi que des données relatives aux nouvelles admissions et aux permis de résidence non permanente.

Les fichiers de la BDIM sont uniquement accessibles aux chercheurs de Statistique Canada et aux personnes réputées y être employées. Cette précaution veille à ce que les mesures adéquates de respect des renseignements personnels soient prises, afin de protéger les renseignements personnels et la confidentialité. Les informations liées à la BDIM sont à la disposition du public sous forme de tableaux sommaires agrégés annuels que produit Statistique Canada et qui sont publiés sur son site Web. De plus, des chercheurs externes peuvent demander des tableaux et analyses ponctuels; Statistique Canada fournit ces services selon une formule de recouvrement des coûts. Les utilisateurs souhaitant des renseignements ne figurant que sur le Fichier des immigrants reçus (FIR) devraient s'adresser à IRCC, puisque la BDIM est uniquement conçue pour l'analyse des immigrants déclarants.

Le présent rapport vise à discuter des sources de données de la BDIM (section 2), de ses concepts et variables (section 3), du couplage d'enregistrements (section 4), du traitement des données (section 5), de sa diffusion (section 6), de l'évaluation des données et des indicateurs de qualité (section 7), de sa comparabilité avec d'autres ensembles de données relatifs à l'immigration (section 8) ainsi que des analyses que la BDIM permet (section 9).

1. Auparavant Citoyenneté et Immigration Canada (CIC).

2 Sources des données

Plusieurs fichiers sont nécessaires pour créer la BDIM. Ces fichiers, décrits dans la présente section, comprennent des données d'immigration, des fichiers de données fiscales d'immigrants et des fichiers auxiliaires couvrant de nombreuses années.

2.1 Données sur l'immigration

Chaque année, Statistique Canada (StatCan) reçoit d'IRCC les données d'admission relatives aux nouveaux titulaires de permis de résidence permanente et non permanente.

2.1.1 Fichier des immigrants reçus (FIR)

Chaque année, les données d'admission sont ajoutées pour créer le Fichier des immigrants reçus (FIR). Ce fichier contient des renseignements tels que la date d'admission, la date de naissance et la catégorie d'immigration. Le FIR peut être considéré comme le recensement des personnes ayant immigré au Canada à titre de résidents permanents depuis 1980; il présente des renseignements sur leurs caractéristiques lors de leur admission. Ce fichier n'est cependant pas à la disposition des utilisateurs de la BDIM. Cette dernière porte uniquement sur les immigrants ayant produit au moins une déclaration de revenus depuis 1982. Les données d'admission de ces immigrants déclarants sont disponibles au sein du Fichier intégré des résidents permanents et non permanents (FRPN).

Le FIR, qui consigne les données administratives relatives à la résidence permanente, surestime le nombre d'immigrants vivant actuellement au Canada. Cet écart s'explique de deux façons. Tout d'abord, le FIR ne recense pas les personnes ayant quitté le pays. Des immigrants peuvent être en effet repartis du Canada après leur admission. Ensuite, les décès des immigrants admis à partir de 1980 ne sont que partiellement consignés. Des renseignements supplémentaires sur les données relatives à la mortalité sont présentés à la section 7.2.

2.1.2 Fichier des résidents non permanents (FRN)

Des données relatives aux immigrants admis depuis 1980 possédant une expérience acquise avant leur admission et détenant un statut de résident non permanent ont été utilisées pour créer le Fichier des résidents non permanents (FRN). Ce fichier comprend le nombre de permis, le type de permis (de travail ou d'études, par exemple) et la première année de résidence temporaire. Il est mis à jour chaque année en y ajoutant les données annuelles relatives aux nouveaux permis de résidence non permanente.

Une personne peut en effet posséder plusieurs permis, puisqu'une même personne peut bénéficier de nombreux permis à différents moments. Le FRN consigne de tels renseignements au niveau personnel. Outre le FRN, un fichier sur les permis (section 2.1.4) a été créé pour consigner des données au niveau des permis. Le FRN n'est pas accessible aux utilisateurs de données; cependant, des variables pertinentes provenant de ce fichier sont disponibles au sein du Fichier intégré des résidents permanents et non permanents (FRPN) (voir section 2.1.3).

2.1.3 Fichier intégré des résidents permanents et non permanents (FRPN)

Les chercheurs peuvent accéder au Fichier intégré des résidents permanents et non permanents (FRPN), l'acronyme anglais est PNRF), qui combine des données du FIR et du FRN au niveau des personnes. Le FRPN permet aux utilisateurs de suivre l'historique de la migration des immigrants, notamment leur expérience avant leur admission au Canada. Le FRPN contient des données détaillées sur les caractéristiques sociodémographiques des immigrants admis au Canada depuis 1980; cela permet, par exemple, de déterminer si une personne était un résident non permanent avant son admission. Il est important de noter que les enregistrements des résidents non permanents n'étant pas devenus des résidents permanents ne sont pas inclus dans ce fichier. Ce dernier contient en revanche le nombre de permis pour chaque résident non permanent devenu résident permanent ainsi que sa date d'admission. Le FRPN peut également comprendre une date de décès, lorsqu'un lien vers l'enregistrement de décès a été établi (voir la section 7.2.2). Pour de plus amples détails sur le contenu de ce fichier, reportez-vous à la composante relative à l'immigration du dictionnaire de la BDIM, sections 3.3 et 3.4 de ce rapport.

Outre le FRPN, un autre fichier intitulé **PNRF_EXTRA** (nom basé sur l'acronyme anglais) est à la disposition des utilisateurs de données; il comprend des variables supprimées qui présentent une faible valeur analytique ou pour lesquelles aucune métadonnée n'est disponible. La liste complète des variables est présentée dans le Dictionnaire des données sur l'immigration de la BDIM.

2.1.4 Fichier des résidents non permanents, permis (FRN, permis)

Pour compléter le FRPN, un fichier comprenant des renseignements au niveau des permis a été créé pour les résidents non permanents ayant obtenu un permis depuis 1980. Une personne donnée peut faire l'objet de plusieurs entrées dans ce fichier; c'est-à-dire, une entrée par permis délivré. Sont incluses dans ce fichier les dates liées à chaque permis, comme la date d'entrée en vigueur et la durée de validité. Des données relatives aux résidents non permanents devenus résidents permanents sont à la disposition des utilisateurs de la BDIM. Elles peuvent être couplées au FRPN au moyen de l'identificateur personnel unique de la BDIM (IMDB_ID). Pour de plus amples détails sur les variables incluses dans ce fichier, veuillez vous reporter à la composante relative à l'immigration du dictionnaire de la BDIM.

2.2 Fichier sur la famille T1

Les fichiers de données fiscales utilisés pour créer les fichiers **IMDB_T1FF** sont ceux que contient le fichier sur la famille T1² (FFT1). Statistique Canada obtient les fichiers T1 individuels annuels, les fichiers fiscaux T4 et les fichiers de la Prestation fiscale canadienne pour enfants (PFCE)³ auprès de l'ARC et crée le Fichier sur la famille T1 pour l'année en question. Le traitement comprend de nombreuses étapes, allant du codage géographique à la formation de familles (p. ex. lorsqu'un déclarant mentionne une conjointe également déclarante, cette dernière est associée au déclarant initial par un identificateur commun). Les données du FFT1 remontent jusqu'à l'année d'imposition de 1982. Grâce à l'expérience acquise au cours des nombreuses années de traitement du FFT1, des règles de vérification ont été créées pour réduire le nombre d'incohérences au sein de la base de données et améliorer la qualité des données.

La disponibilité des variables fiscales dépend des renseignements recueillis pour une année donnée. Le FFT1 produit chaque année pour la BDIM comprend les revenus individuels et familiaux ainsi que les variables de composition de la famille, telles que le nombre d'enfants et l'identificateur de conjoint. La BDIM contient les fichiers IMDB_T1FF à compter de 1982 pour les immigrants déclarants. Le processus de création de ces fichiers est décrit à la section 5.1. Pour de plus amples détails sur les variables incluses dans les fichiers IMDB_T1FF, veuillez vous reporter à la composante relative à l'immigration du dictionnaire de la BDIM.

2.3 Fichiers auxiliaires

Pour créer la BDIM, des fichiers auxiliaires sont nécessaires afin de faciliter le couplage des enregistrements et ajouter des variables à la base de données. Ces fichiers auxiliaires ne sont pas accessibles aux utilisateurs de la BDIM.

Le **Fichier de contrôle des couplages (FCC)** relie les données d'immigration aux données fiscales, permettant ainsi la création des fichiers IMDB_T1FF. Le FCC est une base de données des identificateurs personnels présentant des renseignements sur les déclarants à partir de 1981. Le FCC est dérivé, entre autres, des renseignements fournis dans les feuillets T1 et de la PFCE.

Des variables relatives à la mortalité figurent également dans la BDIM. Ces variables proviennent de renseignements présents dans la **base de mortalité amalgamée (BMA)** de Statistique Canada. La BMA combine des enregistrements provenant de fichiers de statistique de l'état civil et de données fiscales afin de produire un ensemble de données sur la mortalité.

2. Pour obtenir des renseignements sur le plus récent FFT1, consultez le [site Web](#) de Statistique Canada.

3. Le 1^{er} juillet 2016, la Prestation fiscale canadienne pour enfants a été remplacée par l'Allocation canadienne pour enfants.

3 Concepts et variables

3.1 Statut d'immigrant au Canada

La BDIM fournit des données sur un sous-ensemble de la population des immigrants, tel qu'il est décrit à la section 2. Trouvez ci-après les définitions des termes « immigrants » et « résidents non permanents » selon Statistique Canada.

Le terme « **immigrant** » désigne des personnes qui sont, ou ont déjà été, des immigrants reçus ou des résidents permanents; c'est-à-dire auxquelles les autorités de l'immigration ont accordé le droit de résider au Canada en permanence. Les immigrants ayant obtenu la citoyenneté canadienne par naturalisation sont également inclus dans cette catégorie.

Les « **résidents non permanents** » ne sont pas considérés comme des immigrants, même s'il s'agit d'une population sur laquelle porte également la BDIM, comme le décrit la section 2. Le terme « **résident non permanent** » désigne une personne originaire d'un autre pays, titulaire d'un permis de travail ou d'études ou qui revendique le statut de réfugié, ainsi que tout membre de sa famille partageant le même permis et vivant avec elle au Canada⁴. Ces personnes sont autorisées à demeurer au Canada pour la période indiquée sur leur permis.

3.1.1 Aperçu de l'immigration au Canada

Un article de *Mégatendances canadienne*, *150 ans d'immigration au Canada*⁵, paru en 2016, résume la fluctuation des niveaux d'immigration et des pays d'origine au cours du dernier siècle. La migration au Canada a été continue depuis la fondation du pays. Plus de 17 millions d'immigrants se sont installés au Canada depuis 1867. Le nombre d'immigrants reçus a augmenté, passant d'environ 200 000 dans les années 1990 à plus de 250 000 au début des années 2010. La proportion des Canadiens nés à l'étranger a augmenté, passant de 14,7 % en 1951 (2,06 millions de personnes) à 20,6 % en 2011 (6,78 millions de personnes).

En vertu de l'article 95 de la *Loi constitutionnelle de 1867*, l'immigration est un domaine de pouvoir partagé entre le gouvernement fédéral et les gouvernements provinciaux. D'autres lignes directrices, définies dans la *Loi sur l'immigration et la protection des réfugiés* de 2002 (LIPR)⁶, précisent les objectifs et la direction stratégique de la politique d'immigration du gouvernement du Canada, que régit en partie Immigration, Réfugiés et Citoyenneté Canada (IRCC). Avant 2002, la *Loi sur l'immigration de 1976* servait de principale loi en matière d'immigration au Canada.

Conformément à la LIPR, le gouvernement du Canada est chargé « de stipuler les exigences en matière d'admissibilité, de fixer les niveaux nationaux d'immigration, de définir les catégories d'immigration, de prendre des décisions concernant les demandes d'asile présentées au Canada, de réunifier les familles et d'établir les critères d'admissibilité pour les programmes d'établissement⁶ » pour toutes les provinces et les territoires, à l'exception du Québec. La province du Québec est entièrement responsable de ses niveaux, programmes et politiques d'immigration en vertu de l'*Accord Canada-Québec relatif à l'immigration et à l'admission temporaire des aubains*. Le gouvernement fédéral continue, cependant, à sélectionner et à traiter les immigrants que parrainent les familles, les personnes protégées au Canada et les personnes revendiquant le statut de réfugié au Québec⁶.

Les **résidents permanents** sont des personnes ayant été autorisées à vivre au Canada de façon permanente, qui ont le droit d'y travailler et d'y étudier, mais qui ne sont pas devenues des citoyens canadiens⁴. En vertu de la LIPR, il existe trois catégories générales d'immigrants : les immigrants économiques, le regroupement familial et les réfugiés.

Les résidents permanents peuvent devenir citoyens du Canada lorsqu'ils répondent à certaines exigences. La première est une exigence de résidence, selon laquelle le résident permanent doit avoir été présent au Canada pendant une période définie. Les résidents permanents doivent également être âgés de plus de 18 ans; dans le cas de mineurs, la demande doit être effectuée simultanément (concomitante) à celle de l'un ou des deux parents

4. Définitions approuvées comme norme recommandée de Statistique Canada le 21 mars 2016. <http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=139756&CVD=139756&CLV=0&MLV=1&D=1>

5. Voir Statistique Canada 2016 dans la biographie.

6. *Loi sur l'immigration et la protection des réfugiés* (L.C. 2001, ch. 27) <http://laws-lois.justice.gc.ca/fra/lois/l-2.5/index.html>

ou après (non concomitante) que l'un des parents ou les deux sont devenus citoyens canadiens. Les résidents permanents doivent avoir honoré leurs obligations de déclaration de revenus auprès du Canada. Les résidents permanents âgés de 14 à 64 ans doivent en outre présenter une preuve de maîtrise d'au moins l'une des langues officielles du Canada et avoir passé l'examen de citoyenneté (IRCC, 2016; gouvernement du Canada, 2016).

La LIPR stipule que tous les ressortissants étrangers, à l'exception des résidents permanents, qui entrent au Canada doivent posséder un visa de résidence temporaire. Les visas de résidence temporaire sont délivrés aux travailleurs et étudiants « de manière à optimiser leur apport au développement économique, social et culturel du Canada tout en protégeant la santé des Canadiens et en assurant leur sécurité » (IRCC, 2015). Les résidents non permanents peuvent demander la résidence permanente dans le cadre de divers programmes et peuvent bénéficier d'un avantage par rapport aux demandeurs de l'étranger, s'ils ont étudié au Canada ou y ont acquis une expérience professionnelle.

Quant aux **personnes qui revendiquent le statut de réfugié**, « la Section de la protection des réfugiés (SPR) est la section de la Commission de l'immigration et du statut de réfugié du Canada (CISR) qui instruit les demandes d'asile présentées au Canada et qui décide de les accepter ou non » (CISR, 2015). Dans le cadre de la BDIM, ces demandeurs sont classés comme résidents non permanents possédant un permis de personne revendiquant le statut de réfugié.

3.2 Population cible et période de couverture

La BDIM est une base de données d'immigrants admis au Canada depuis 1980, pour lesquels au moins un enregistrement FFT1 est disponible depuis 1982. Cette base de données fournit également des renseignements sur les permis des immigrants qui étaient résidents non permanents avant leur admission à titre de résidents permanents.

La BDIM rassemble, par couplage d'enregistrements hiérarchique déterministe (voir la section 4), les données administratives d'Immigration, Réfugiés et Citoyenneté Canada (IRCC) et les fichiers fiscaux du FFT1. Les données d'immigration incluses dans la BDIM proviennent du Fichier des immigrants reçus (FIR) qui contient les enregistrements des immigrants admis au Canada depuis 1980 ainsi que les données provenant du Fichier des résidents non permanents (FRN), qui contient les enregistrements des résidents non permanents ayant obtenu un permis de résidence temporaire depuis 1980.

3.3 Variables relatives à l'admission

Les immigrants sont admis au Canada dans le cadre de différents programmes ayant tous des objectifs précis. Ces programmes sont des indicateurs des conditions sous lesquelles les immigrants sont admis au pays et du type de soutien reçu lors de leur établissement. Par conséquent, les analyses visant à orienter les politiques publiques devraient normalement en tenir compte. Afin de répondre à des questions de recherche variées, la BDIM comporte plusieurs variables relatives à l'admission des immigrants, toutes dérivées de deux concepts principaux, soit la catégorie d'admission et le type de demandeur.

3.3.1 Catégorie d'admission

La **catégorie d'admission** désigne le nom du programme ou du groupe de programmes d'immigration en vertu duquel un immigrant a obtenu pour la première fois depuis 1980 le droit de vivre au Canada en permanence par les autorités de l'immigration. Au fil des ans, les immigrants ont été admis au pays dans le cadre de plusieurs dizaines de programmes distincts. Afin de faciliter l'utilisation des données sur le sujet, la BDIM met à la disposition de ses utilisateurs différentes variables à l'intérieur desquelles les programmes ayant des buts similaires sont agrégés. Le plus haut niveau d'agrégation repose sur les trois objectifs principaux de la politique d'immigration du Canada, soit contribuer au développement économique du pays, réunir les familles et protéger les réfugiés.

3.3.1.1 Immigrants économiques

L'admission d'immigrants économiques vise à atteindre le premier des objectifs susmentionnés de la politique d'immigration, c'est-à-dire celui de contribuer à l'économie canadienne. Les immigrants économiques sont répartis dans trois groupes de programmes principaux, soit les programmes des travailleurs, les programmes des gens d'affaires et les programmes de candidats des provinces et territoires.

Les **programmes des travailleurs** comprennent les immigrants qui ont été sélectionnés pour leur capacité à prendre part au marché du travail. Ceux-ci sont évalués en fonction de leurs compétences et de leur expérience professionnelle et sont répartis en quatre grandes catégories :

1. Les travailleurs qualifiés sont sélectionnés pour leurs compétences et leur expérience de travail dans des postes de gestion, des postes professionnels, des emplois techniques ou des métiers spécialisés.
2. Les travailleurs exerçant un métier spécialisé sont sélectionnés plus particulièrement pour leurs compétences et leur expérience de travail dans le cadre d'un métier spécialisé admissible. Ils se distinguent des travailleurs qualifiés par l'exigence de détenir une offre d'emploi valide d'un employeur canadien ou un certificat de compétence délivré par un organisme provincial ou territorial canadien.
3. Les immigrants admis dans la catégorie de l'expérience canadienne se distinguent des deux premiers groupes par l'exigence de posséder de l'expérience de travail canadienne. Celle-ci doit avoir été acquise dans des postes de gestion, des postes professionnels, des emplois techniques ou des métiers spécialisés.
4. Les aides familiaux et les personnes soignantes obtiennent leur statut de résident permanent après avoir offert, au Canada et pour une période déterminée, des soins à domicile à des enfants ou à des personnes ayant des besoins spéciaux, telles que des personnes âgées ou des personnes souffrant d'un handicap ou d'une maladie chronique.

Les immigrants économiques admis au Canada dans le cadre des **programmes des gens d'affaires** sont quant à eux répartis en trois grandes catégories :

1. Les entrepreneurs sont sélectionnés pour leurs compétences et leur capacité à soit posséder et gérer une entreprise, soit mettre sur pied une entreprise admissible au Canada. Certains possèdent un avoir net minimal alors que d'autres doivent avoir obtenu l'appui d'une organisation désignée pour leur idée d'entreprise.
2. Les investisseurs obtiennent leur statut de résident permanent à la condition d'investir une somme importante au Canada. Leur placement est attribué aux provinces et aux territoires participants afin de stimuler le développement économique et la création d'emplois.
3. Les travailleurs autonomes obtiennent leur statut de résident permanent à la condition d'être en mesure — et d'avoir l'intention — de créer leur propre emploi au Canada et de contribuer de façon appréciable à l'économie canadienne. Cette grande catégorie comprend aussi les personnes qui ont l'intention de contribuer de façon appréciable à la vie culturelle ou sportive du pays (p. ex. en tant qu'artiste, acteur, écrivain ou athlète professionnel).

Le dernier groupe de programmes principal dans le cadre duquel les immigrants économiques sont admis est celui des **programmes de candidats des provinces et territoires**. Comme l'indique le nom du groupe de programmes, les immigrants sont sélectionnés par une province ou un territoire pour contribuer à l'économie locale en répondant à des besoins précis en matière de main-d'œuvre. Ils sont évalués en fonction de critères de sélection relatifs à leur scolarité, à leur expérience de travail et à leurs compétences particulières. Chaque province ou territoire participant possède ses propres critères pour sélectionner son groupe d'intérêt, par exemple les étudiants, les gens d'affaires, les travailleurs spécialisés ou les travailleurs semi-spécialisés.

3.3.1.2 Immigrants parrainés par la famille

L'admission d'immigrants parrainés par la famille vise la réunification des familles et permet aux citoyens canadiens et aux résidents permanents de parrainer des personnes qui leur sont apparentées. Les immigrants admis dans le cadre de tels programmes peuvent recevoir leur statut de résident permanent en raison de leur lien soit comme conjoint, partenaire, parent, grand-parent ou enfant.

Sous certaines conditions, un immigrant admis dans le cadre de tels programmes peut aussi être parrainé en raison d'un autre lien de parenté. C'est le cas par exemple des jeunes frères, sœurs, neveux, nièces et petits-enfants orphelins. Un citoyen canadien ou un résident permanent vivant au Canada peut aussi parrainer quelqu'un sur la base d'un autre lien si personne ne lui est apparenté par les liens décrits plus haut.

À cela s'ajoutent aussi les immigrants qui ont été parrainés par un citoyen canadien ou par un résident permanent vivant au Canada et qui n'auraient normalement pu se qualifier dans le cadre d'aucun programme, mais qui ont reçu, à titre exceptionnel, le statut de résident permanent pour des considérations d'ordre humanitaire.

3.3.1.3 Réfugiés

Le troisième et dernier objectif principal de la politique d'immigration du Canada est celui de protéger les réfugiés, c'est-à-dire les personnes qui ont une crainte fondée de retourner dans leur pays d'origine. Cette catégorie comprend les personnes qui craignent avec raison d'être persécutées pour des motifs liés à leur race, à leur religion, à leur nationalité, à leur appartenance à un groupe social particulier ou à leurs opinions politiques (les réfugiés au sens de la Convention de Genève), de même que les personnes qui ont subi des conséquences graves et personnelles en raison d'une guerre civile, d'un conflit armé ou d'une violation massive des droits de la personne. Certains réfugiés étaient au Canada lorsqu'ils ont demandé l'asile pour eux-mêmes et pour les membres de leur famille avec eux au Canada ou à l'extérieur du pays. D'autres étaient à l'étranger et ont été recommandés par le Bureau du Haut Commissariat des Nations Unies pour les réfugiés (HCR) ou par un autre organisme de recommandation aux fins d'établissement au Canada. Les immigrants qui ont été recommandés reçoivent un soutien pour leur rétablissement de la part de sources gouvernementales, d'organisations, d'individus ou de groupes d'individus du secteur privé, ou encore un soutien provenant à la fois du gouvernement du Canada et d'intervenants du secteur privé.

3.3.1.4 Autres immigrants

En plus des trois objectifs principaux susmentionnés, la politique d'immigration du Canada attribue aux autorités de l'immigration un certain pouvoir discrétionnaire pour octroyer le statut de résident permanent dans le cadre d'un programme qui ne vise ni les immigrants économiques, ni les immigrants parrainés par la famille, ni les réfugiés. Ce programme comprend notamment les immigrants qui ont reçu, à titre exceptionnel, le statut de résident permanent pour des considérations d'ordre humanitaire ou des raisons d'intérêt public.

3.3.2 Type de demandeur

En plus de la catégorie d'admission, la BDIM fournit aux utilisateurs de données des renseignements sur le type de demandeur. Ce concept indique si l'immigrant est identifié comme demandeur principal, comme conjoint ou comme personne à charge dans la demande de visa de résident permanent.

En règle générale, les renseignements sur le type de demandeur sont utilisés aux fins d'analyse dans le cadre de l'étude sur les immigrants économiques. Comme les demandeurs principaux admis en vertu de ces programmes sont sélectionnés pour leur capacité à contribuer à l'économie canadienne, il est pertinent de les distinguer de leurs conjoints et de leurs personnes à charge qui, eux, n'ont pas été évalués en fonction de cette même capacité. En isolant les demandeurs principaux des autres demandeurs, l'efficacité de ces programmes peut alors être étudiée de façon plus directe.

Par contre, en ce qui concerne la réunification des familles et la protection des réfugiés, le but de la politique d'immigration demeure le même pour toutes les personnes qui présentent une même demande, et ce, sans égard au type de demandeur. Dans le cas des immigrants admis en vertu de ces deux objectifs, le concept du « type de demandeur » revêt donc plutôt une valeur administrative.

La valeur administrative du concept est particulièrement marquée pour les immigrants ayant le statut de demandeur principal, lequel ne dépend pas systématiquement du lien juridique entre les demandeurs de résidence permanente. Par exemple, dans le cas de la catégorie d'admission des conjoints et des partenaires parrainés, le conjoint est répertorié comme demandeur principal, alors que le type de demandeur « conjoint » n'apparaît pas sur la demande de résidence. De plus, dans le cas de la catégorie d'admission des enfants parrainés, le statut de demandeur principal sera désigné de façon nominative pour l'un des enfants, alors que les autres enfants seront désignés comme étant les personnes à charge. Enfin, dans certaines circonstances, la

demande de résidence permanente se déroule en deux temps : à partir du Canada pour le demandeur principal, puis à partir de l'étranger pour les autres membres de la famille. Ce type de processus existe pour les aides familiaux résidants et pour les personnes protégées au Canada. Dans ces circonstances, l'un des membres de la famille faisant la demande à partir de l'étranger obtiendra le statut de demandeur principal, et ce, même s'il est le conjoint d'un immigrant approuvé antérieurement suite à une demande faite au Canada.

3.3.3 Changements au fil du temps

La BDIM comprend des renseignements sur les immigrants admis au pays depuis plus d'une trentaine d'années, au cours desquelles plusieurs changements relatifs aux politiques et aux programmes d'immigration sont survenus. De nouveaux programmes ont été créés, d'autres ont été abolis, alors que pour d'autres, les critères de sélection et d'admissibilité ont été modifiés. Ainsi, une personne admise en tant que travailleur qualifié en 1980 n'a pas nécessairement été évaluée selon les mêmes critères que ses homologues admis dans les années 2000. Même si tous les efforts ont été déployés pour tenter de créer les agrégations de programmes les plus similaires possible, l'utilisateur de la BDIM devrait garder en tête ces différences lorsqu'il tire des conclusions au sujet de différentes cohortes d'établissement.

Le changement le plus marquant survenu dans la période couverte par la BDIM est sans doute l'entrée en vigueur de la *Loi sur l'immigration et la protection des réfugiés* de 2002, qui a remplacé la *Loi sur l'immigration* de 1976. Même si ces deux lois visent les trois mêmes groupes principaux (immigrants économiques, immigrants parrainés par la famille et réfugiés), l'administration des programmes dans le cadre de celles-ci a subi plusieurs modifications. De plus, les programmes d'admission ont aussi été modifiés en fonction des besoins sociodémographiques et des priorités établies par les gouvernements successifs à l'intérieur même de ces deux cadres législatifs. Pour ces raisons, il est fortement conseillé aux utilisateurs de données s'intéressant à des programmes spécifiques ou à plusieurs cohortes d'établissement de se renseigner sur les changements apportés aux politiques ou aux programmes applicables à leur sujet d'étude.

Il est aussi important de souligner que quelques années peuvent s'écouler avant que les effets d'un changement administratif puissent être observés à l'intérieur de la base de données. Par exemple, lorsqu'un nouveau programme est créé, il peut s'écouler plusieurs mois — voire plusieurs années — entre le moment où celui-ci entre en vigueur (moment où il est possible de poser sa candidature) et le moment où les premiers immigrants sont admis au pays dans le cadre de celui-ci. Un scénario similaire existe pour les programmes abolis. Il peut s'écouler un certain temps entre le moment où toutes les demandes sont évaluées et où toutes les personnes admissibles sont reçues au pays et celui où les programmes abolis disparaissent complètement des statistiques sur les admissions annuelles.

3.3.4 Variables de catégorie d'admission du FRPN

Plusieurs catégories d'admission existent au sein du FRPN. Elles sont décrites dans la composante relative à l'immigration du dictionnaire de la BDIM. Cette section fournit de plus amples détails sur certaines de ces variables.

La variable la plus détaillée est celle des catégories d'immigration **IMMIGRATION_CATEGORY**, qui comprend plus de 100 catégories ayant existé de 1980 jusqu'à la BDIM actuelle. Des agrégats des renseignements disponibles dans la variable **IMMIGRATION_CATEGORY** figurent dans les variables dérivées de groupement de catégories d'immigration, **IMM_CATEGORY_STC_ROLLUP1** et **IMM_CATEGORY_STC_ROLLUP2**, qui comprennent un moins grand nombre de catégories.

La variable agrégée 1 de groupement de catégories **IMM_CATEGORY_STC_ROLLUP1** est une catégorisation conforme à la publication officielle *Faits et chiffres* d'Immigration, Réfugiés et Citoyenneté Canada. Elle ne précise, cependant, pas que certains immigrants ont été admis dans le cadre des **programmes d'élimination de l'arriéré et d'examen administratif**. Ces programmes visaient à accélérer le traitement des dossiers d'immigration à la fin des années 1980, afin de faire face aux crises géopolitiques se déroulant à l'étranger et touchant la capacité de résidents temporaires à retourner dans leur pays (p. ex. manifestations de la place Tian'anmen, dissolution de l'URSS et de la Yougoslavie). Du fait de l'absence de distinction de ces catégories, ces personnes, dont le dossier a été traité rapidement selon des critères différents, ne peuvent être comparées aux autres immigrants dont les dossiers figurent dans les mêmes catégories. Pour recenser les immigrants admis dans le cadre de ces

programmes, les utilisateurs doivent se reporter aux variables d'élimination de l'arriéré BACKLOG_CLEARANCE_IND et d'examen administratif ADMINISTRATIVE_REVIEW_IND (disponibles dans le fichier PNRF_extra, pour les années d'admission antérieures à 2014).

L'utilisateur peut également utiliser les renseignements agrégés sur l'immigration de la variable 2 de groupement de catégories d'immigration IMM_CATEGORY_STC_ROLLUP2. Cette variable vise à fournir des informations cohérentes indépendamment des divers changements de politique et de réglementation (p. ex. *Loi sur l'immigration et la protection des réfugiés* de 2002 et *Loi sur l'immigration de 1976*) et pour maintenir au cours du temps des programmes particuliers relatifs à l'immigration (p. ex. travailleurs qualifiés). Cette variable offre des renseignements détaillés sur les initiatives d'élimination de l'arriéré et d'examen administratif. Des détails relatifs au groupement pour les variables dérivées figurent dans le dictionnaire des données sur l'immigration de la BDIM.

Un autre point à prendre en compte quant aux variables de catégories d'admission est leur relation avec le **type de demandeur** (variable de statut dans la famille FAMILY_STATUS du FRPN). En règle générale, les demandeurs principaux sont ceux étant évalués à l'aide des critères d'admission dans chacune des catégories, alors que le conjoint ou la conjointe et les personnes à charge qui les accompagnent sont automatiquement admis avec le demandeur principal (même si les aptitudes linguistiques du conjoint ou de la conjointe peuvent également être évaluées dans les demandes d'immigration de catégorie économique). Au sein des variables de groupement, certaines catégories d'admission précisent explicitement s'il s'agit 1) d'un demandeur principal ou 2) d'un conjoint ou d'une personne à charge, alors que pour d'autres catégories (p. ex. regroupement familial), il est nécessaire de se référer à la variable FAMILY_STATUS pour déterminer le statut de la personne en tant que demandeur principal ou conjoint/personnes à charge.

Deux catégories font exception à ce qui précède : la catégorie personnes à charge de type aide familial résidant et la catégorie réfugiés. Lorsqu'elles sont comparées à la catégorie de groupement de statut dans la famille FAMILY_STATUS_ROLLUP, ces variables font référence à des demandeurs principaux ainsi qu'à des personnes à charge. Cela peut se produire lorsque le demandeur principal se trouve déjà dans le pays et qu'une personne à sa charge soumet une demande distincte de résidence permanente de l'étranger. Puisque chaque demande distincte doit présenter un demandeur principal, même nominal, l'une des personnes à charge (généralement le conjoint) est considérée comme le « demandeur principal » de la demande des personnes à charge. Il n'existe, cependant, aucune différence de traitement entre les demandeurs principaux, les conjoints et les personnes à charge dans ces deux catégories d'admission.

3.4 Variables d'intérêt

La **BDIM** est une base de données complète fournissant aux chercheurs une myriade de variables permettant d'étudier les résultats relatifs aux caractéristiques d'immigration ainsi que diverses répercussions à long terme. Le nombre de variables au sein des fichiers de données fiscales dépasse 600 (la moitié environ au niveau individuel et l'autre moitié au niveau familial de l'agrégation). Le Fichier intégré des résidents permanents et non permanents (FRPN) contient plus de 50 variables. Même si les définitions exactes de ces variables figurent dans la composante relative à l'immigration du dictionnaire de la BDIM, certains des concepts plus nuancés sont décrits dans le présent rapport. Les sections suivantes portent sur les variables géographiques, temporelles, de scolarité et de profession envisagée, afin d'en présenter plus précisément la signification et l'utilité. De plus amples détails sur les variables de revenu figurent dans la composante relative aux impôts du dictionnaire de la BDIM.

Les variables du FRPN (PNRF) font référence aux caractéristiques des immigrants lors de leur admission ou à la réception d'un document de résidence temporaire, alors que les variables des fichiers de données fiscales font référence aux caractéristiques au cours de l'année d'imposition. Alors que certaines variables sont disponibles dans les deux fichiers, les variables relatives aux impôts changent au cours du temps. L'âge, par exemple, figure dans les deux fichiers et devrait changer chaque année dans le fichier de données fiscales. L'état matrimonial des immigrants (MARITAL_STATUS) et la province de destination (DESTINATION_PROVINCE) lors de la demande de résidence permanente peuvent également différer de l'état matrimonial (MTSCO) et de la province de résidence (PRCO) au moment de la déclaration de revenus. Pour les variables qui ne devraient pas changer au cours du temps, il convient d'utiliser le FRPN par souci de cohérence.

3.4.1 Variables géographiques

La BDIM permet d'étudier la mobilité et la rétention des immigrants déclarants au Canada. Il n'est cependant pas possible de capturer complètement les émigrations, car les immigrants ou déclarants ne sont pas tenus de déclarer qu'ils quittent ou comptent quitter le pays. Les fichiers FRPN et de données fiscales présentent diverses mesures du lieu géographique qui permet aux chercheurs d'établir une destination envisagée à l'admission et les lieux de résidences suivants des immigrants. Dans le FRPN, la **destination envisagée** est mesurée à l'échelle provinciale, de la région métropolitaine de recensement, de la division de recensement et de la subdivision de recensement. Ces variables proviennent de la destination autodéclarée à l'admission figurant sur la demande d'immigration. Contrairement aux variables géographiques du FFT1, les variables du fichier d'établissement sont uniquement disponibles pour les géographies définies dans le dernier recensement disponible; cela signifie qu'elles ne reflètent que les limites de recensement les plus récentes.

L'autre renseignement géographique figurant sur le fichier d'établissement est la **province de nomination**, disponible pour les candidats des provinces. La province indiquée est celle selon les critères de laquelle le demandeur a été admis; elle ne correspond cependant pas nécessairement à la variable de province de destination.

Sous **régions géographiques d'origine**, les variables de pays du fichier d'établissement indiquent le pays de naissance de la personne, son pays de citoyenneté et sa résidence la plus récente au moment de son admission. Il convient de préciser que ces régions géographiques peuvent ne pas demeurer comparables au fil du temps, car les frontières nationales peuvent changer d'une année à l'autre. On peut penser par exemple à la dissolution de l'URSS, de la Yougoslavie et de la Tchécoslovaquie, à l'union du Sikkim et de l'Inde et à celle du Vietnam et du Nord, ou encore à la création du Soudan du Sud.

Certaines personnes indiquent dans le fichier d'établissement le Canada comme pays de naissance. Normalement, les personnes nées au Canada reçoivent cette citoyenneté à la naissance et n'ont pas à faire de demande de résidence permanente. Les personnes indiquant être nées au Canada sur le fichier d'établissement sont probablement des enfants de diplomates étrangers nés lorsqu'ils résidaient au Canada et qui ont choisi ultérieurement de demander le statut de résident permanent.

Plusieurs variables géographiques des ensembles de données du FFT1 font référence à des notions de régions géographiques légèrement différentes de celles du fichier d'établissement. Le renseignement géographique le plus détaillé du FFT1 est disponible au niveau du secteur de recensement; il est dérivé du **code postal de l'adresse postale (PSCO_)**. Le code postal indique généralement l'adresse de résidence lors de la déclaration de revenus au printemps de l'année suivante. L'adresse postale peut également faire référence à une entreprise, comme un cabinet comptable ou d'avocats, et n'est pas nécessairement l'adresse actuelle de la personne en question. La **province de résidence** au 31 décembre de l'année d'imposition (PRCO_) peut ne pas être identique à la province de l'adresse postale. Cette distinction est importante, car la variable dérivée de la région géographique de recensement peut ne pas correspondre à la province de résidence au 31 décembre (PRCO_); elle devrait cependant correspondre au **code de province (PR_)**. La variable PRHO_ indique une autre valeur d'adresse postale et existe uniquement à partir de 2008. De plus, alors que la variable de **code de province d'imposition (TXPCO_)** est, par définition, identique à la province de résidence au 31 décembre (PRCO_), ce code de province d'imposition (TXPCO_) est moins fiable (car des problèmes connus de qualité de données concernent cette variable, pour laquelle les valeurs manquantes et celles relatives à Terre-Neuve-et-Labrador reçoivent le code « 0 »).

Utiliser les variables des fichiers de données fiscales pour étudier la mobilité géographique des immigrants exige de tenir attentivement compte du moment en question avant de déduire un déménagement et l'emplacement de travail. Un guide du chercheur portant sur l'étude de la mobilité et de la présence est inclus en annexe D.

3.4.2 Variables temporelles

L'« année d'admission » et l'« année d'imposition » sont des variables temporelles souvent utilisées pour produire des tableaux et effectuer des analyses à l'aide de la BDIM. L'année d'admission est l'année où l'immigrant a reçu le statut de résident permanent, alors que l'année d'imposition est l'année pour laquelle la déclaration de revenus est produite.

Il est recommandé de considérer la première année de revenu comme l'« année d'admission + 1 », puisqu'il s'agit de la première année complète où la personne se trouvera dans le pays. Il convient d'interpréter avec précaution les revenus déclarés pour l'année d'admission. Tout d'abord, environ 50 % de chaque cohorte d'admission produisent une première déclaration de revenus pour l'année d'admission (proportion calculée en fonction des déclarants à partir des immigrants couplés). Ensuite, les revenus déclarés l'année d'admission peuvent ne pas représenter une année de revenus complète. Une personne admise en octobre 2010 déclarera, au printemps 2011, des revenus pour seulement 3 mois, alors qu'une personne admise en janvier 2010 déclarera des revenus pour 12 mois.

Des déclarations de revenus sont également possibles après le décès d'une personne, lorsqu'une personne apparentée au décédé produit une déclaration de revenus en son nom. La variable de type de famille (FCMP_) du FFT1 doit être utilisée dans ce cas. Veuillez consulter le dictionnaire des données pour plus de précisions.

Le tableau 1 illustre, par exemple, les scénarios possibles et décrit les enregistrements à inclure dans une étude visant à évaluer les résultats socioéconomiques de la cohorte des immigrants de 1995 à 2000, cinq ans après leur admission. Pour qu'un enregistrement soit inclus, l'immigrant doit avoir été admis entre 1995 et 2000 et avoir produit une déclaration de revenus cinq ans après son admission. Cette analyse inclurait les enregistrements de la BDIM suivants : IM19952 et IM19963.

Tableau 1
Exemple de définition d'une cohorte d'intérêt

IMDB_ID	Année d'admission	Années d'imposition disponibles	Inclus dans le champ d'enquête
IM19801	1980	1982 à 2013	Non; admis avant 1995
IM19952	1995	1988 à 2011	Oui
IM19963	1996	1996 à 2013	Oui
IM19974	1997	2010 à 2013	Oui, mais pas de fichiers de données fiscales disponibles cinq ans après l'admission
IM20095	2009	2011 à 2013	Non; admis après 2000

Note : Cet exemple se fonde sur des données fictives.

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration.

L'un des inconvénients d'utiliser les données fiscales administratives est leur manque de précision quant aux dates. Outre l'année pour laquelle les revenus sont déclarés, aucune variable de temps n'existe dans le FFT1. Cela présente des difficultés pour étudier les durées d'emploi et de chômage, le moment des déménagements et les dates de mariage. Cette lacune ne permet pas non plus de distinguer facilement les travailleurs autonomes qui sont également des employés saisonniers de ceux qui reçoivent un revenu simultanément de deux sources. Malgré ces limites, il est tout de même possible de prendre des décisions en matière de date en tenant compte des points suivants.

Puisque les revenus de l'année précédente sont généralement déclarés au printemps de l'année civile suivante, une incertitude peut exister quant à l'année précise d'un changement d'adresse. Une personne A, par exemple, pourrait déclarer ses revenus de 2011 avec une adresse postale à Toronto, puis déclarer ses revenus de 2012 avec une adresse postale à Ottawa. On peut en déduire que cette personne A vivait à Toronto lors de sa déclaration de revenus de 2011 au printemps 2012 et vivait à Ottawa au printemps 2013. La personne A a pu déménager en 2012, après avoir produit sa déclaration de revenus pour l'année précédente, ou au cours des premiers mois de 2013, avant de produire sa déclaration de revenus pour 2012. Si cette personne A a changé de province, la variable de la province de résidence au 31 décembre (PRCO_) pourrait être utile pour déduire l'année du déménagement, puisqu'elle porte sur la province de résidence de cette personne en décembre de l'année d'imposition. L'adresse postale ne correspond pas nécessairement au lieu de résidence.

3.4.3 Variables relatives à la scolarité

Plusieurs variables du fichier d'établissement, comme les années d'études et le niveau de scolarité, permettent de mesurer le niveau de scolarité à l'admission. La première prend la forme d'une réponse écrite à la question « Combien d'années d'études avez-vous terminées? » La deuxième est un choix parmi des options correspondant au niveau de scolarité le plus élevé atteint. La variable de « niveau de scolarité » dérivée combine les renseignements provenant de ces deux questions.

Un problème de qualité des données a été relevé pour ces variables relatives à la scolarité depuis 2011, car une proportion significative des personnes pour cette période n'ont pas précisé leur niveau de scolarité ou leur nombre d'années d'études; le code « 0 » (« aucun ») a été attribué à ces variables EDUCATION_QUALIFICATIONS et YEAR_OF_SCHOOLING plutôt que « donnée manquante ». Ce problème est devenu courant depuis 2011. En 2011, jusqu'à 35 % des immigrants ont indiqué ne pas avoir suivi d'études, par rapport à environ 10 % dans les années 1990. Les variables relatives à la scolarité depuis 2011 ont fait l'objet d'une imputation pour résoudre ce problème. Si vous souhaitez davantage de détails sur l'imputation, consultez la section 7.3.

3.4.4 Variables de profession envisagée

IRCC recueille la profession envisagée à partir de l'enregistrement d'admission et lui attribue une classification correspondant aux codes de Classification nationale des professions (CNP) du fichier d'établissement. Il s'agit de la classification la plus large en matière de compétences; les codes CNP à cinq chiffres étant les plus précis (voir le dictionnaire en annexe pour des définitions complètes).

Même si la profession envisagée est également considérée comme un remplacement valable de la profession de la personne dans son pays d'origine, des précautions sont recommandées. Pour pouvoir indiquer une profession envisagée, le demandeur doit démontrer qu'il possède les qualifications requises en matière d'études ainsi qu'au moins une année d'expérience dans le domaine. IRCC considère, par conséquent, ce renseignement comme une mesure prudente de la variable envisagée après l'admission, puisque les exigences qui l'accompagnent sont relativement strictes. Les étudiants venant d'obtenir un diplôme de génie, par exemple, ne peuvent pas indiquer « ingénieur » comme profession envisagée (du fait de leur manque d'expérience professionnelle) et sont plutôt classés comme « étudiants ». D'autres variables, comme **l'intention sur le marché du travail** (LM_INTENTION) et le **niveau de compétence** (SKILL_LEVEL), peuvent servir à obtenir des renseignements sur le domaine professionnel de la personne dans son pays d'origine. Il convient, en outre, de souligner que le champ de profession envisagée est obligatoire pour les immigrants demandeurs principaux au sein des catégories économiques. Pour les autres immigrants, ce renseignement est facultatif. Il peut ainsi ne pas être une mesure aussi fiable de la profession que ces personnes envisagent.

3.4.5 Autres variables de la BDIM

Seules les variables nécessitant une explication détaillée et présentant d'éventuelles difficultés importantes pour les analystes ont été incluses à la section 3.4. Pour de plus amples détails sur les variables incluses dans la BDIM, veuillez vous reporter aux dictionnaires de la BDIM. La composante relative aux impôts décrit les variables incluses dans les fichiers IMDB_T1FF, alors que la composante relative à l'immigration décrit les variables incluses dans les autres fichiers. Ces dictionnaires sont à la disposition des utilisateurs des données, mais il est également possible de les obtenir en en faisant la demande par écrit à Statistique Canada à l'adresse STATCAN.infostats-infostats.STATCAN@canada.ca.

Les utilisateurs des données de la BDIM doivent garder à l'esprit que les données des fichiers relatifs à l'immigration et de données fiscales sont recueillies à divers moments et que, dans certains cas, les caractéristiques des personnes évoluent au cours du temps. Par conséquent, l'état matrimonial et la composition de la famille d'une personne peuvent changer au cours de l'année et différer entre le FRPN et le FFT1. Les variables à utiliser dans l'analyse dépendent de l'objet de l'étude.

4 Couplage d'enregistrements

Comme le décrit le présent document, la BDIM est le produit de nombreux couplages d'enregistrements. Elle a été créée en vue de fournir des renseignements statistiques anonymes. Cette section fournit un aperçu des méthodes de couplage d'enregistrements utilisées pour créer la BDIM. Pour de plus amples détails sur le traitement des données relatif au couplage d'enregistrements, reportez-vous à la section 5.

Le couplage d'enregistrements est le processus de mise en correspondance d'enregistrements entre ou au sein de bases de données. Cette approche est couramment utilisée pour combler des lacunes statistiques et créer un ensemble de données présentant de vastes applications (Rotermann et coll., 2015).

La méthode hiérarchique déterministe est utilisée pour coupler des enregistrements des fichiers FIR et FRN au Fichier de contrôle des couplages (FCC), base de données d'identificateurs personnels (voir la section 2 pour la description de ces fichiers). Cette méthode consiste à coupler des enregistrements entre plusieurs fichiers (ou au sein d'un fichier donné) au moyen de variables communes (Dusetzina et coll., 2014). Au fur et à mesure des vagues de correspondance, les critères de couplage deviennent de moins en moins stricts. Le FCC n'est pas accessible aux chercheurs; il ne sert qu'à produire la BDIM.

Les utilisateurs d'un ensemble de données créé à partir d'un couplage d'enregistrements doivent être conscients que des erreurs de couplage sont possibles. Un couplage d'enregistrements peut produire quatre résultats : de réelles correspondances correctement classées comme des correspondances, de réelles correspondances classées à tort comme des non-correspondances, de réelles non-correspondances classées à tort comme des correspondances ou de réelles non-correspondances correctement classées comme des non-correspondances (Winkler, 2009). Ainsi que le montre l'exemple du tableau 2, dans lequel les enregistrements du fichier 1 sont couplés aux enregistrements du fichier 2, le résultat du couplage d'enregistrements entre deux variables sera soit une correspondance soit une non-correspondance. Un bon couplage d'enregistrements maximisera la proportion des réelles correspondances correctement classées comme des correspondances et la proportion des réelles non-correspondances correctement classées comme des non-correspondances, et minimisera les autres résultats éventuels du couplage d'enregistrements.

Tableau 2
Exemple de résultats de couplage d'enregistrements

	Enregistrement	Fichier 2			Type de résultat
		A	B	D	
Fichier 1	A	Correspondance	Non-correspondance	Non-correspondance	Réelle correspondance
	C	Non-correspondance	Correspondance	Non-correspondance	Fausse correspondance
	D	Non-correspondance	Non-correspondance	Non-correspondance	Fausse non-correspondance
	E	Non-correspondance	Non-correspondance	Non-correspondance	Réelle non-correspondance

Source : Statistique Canada, exemple de résultats de couplage d'enregistrements.

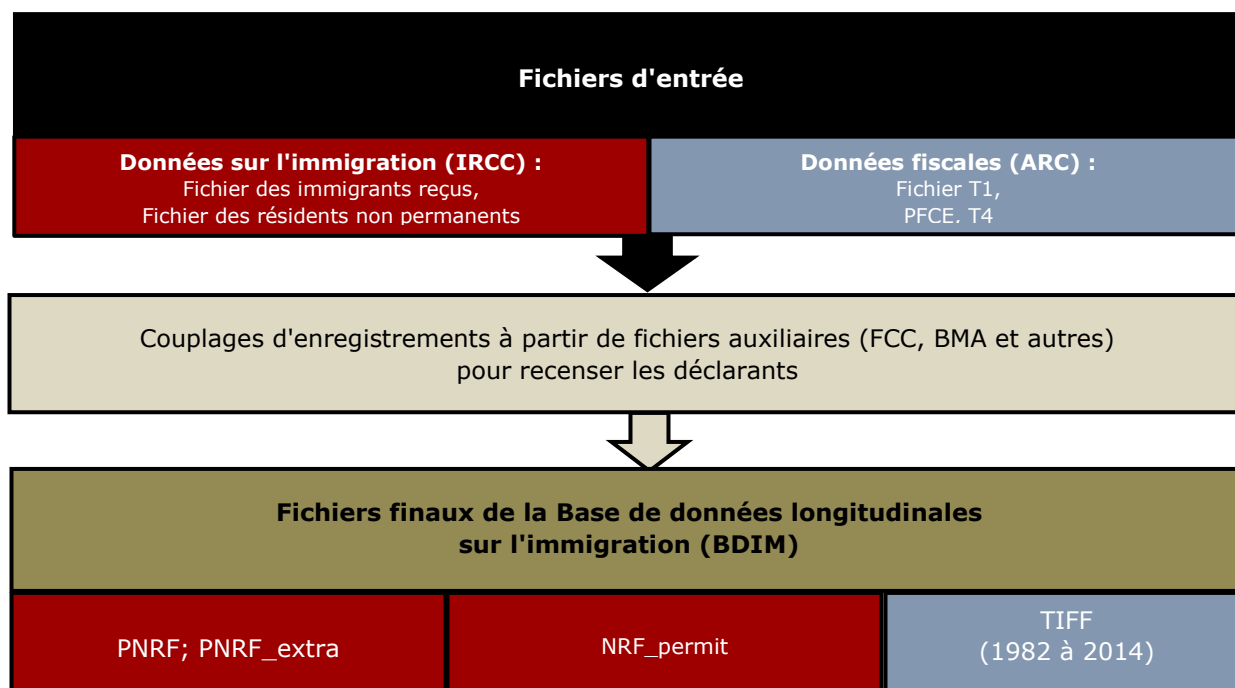
Les résultats d'un couplage d'enregistrements déterministe dépendent de la qualité des variables de couplage. Des noms mal orthographiés, par exemple, ou des coquilles dans la date de naissance peuvent empêcher des correspondances ou créer des erreurs de correspondance. Une non-correspondance ne signifie pas nécessairement que la personne n'a pas produit de déclaration de revenus. Les taux de couplage d'enregistrements pour la BDIM la plus récente sont fournis à la section 7.2.1.

5 Traitement des données

5.1 Traitement

Plusieurs organismes gouvernementaux participent à la création et au traitement de la BDIM. De la collecte initiale des données à la diffusion en passant par le traitement, la coopération de ces organismes est nécessaire pour fournir la qualité élevée des données qu'attendent les utilisateurs des données de Statistique Canada. À chaque étape de la séquence de traitement, des vérifications attentives, manuelles et automatiques, de la qualité des données sont effectuées et des boucles de rétroaction servent à corriger toute erreur détectée à la source. La section suivante décrit brièvement le processus annuel de traitement qui permet de mettre la BDIM à jour.

Figure 1
Résumé du déroulement du processus de création de la BDIM



Note : Voir le glossaire pour connaître les définitions des acronymes.

Source : Statistique Canada, description du déroulement du processus de création de la BDIM.

Comme le montre la figure 1 (les fichiers finaux sont nommés avec l'acronyme anglais dans la BDIM), Statistique Canada reçoit d'abord les données de T1 de l'Agence du revenu du Canada (ARC), dans un fichier intitulé « Fichier maître des particuliers » et d'autres fichiers de données fiscales. Les fichiers de données fiscales servent alors à créer le FFT1, au sein duquel les personnes sont couplées à leurs conjoints et enfants par un identificateur commun et des variables géographiques sont créées. Statistique Canada effectue des vérifications manuelles de la qualité et compare les estimations du FFT1 à d'autres sources de données, comme le recensement (pour les années de recensement) et l'Enquête sur la dynamique du travail et du revenu, ainsi que des statistiques de revenu annuel que produit l'ARC⁷.

En matière d'immigration, IRCC fournit les données relatives aux immigrants admis et aux résidents non permanents qui sont utilisées pour produire la BDIM. Ces données servent à créer le Fichier des immigrants reçus (FIR) et le Fichier des résidents non permanents (FRN). On considère que les fichiers FIR et FRN sont des recensements complets des permis de résidence permanente et temporaire qu'a délivrés IRCC depuis 1980.

7. Pour de plus amples détails sur le traitement et la qualité des données du FFT1, consultez le site http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&ld=240717.

Les enregistrements des fichiers FIR et FRN sont couplés au Fichier de contrôle des couplages (FCC) par des variables de couplage principales, telles que le nom, la date de naissance et le sexe. Des couplages hiérarchiques déterministes sont créés au cours de plusieurs vagues, selon des critères de sévérité décroissant à chaque vague. Cependant, même les vagues les moins strictes sont de nature prudente, afin d'éviter les fausses correspondances (coupler deux enregistrements ne faisant pas référence à la même personne), tout en générant le moins de fausses non-correspondances possibles (couplages manqués). Le fichier de couplage ainsi obtenu sert à l'étape de traitement finale, qui consiste à coupler les données relatives à l'immigration à celles du FFT1.

Outre l'ajout des renseignements de l'année d'imposition la plus récente, une vérification complète des années précédentes a lieu afin d'ajouter des données fiscales pour toute nouvelle personne couplée. Des améliorations annuelles de la méthodologie ainsi que la disponibilité de nouveaux renseignements sur le couplage accroissent les taux de couplage d'une année à l'autre et, ainsi, augmentent le nombre d'enregistrements couplés même pour les années précédentes. Cela peut signifier que les enregistrements de déclaration de revenus d'un immigrant admis ou d'un résident non permanent peuvent ne pas être couplés dans la BDIM une certaine année, mais que ses déclarations de revenus suivantes peuvent être couplées pour une année ultérieure. À mesure que la méthodologie s'améliore, la vérification rétroactive peut permettre de coupler également toutes ses déclarations de revenus antérieures, si elles figurent dans le FFT1. C'est ainsi qu'après le traitement des données fiscales les plus récentes, des personnes admises et ayant produit une déclaration de revenus plusieurs années auparavant peuvent encore être ajoutées à la BDIM. Lorsque plusieurs admissions depuis 1980 existent pour une personne, on conserve les données de sa première admission.

Bien que les déclarations de revenus d'une année donnée soient généralement produites au printemps de l'année suivante (p. ex. déclaration des revenus de 2013 en 2014), des exceptions existent. Parfois, une personne produit sa déclaration de revenus plus tard dans l'année et n'est alors pas incluse dans le traitement des T1 qu'effectue Statistique Canada pour cette année-là. Lorsque ce fichier est transmis aux fins de traitement pour la BDIM, ces déclarants retardataires ne sont pas inclus et ne le seront pas non plus lors du traitement de l'année suivante, puisque le FFT1 n'est pas mis à jour. De la même manière, les personnes qui produisent leur déclaration de revenus pour des années antérieures ne sont pas ajoutées à la BDIM pour ces années, puisque les FFT1 des années précédentes ne sont pas mis à jour. Dans ces cas, la première déclaration de ces personnes effectuée au moment attendu s'affichera comme leur première année dans la base de données.

Pour l'instant, on exécute une série de programmes pour évaluer la qualité des données et les taux de couplage, afin d'éviter tout doublon et de signaler les valeurs aberrantes. Une fois la base de données couplée, elle est jugée complète et prête à être mise à disposition.

Finalement, la base de données consiste en des fichiers SAS, un fichier de données fiscales par année depuis 1982 (IMDB_T1FF_&year_OUTLIERS) et des fichiers sur les données relatives à l'immigration (PNRF_&year, PNRF_EXTRA_&year et NRF_PERMIT). Tous ces fichiers sont décrits dans la section 2. L'identificateur personnel unique de la BDIM (IMDB_ID) sert à relier tous ces fichiers (voir l'annexe D.1 pour des conseils de programmation).

5.2 Couplage du Fichier des résidents non permanents (FRN)

Le Fichier des résidents non permanents (FRN) qu'IRCC fournit couvre les enregistrements de permis de résidence temporaire délivrés à partir de 1980. Il fournit des données démographiques sur les résidents non permanents ainsi que des renseignements détaillés sur leur permis, tels que le type de permis et leur plage de validité.

Le FRN contient des millions d'observations. Elles comprennent cependant des doublons, selon lesquels une même personne peut compter plusieurs identifiants différents. Le problème provient principalement d'enregistrements à la fin des années 1980 pour lesquels le numéro d'identification de la personne initiale a été perdu. Ces enregistrements ont été supprimés en couplant le FRN à lui-même. Cela a permis de déterminer qu'environ 220 000 enregistrements (environ 400 000 observations) étaient des doublons. Dans les cas où les deux enregistrements de résidents non permanents possédaient chacun un enregistrement d'admission, le lien de doublon a été supprimé (ce qui s'applique à moins de 1 000 enregistrements), puisqu'on considère que le fichier d'établissement contient des identifiants uniques. Après nettoyage, il ne reste que des résidents non permanents distincts.

Les deux fichiers relatifs à l'immigration (FIR et FRN) contiennent des données démographiques. Toutefois, ces données démographiques ne sont pas toujours cohérentes. Cela se produit lorsque plusieurs sources sont disponibles ou en cas de conflit. Il a été décidé que les données du FIR relatives au Fichier intégré des résidents permanents et non permanents (FRPN) seraient conservées, du fait de problèmes de qualité concernant les données du FNR au cours de ses premières années.

5.3 Variables dérivées incluses dans le FFT1

Une fois les couplages d'enregistrements effectués, on ajoute au FFT1 des variables relatives à l'immigration pour les immigrants et les résidents temporaires.

On crée un indicateur de statut d'immigrant pour les déclarants en fonction du couplage avec les données sur l'immigration d'IRCC. Pour déterminer si une déclaration de revenus correspond à un déclarant qui était un résident non permanent, on utilise l'indicateur **TR_IND** (1 : oui; 0 : non).

Des variables dérivées indiquant et décrivant des familles sont également créées. Dans chaque FFT1 annuel, il est possible d'estimer le nombre d'immigrants dans une famille admise depuis 1980 (variable **IMM80F&year**). Il est également possible de déterminer si l'immigrant est un conjoint (pour l'année d'imposition donnée) et si ce conjoint est immigrant ou résident non permanent (variable **SP_IDI&year**). Les utilisateurs de données peuvent recenser les immigrants d'une même famille, pour chaque année d'imposition, à l'aide de la variable de **numéro d'identification de la famille** (FIN_). Cette variable présente la même valeur pour tous les membres d'une famille, c'est-à-dire l'identifiant IMDB_ID du membre de la famille le plus âgé admis depuis 1980. La qualité de ces variables dépend de la qualité du couplage d'enregistrements et des fichiers FFT1, puisque seules les personnes couplées sont comptabilisées (voir la section 7.2.5).

Les variables ayant le préfixe **TNK** représentent le nombre d'enfants déclarés d'un certain âge dans les familles d'immigrants et de résidents non permanents (voir la composante relative aux impôts du dictionnaire des données pour de plus amples détails). Le terme « enfant » désigne toute personne célibataire vivant avec un ou deux parents; un enfant peut être de tout âge. Au tableau 3, par exemple, la famille d'immigrants dont le numéro est IM19801 a deux enfants de 1 an en 2011 (TNK01I2011), alors que la famille IM19873 a trois enfants en 2011 (TNKIDI2011), un de 0 an (TNK00I2011), un de 1 an (TNK01I2011) et un de plus de 18 ans (TNK19I2011). L'immigrant IM20105 n'a pas d'enfant en 2011.

Tableau 3
Exemple de variables relatives au nombre d'enfants dans une famille

IMDB_ID	TNK00I2011	TNK01I2011	TNKxxI2011	TNK19I2011	TNKIDI2011
	nombre				
IM19801	0	2	0	0	2
IM19802	0	1	0	0	1
IM19873	1	1	0	1	3
IM19994	0	0	0	1	1
IM20105	0	0	0	0	0

Note : Toutes les variables ne figurent pas dans ce tableau. Cet exemple se fonde sur des données fictives.

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration.

Une autre variable ajoutée au FFT1 est l'identificateur de valeur aberrante **OUTLIER_IND** (1 : aberrante; 0 : non). Il s'agit d'un indicateur ajouté pour détecter les enregistrements présentant des résultats extrêmes (voir la section 5.5 pour de plus amples détails) et devant être supprimés de tout tableau ou calcul. Les enregistrements marqués comme aberrants présentent des revenus extrêmes qui ajouteraient un biais aux résultats de l'analyse.

5.4 Variables dérivées incluses dans le FRPN

Lorsque le FRPN est produit, certaines variables relatives aux tendances de déclaration de revenus sont dérivées et incluses au fichier. La variable **FIRST_TAX_YEAR** indique la première année pour laquelle les données fiscales sont disponibles pour une personne donnée, alors que la variable **LAST_TAX_YEAR** indique la dernière année pour laquelle les données fiscales sont disponibles. Il convient de noter que des données fiscales ne sont pas nécessairement disponibles chaque année entre la première et la dernière année d'imposition. Par exemple, un

cas où les variables `First_tax_year=1982` et `Last_tax_year=2012` existent ne signifie pas nécessairement que le déclarant a produit une déclaration de revenus de manière continue; les données fiscales pour 2006 peuvent manquer, par exemple.

La variable de déclarant avant l'admission **PREFILER_IND** sert à signaler les immigrants pour lesquels des données existent dans le FFT1 avant leur année d'admission. La plupart de ces cas ont été couplés à des enregistrements de résidents non permanents, comme on s'y attend (voir la section 7.2.4 pour de plus amples détails).

5.5 Détection des valeurs aberrantes

Après la création des fichiers `IMDB_T1FF`, on procède à la détection des valeurs aberrantes dans tous les fichiers de données fiscales, afin de relever tous les enregistrements aberrants. Un enregistrement est jugé aberrant lorsque l'on détermine qu'il contient une valeur ou plus de revenus extrême par rapport aux autres enregistrements. Les critères utilisés pour déterminer les valeurs aberrantes sont confidentiels. La variable **OUTLIER_IND** est créée pour recenser les enregistrements présentant des valeurs extrêmes.

L'indicateur de valeur aberrante `OUTLIER_IND` figure dans les fichiers de données fiscales, mais non dans le FRPN. Cet indicateur peut être attribué à l'enregistrement d'une personne donnée pour une année particulière, sans que cela ne soit nécessairement le cas pour toutes les années pour lesquelles la personne a produit une déclaration de revenus. Toutes les valeurs aberrantes doivent être supprimées de toute analyse. Comme le montre le tableau 4, pour la personne `IM19802`, l'indicateur de valeur aberrante a seulement été attribué à l'enregistrement de 1983, alors qu'aucun des fichiers de données fiscales de la personne `IM19801` ne présente cet indicateur. Aucun indicateur de valeur aberrante n'apparaît en 2012 pour la personne `IM19994`, car aucune donnée fiscale n'est disponible pour cette personne cette année-là.

Tableau 4
Exemple relatif à l'indicateur de valeur aberrante

IMDB_ID	OUTLIER_IND1982	OUTLIER_IND1983	OUTLIER_INDyyyy	OUTLIER_IND2012	OUTLIER_IND2014
	nombre				
IM19801	0	0	0	0	0
IM19802	0	1	0	0	0
IM19873	1	1	1	0	...
IM19994	0	0	0	...	0

... n'ayant pas lieu de figurer

Note : Toutes les variables ne figurent pas dans ce tableau. Cet exemple se fonde sur des données fictives.

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration.

Les valeurs aberrantes sont supprimées des totalisations et de toute analyse. La BDIM exclut (les très rares) revenus importants, car ils influeraient sur les moyennes et donneraient une impression incorrecte de la situation en matière de revenus pour certains types d'immigrants. Prenons l'exemple fictif d'un revenu moyen de 40 000 \$ pour les Canadiens tchèques pour une année donnée; l'année suivante, cette moyenne atteint subitement 500 000 \$ parce qu'un joueur de hockey tchèque a été admis. Cela inclurait un biais dans la situation « réelle » des Canadiens tchèques en matière de revenus. Ainsi, le revenu « non représentatif » du joueur de hockey tchèque serait retiré des calculs. De plus, cet exemple présente une composante de confidentialité. Si une telle hausse du revenu était observée, il serait possible de déduire le revenu du joueur de hockey; ce qui constituerait une atteinte à la vie privée. Dans certains produits de la BDIM, le revenu médian, qui résiste mieux à l'influence d'importantes valeurs individuelles, est également fourni.

Lorsqu'une personne produit des tableaux ou analyse des données, les enregistrements jugés être des valeurs aberrantes pour une année donnée doivent être retirés des calculs pour l'année en question pour les raisons susmentionnées. Pour de plus amples détails, voir l'annexe D.6.

6 Diffusion

Une fois le couplage terminé, les fichiers de données (y compris un ensemble de données synthétique, voir la section 6.3) sont enregistrés sur les serveurs de Statistique Canada, afin que les utilisateurs de données puissent créer des tableaux et des modèles à partir des résultats. Statistique Canada diffuse les résultats au moyen de produits tabulaires et analytiques, tout en respectant strictement les normes de confidentialité des données. Les utilisateurs souhaitant des renseignements ne figurant que sur le Fichier des immigrants reçus (FIR) devraient s'adresser à IRCC, puisque la BDIM ne présente que des données sur les immigrants déclarants.

6.1 Produits analytiques

À Statistique Canada, les principaux produits analytiques sont les **tableaux du Système canadien d'information socioéconomique (CANSIM)** qui présentent le revenu des immigrants selon diverses caractéristiques individuelles et régions. Ces tableaux figurent sur le site Web de Statistique Canada sous le numéro de série [054](#). Les mesures de revenu (moyen et médian) disponibles dans les tableaux CANSIM sont le revenu d'emploi, l'assurance-emploi, les revenus d'investissements, les revenus de travailleurs autonomes et l'assistance sociale (pour de plus amples détails sur la façon dont ces mesures sont dérivées, voir l'annexe D.8).

En 2014, quatre tableaux CANSIM fondés sur les données de la BDIM ont été produits à l'échelle nationale :

Tableau 054-0001 : Revenu des immigrants, selon le sexe, le groupe d'âge lors de l'admission, la catégorie d'admission d'immigrant, le nombre d'années écoulées depuis l'admission et l'année d'admission, dollars constants de 2014

Tableau 054-0002 : Revenu des immigrants, selon la région du monde, le sexe, la catégorie d'admission d'immigrant, le niveau de scolarité, la connaissance des langues officielles et l'année d'admission pour l'année d'imposition de 2014

Tableau 054-0003 : Nombre d'immigrants déclarants, selon la province d'admission, la province de résidence, le sexe, le groupe d'âge lors de l'admission, la catégorie d'admission d'immigrant et l'année d'admission pour les années d'imposition 1996 à 2014

Tableau 054-0018 : Revenu des immigrants, selon le sexe, le groupe d'âge lors de l'admission, la catégorie d'admission d'immigrant, le statut familial, la période d'immigration et l'année d'imposition, dollars constants de 2014

Pour tous ces tableaux, à l'exception du tableau 054-0003, il existe une série de tableaux à l'échelle provinciale, comme l'indique le tableau 5. Les taux de rétention peuvent être calculés à partir du tableau CANSIM 054-003 (voir l'annexe D.4).

À l'échelle provinciale, toutes les provinces atlantiques sont incluses dans un seul tableau; il n'existe pas de tableau individuel à l'échelle territoriale. La valeur de la province se fonde sur la province de résidence au 31 décembre de l'année d'imposition (variable PRCO).

Tableau 5

Liste de tableaux CANSIM relatifs à la Base de données longitudinales sur l'immigration

Tableau national	Tableau provincial
054-0001	0004, 0006, 0008, 0010, 0012, 0014, 0016
054-0002	0005, 0007, 0009, 0011, 0013, 0015, 0017
054-0003	Aucun
054-0018	0019, 0020, 0021, 0022, 0023, 0024, 0025

Source : Statistique Canada, CANSIM.

Plusieurs articles analytiques relatifs à la BDIM ont en outre été publiés au cours des années (voir l'annexe C). Les analystes de Statistique Canada acceptent, en outre, des demandes de données ponctuelles de chercheurs et d'utilisateurs de données. Ils y répondent selon une formule de recouvrement des coûts.

6.2 Demande de fichiers analytiques

Une fois la BDIM publiée, tous les fichiers analytiques décrits dans le présent rapport (FRPN, FRPN_extra, FFT1 et FRPN_permiss) sont également mis à la disposition des chercheurs sur place (les fichiers sont disponibles sous l'acronyme anglais), qui obtiennent un droit d'accès lorsqu'ils reçoivent le statut de personne réputée être employée de Statistique Canada. Tout renseignement permettant l'identification est supprimé de ces microdonnées individuelles (comme une date de naissance exacte, la date d'admission, le numéro d'assurance sociale et le nom). Les chercheurs qui ne peuvent se rendre physiquement au bureau central de Statistique Canada peuvent accéder à un **fichier synthétique**, comprenant les variables et formats de variable, mais contenant des données fictives. Ces chercheurs peuvent utiliser ce fichier synthétique pour créer les programmes souhaités, puis envoyer la syntaxe à Statistique Canada aux fins d'exécution. Ce fichier doit être utilisé conjointement aux dictionnaires les plus à jour, car certaines variables du fichier synthétique ont été supprimées depuis sa création et d'autres ont été ajoutées. Les utilisateurs de la BDIM peuvent demander des totalisations personnalisées à Statistique Canada; ces demandes sont traitées selon une formule de recouvrement des coûts et ces coûts peuvent varier selon la nature et le type de chaque demande.

Avant la parution de tout résultat, Statistique Canada en vérifie la confidentialité. Des exigences de taille de cellule minimale et l'arrondissement limitent les risques d'atteinte à la vie privée.

6.3 Autres programmes statistiques utilisant les données de la BDIM

Les données de la BDIM sont utilisées à diverses fins dans le cadre de nombreux programmes de Statistique Canada. La **Banque de données administratives longitudinales** (DAL) inclut dans son échantillon un échantillon de 20 % des enregistrements de la BDIM. La DAL utilise également les enregistrements de la BDIM pour ajouter des variables propres aux immigrants (comme l'année d'admission) à sa banque de données.

La **Base de données canadienne sur la dynamique employeurs-employés** (BDCDEE) est un ensemble de fichiers de données analytiques longitudinales que Statistique Canada maintient à jour pour fournir des correspondances de données entre employés et employeurs du marché du travail canadien depuis l'an 2000. Les fichiers de la BDCDEE couvrent toutes les personnes dont l'identité peut être établie dans les fichiers T1 et T4, ainsi que les renseignements sur les employeurs ou tout travail autonome auxquels les personnes peuvent être reliées. La BDIM est l'un des fichiers composant la BDCDEE et ce couplage permet aux chercheurs d'effectuer une analyse de la situation sur le marché du travail et des dynamiques de l'emploi relativement à la population immigrante au Canada.

L'**Enquête sociale générale (ESG) de 2013 sur l'identité sociale** (IS) recueille des renseignements détaillés sur les réseaux sociaux, ainsi que sur la participation et l'engagement civiques de la population canadienne. L'ESG de 2013 sur l'identité sociale a été couplée à la BDIM dans le but de sélectionner un échantillon représentatif de la population immigrante, afin de soutenir et d'évaluer les politiques et programmes relatifs aux immigrants. Immigration, Réfugiés et Citoyenneté Canada (IRCC) a en particulier utilisé ce fichier de données couplées pour élaborer un profil descriptif des relations sociales et de l'engagement civique des immigrants selon les catégories d'admission.

DEMOSIM⁸, modèle de microsimulation de Statistique Canada, utilise le fichier BDIM-DAL pour créer des projections de population pour les provinces, territoires, régions métropolitaines de recensement et certains découpages géographiques moins étendus, selon plusieurs caractéristiques. Les **programmes de recensement** ont recours à la base de données pour attester les données sur l'immigration.

6.4 Confidentialité⁹

Statistique Canada s'est engagé à assurer la protection des renseignements personnels. Tous les renseignements personnels créés, détenus ou recueillis par Statistique Canada sont protégés par la [Loi sur la protection des renseignements personnels](#) et par la [Loi sur la statistique](#) dans le cas des répondants à nos enquêtes.

8. <http://www.statcan.gc.ca/fra/microsimulation/demosim/demosim>.

9. Source : <http://www.statcan.gc.ca/fra/reference/privée>.

Du fait de son mandat unique à titre d'agence statistique nationale de collecte de renseignements personnels aux seules fins d'études statistiques et de recherche, Statistique Canada a préparé des [évaluations des facteurs relatifs à la vie privée](#) qui portent sur les problèmes de protection de la vie privée associés aux activités d'enquête.

Statistique Canada a mis en place une évaluation des facteurs relatifs à la vie privée¹⁰ lorsque le Comité des politiques (le haut comité exécutif de l'agence que préside le statisticien en chef) a approuvé les modifications importantes apportées à la base de données longitudinales sur les immigrants. L'objectif de cette évaluation était de déterminer s'il existait des problèmes de confidentialité, de sécurité et de protection de la vie privée liés à ces changements et, le cas échéant, de formuler des recommandations en vue de les résoudre ou de les atténuer.

Cette évaluation a conclu que, grâce aux mesures de protection existantes au sein de Statistique Canada ainsi qu'aux mesures additionnelles mises en place pour la Base de données longitudinales sur l'immigration, le risque de divulgation involontaire était extrêmement faible. L'importance des données sur le plan des politiques publiques l'emporte sur les préoccupations relatives à la vie privée. Les mécanismes de gouvernance en place constituent des mesures de protection contre l'utilisation inappropriée des données. Statistique Canada évalue régulièrement la pertinence de la BDIM au moyen d'examen périodiques qu'effectue le Comité des politiques, s'appesantissant chaque fois sur la pertinence des renseignements par rapport à l'atteinte implicite portée à la vie privée.

Le travail statistique de l'agence comprend des projets de couplage d'enregistrements qui rassemblent des renseignements relatifs aux répondants individuels à des fins de recherche. Il s'agit d'une source reconnue de données statistiques précieuses, mais le couplage doit toujours être d'utilité publique. Afin de faire face à d'éventuelles atteintes à la vie privée dues à ce type de recherche, Statistique Canada respecte non seulement une directive, mais procède en outre à un examen bien défini et exige l'approbation de tous les [couplages d'enregistrements](#).

Pour protéger la vie privée, il est ainsi obligatoire d'arrondir les résultats tabulaires et descriptifs lors de la production de tableaux à partir des données de la BDIM (voir l'annexe D.5).

10. <http://www.statcan.gc.ca/fr/aperçu/efrvp/bdi>.

7 Évaluation des données et indicateurs de qualité

7.1 Sources d'erreurs

Parce que la BDIM est produite à partir de plusieurs couplages d'enregistrements, elle comprend différentes sources d'erreurs, notamment des erreurs de couplage d'enregistrements, de mesure et de couverture. Nous expliquons, dans cette section, les sources d'erreurs et indiquerons l'importance de certaines d'entre elles.

Il convient de noter que, puisqu'il s'agit du recensement des immigrants déclarants admis depuis 1980, aucune pondération n'a été créée au sein de la BDIM. Aucune rectification n'est apportée aux années d'imposition manquantes ni aux erreurs de couplage, aucun échantillonnage n'a lieu et tous les déclarants couplés sont conservés dans l'ensemble de données final. Cependant, le couplage lui-même présente une forme d'erreur d'échantillonnage dans le cas d'échecs de couplage.

7.1.1 Erreurs de couplage d'enregistrements

Les ensembles de données produits à partir des résultats des couplages d'enregistrements peuvent comprendre des erreurs de couplage d'enregistrements. Deux **types d'erreurs** sont ainsi possibles : des faux positifs (fausses correspondances) et des faux négatifs (fausses non-correspondances). Un couplage est un faux positif lorsque deux enregistrements ne concernant pas la même personne sont comptés comme une correspondance. Un couplage est un faux négatif lorsque deux enregistrements concernant la même personne sont comptés comme une non-correspondance. Selon un examen manuel du résultat du couplage d'enregistrements de la BDIM de 2014, le taux de faux positifs était estimé à 0,27 % (Diaz-Papkovich 2016a). Nous ne disposons pas actuellement du taux de faux négatifs.

Le **taux de couplage** fournit une indication de la qualité de la base de données; le taux de couplage est de 89 % (Diaz-Papkovich, 2016b) pour les immigrants admis après 1980. La mise en correspondance fait appel à des variables communes au Fichier de contrôle des couplages (FCC) et aux ensembles de données sur l'immigration (p. ex. prénom et nom, sexe et date de naissance). La plupart des correspondances ont été établies au cours des premières vagues; ce qui a présenté peu de différences de valeur de ces variables ou aucune différence. Cela signifie également que la plupart des correspondances sont de qualité élevée et cela explique le taux peu élevé de faux positifs. Pour de plus amples détails sur les taux de couplage, reportez-vous à la section 7.2.1.

Il est en outre possible de manquer une partie de l'historique d'un immigrant, puisque certains possèdent plusieurs numéros d'assurance sociale (NAS) au cours du temps (NAS temporaire attribué à l'admission au résident non permanent, puis NAS permanent attribué après l'admission). Ces deux NAS sont nécessaires pour constituer l'historique fiscal complet depuis l'admission de la personne au Canada. Le FCC (décrit à la section 2.3) permet de recenser ces deux NAS. Il est possible que, dans quelques cas, des liens entre les NAS échouent ou que des liens erronés soient créés.

7.1.2 Erreurs de mesure

Une **erreur de mesure** est la différence entre la valeur mesurée d'une variable et sa valeur réelle. On peut attribuer ce type d'erreur à plusieurs facteurs, notamment la saisie de données (p. ex. coquilles) et les erreurs des répondants (p. ex. mauvaise interprétation de la question posée). On a tenu compte de ce type d'erreur lors de la création du Fichier intégré des résidents permanents et non permanents (FRPN) pour éviter les renseignements contradictoires pour toute personne. Lorsqu'un enregistrement existe pour une personne dans le FIR et dans le FNR, par exemple, et que les variables sociodémographiques présentent des valeurs incohérentes, les valeurs à l'admission (celles du FIR) sont conservées. Voir les sections 7.2 et 7.5 qui fournissent certains chiffres.

7.1.3 Erreurs de couverture

Les **erreurs de couverture** proviennent d'omissions, de l'inclusion d'unités erronées, d'enregistrements en double et d'erreurs de classification d'enregistrements dans la base de données. Ces erreurs de couverture peuvent provenir d'une couverture inadéquate de la population. Elles peuvent se traduire par des estimations biaisées et les répercussions peuvent varier pour différents sous-groupes de la population. Ces erreurs entraînent souvent un sous-dénombrement. Le **sous-dénombrement** dans le cadre de la BDIM provient en partie de l'exclusion de la base de données des fichiers de données fiscales d'immigrants déclarants. Aucun enregistrement

ne figure dans le fichier IMDB_T1FF pour les immigrants qui ne produisent pas de déclaration de revenus pour une année donnée ou qui le font en retard, même s'ils sont couplés aux données fiscales et font partie de la population d'intérêt. Si, pour une raison quelconque, l'enregistrement d'un immigrant n'est pas inclus dans le Fichier des immigrants reçus (FIR), il ne fera pas partie de la BDIM. Un **surdénombrement** survient lors de l'ajout à la base de données d'enregistrements exclus de la population cible. Un immigrant peut faire l'objet de plusieurs enregistrements dans le FIR du fait de plusieurs admissions non reconnues comme telles, par exemple. Veuillez vous reporter à la section 7.4 et à l'annexe B pour de plus amples détails relatifs à la couverture de la BDIM.

7.2 Exactitude des données

Cette section porte sur l'exactitude des données relatives à l'immigration. Pour de plus amples détails sur l'exactitude du FFT1, veuillez vous reporter au document sur le [FFT1](#) (numéro d'enregistrement 4105).

L'exactitude de la BDIM dépend de la représentativité de la population qui y est incluse. Une étude menée au cours des premières années de la BDIM a conclu que cette dernière « semble être représentative de la population la plus susceptible de produire des déclarations de revenus. Par conséquent, les résultats obtenus à partir de la BDIM ne devraient pas être appliqués à la population immigrante dans son ensemble, mais plutôt à l'univers des immigrants qui produisent des déclarations de revenus » (Carpentier et Pinsonneault, 1994).

Les raisons expliquant les différences entre les déclarants et la population totale des personnes nées à l'étranger sont expliquées dans un article de Badets et Langlois (2000) décrivant les défis d'utilisation de la BDIM :

Les caractéristiques de la population immigrante qui produit des déclarations de revenus diffèrent de celles de l'ensemble de la population née à l'étranger, étant donné que la probabilité ou la nécessité de produire une déclaration de revenus varie selon l'âge, la situation familiale de la personne et d'autres facteurs. On pourrait s'attendre à ce qu'un pourcentage plus élevé d'immigrants de sexe masculin produisent une déclaration de revenus, notamment parce qu'ils ont des taux de participation au marché du travail plus élevés que les femmes. Le niveau de « saisie » des immigrants dans la BDIM est aussi influencé par les changements touchant l'impôt sur le revenu. Par exemple, la mise en place de programmes de crédits d'impôt non remboursables à l'échelle fédérale et provinciale incite les personnes qui n'ont pas de revenus imposables à produire une déclaration de revenus pour être admissibles à certains crédits d'impôt.

Ces commentaires sur la représentativité de la BDIM s'appliquent toujours. Comme le montre la section suivante, les femmes représentent des taux de couplage inférieurs à ceux des hommes, par exemple.

7.2.1 BDIM de 2014 : taux de couplage

Cette section porte sur la BDIM de 2014. Le taux de couplage global pour la BDIM de 2014 était de 89 % et incluait des différences selon le sexe et l'âge. Une paire ne signifie pas qu'une déclaration de revenus est disponible, puisqu'il est possible de coupler des personnes à la charge de déclarants ou des immigrants n'ayant pas encore produit de déclaration de revenus. Parmi les immigrants admis de 1980 à 2014, 82,5 % ont été couplés à au moins un enregistrement du FFT1.

Le tableau 6 présente la proportion de déclarants couplés par tranche d'âge à l'admission et par sexe. Des taux inférieurs sont enregistrés pour les femmes immigrantes par rapport aux immigrants masculins pour toutes les tranches d'âge. Les taux inférieurs pour la tranche d'âge de 0 à 14 ans s'expliquent bien sûr par le fait que ce groupe n'est pas en âge de travailler. Voir l'annexe B pour des taux par sexe, tranche d'âge et cohorte d'admission.

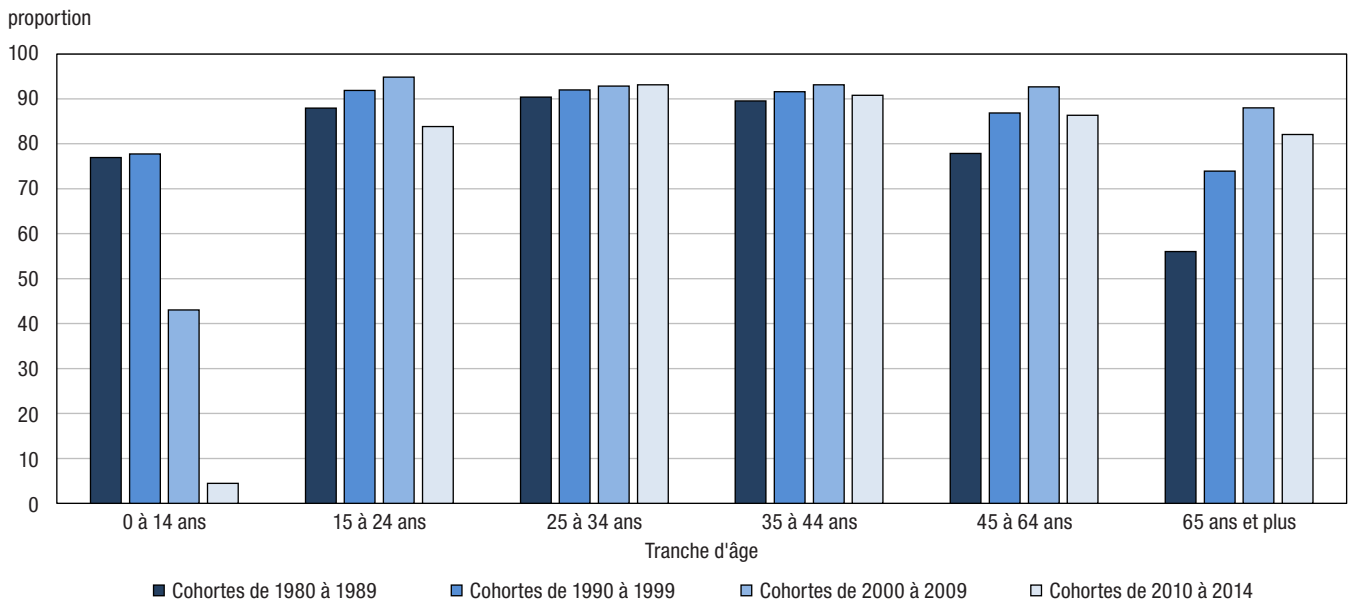
Tableau 6
Proportion de déclarants couplés, par tranche d'âge à l'admission et par sexe

	Âge à l'admission						Total
	0 à 14 ans	15 à 24 ans	25 à 34 ans	35 à 44 ans	45 à 64 ans	65 ans ou plus	
	pourcentage						
Hommes	53,5	91,8	92,8	92,0	87,8	75,3	82,8
Femmes	52,6	89,8	91,7	91,5	85,2	73,6	82,3
Total	53,1	90,7	92,3	91,8	86,4	74,3	82,5

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

À mesure que les immigrants vieillissent, ils commencent à produire des déclarations de revenus et sont inclus dans la BDIM. Le graphique 1 montre que, parmi les immigrants admis à tout âge de 0 à 14 ans, la proportion de déclarants couplés est supérieure pour les immigrants admis avant 2000 que pour ceux admis depuis 2000. Les récents immigrants présentent également des taux de couplage inférieurs. Voir l'annexe B pour un tableau présentant la proportion de déclarants couplés par groupe d'âge à l'admission, sexe et décennie d'admission.

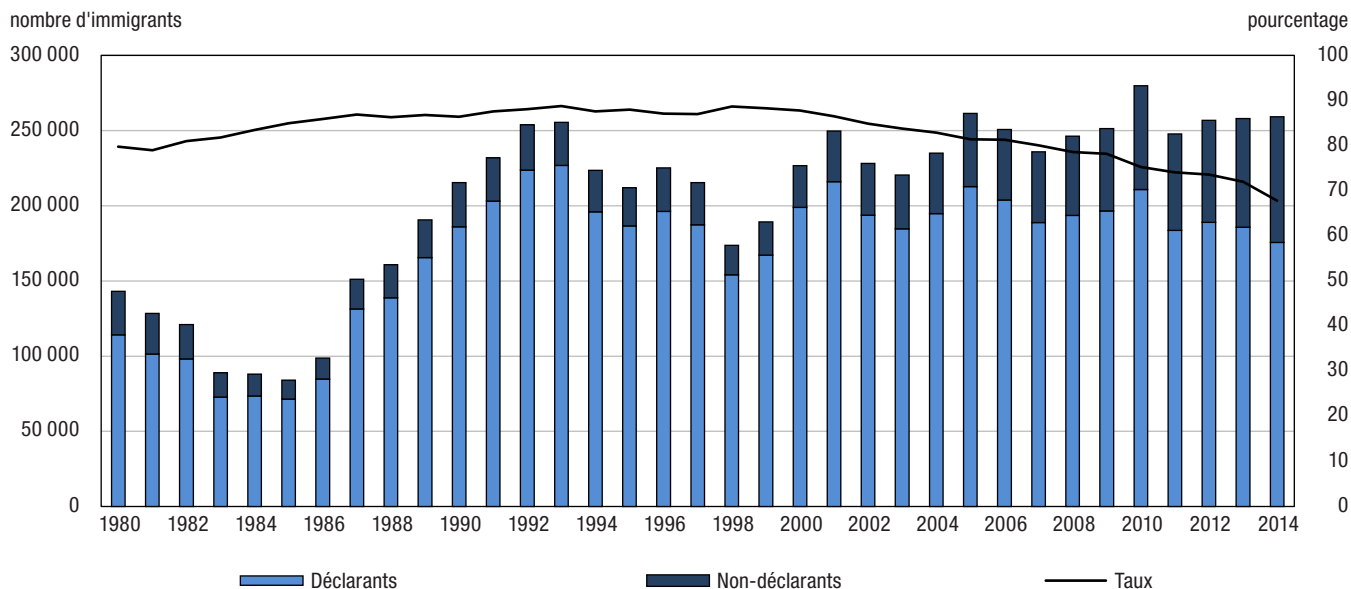
Graphique 1
Proportion de déclarants couplés, par tranche d'âge à l'admission et décennie d'admission



Source : Base de données longitudinales sur l'immigration de 2014.

Le tableau 2 illustre la proportion de déclarants ainsi que le nombre de déclarants et de non-déclarants par année d'admission; la mention « non-déclarant » indique qu'aucun enregistrement dans le FFT1 n'est disponible. Dans la BDIM de 2014, le taux de déclaration varie en fonction de l'année d'admission, de 67,8 % pour les personnes admises en 2014 à 88,8 % pour celles admises en 1993. Le taux de déclaration augmente à mesure du nombre d'années de présence des immigrants au Canada; cela peut expliquer pourquoi les taux de couplage sont supérieurs pour les immigrants admis dans les années 1980 et 1990. Voir l'annexe B, tableaux 15 et 16, pour connaître des chiffres de répartition détaillés par année d'admission.

Graphique 2
Répartition des déclarants et des non-déclarants par année d'admission



Source : Statistique Canada, Base de données longitudinales sur l'immigration, 2014.

7.2.2 Disponibilité de la date de décès

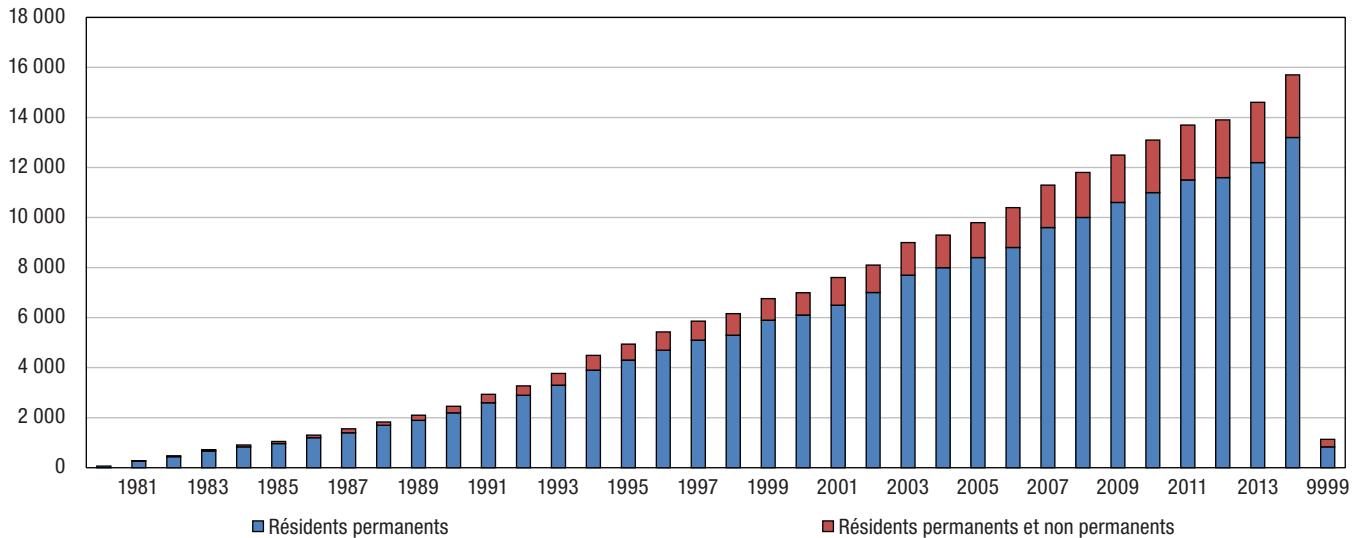
L'année et le mois d'un décès, ainsi que l'indicateur de décès, sont inclus dans le FRPN. Ces variables peuvent être ajoutées à l'aide des renseignements figurant dans la base de mortalité amalgamée (BMA) de Statistique Canada. La BMA combine des enregistrements provenant de fichiers de statistique de l'état civil et de données fiscales afin de produire un ensemble de données sur la mortalité.

Ces renseignements proviennent d'un couplage d'enregistrements. Il est probable qu'une faible proportion de ces couplages soient faux; c'est-à-dire que le couplage d'enregistrements a créé des couplages faux positifs. Tous les couplages présentant une date de décès avant la date d'admission ont été supprimés (moins de 200 enregistrements). Les personnes décédées avant 1981 ou qui n'ont jamais produit de déclaration de revenus sont moins susceptibles de faire l'objet d'un couplage du fait des fichiers utilisés pour les couplages.

Le graphique 3 décrit la tendance générale du nombre de décès par an depuis 1980. Un décès de « résident permanent » désigne le décès de tout immigrant admis depuis 1980 et ne possédant pas d'expérience avant son admission. Un décès de « résident permanent et non permanent » désigne le décès de tout immigrant ayant une expérience avant son admission. La valeur « 9999 » représente les enregistrements d'immigrants décédés pour lesquels l'année de décès n'est pas disponible.

Graphique 3 Résidents permanents et non permanents par année de décès

nombre de décès



Note : La valeur 9999 a été utilisée lorsque la date de décès n'était pas disponible.

Source : Statistique Canada, Base de données longitudinales sur l'immigration, 2014.

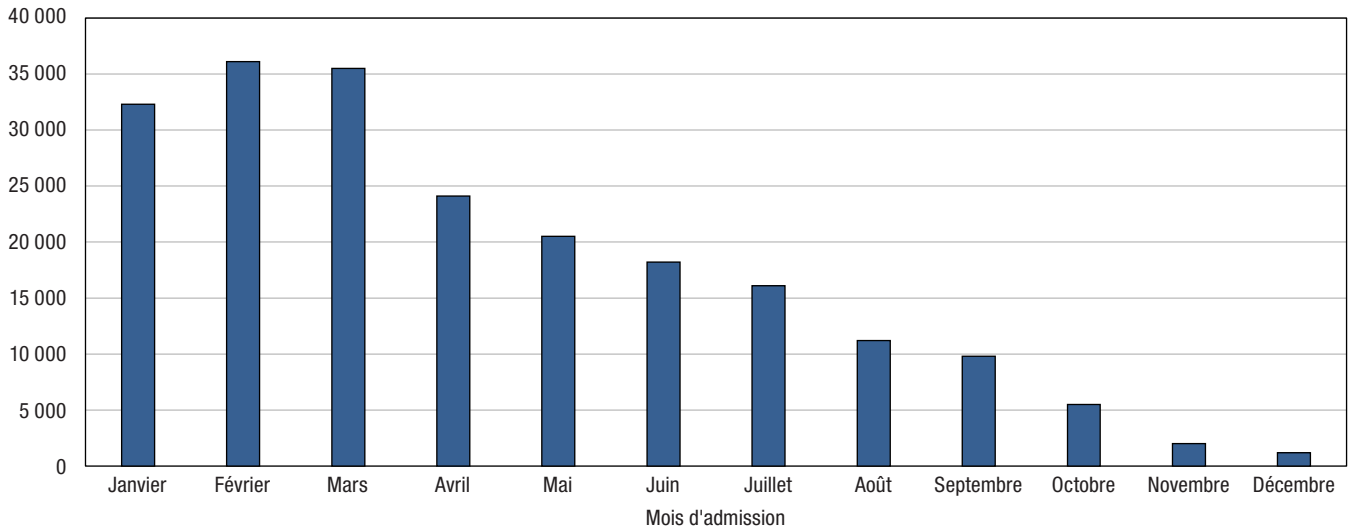
7.2.3 Déclarants avant l'admission par rapport aux enregistrements du Fichier des résidents non permanents (FRN)

Les résultats inclus dans cette section sont tirés d'une étude fondée sur la BDIM de 2013. Les **déclarants avant l'admission** sont des immigrants qui ont produit une déclaration de revenus avant leur année d'admission. Il est parfois déduit que tous les déclarants avant l'admission sont des immigrants qui étaient des résidents non permanents avant leur admission. Cette section explique pourquoi ce n'est pas le cas. Au total, 1,26 million de personnes ont produit une déclaration de revenus avant leur admission officielle depuis 1980; sur ce total, 212 500 ne sont pas couplées à un enregistrement de résident non permanent, contrairement à ce à quoi l'on pourrait s'attendre. Après une enquête approfondie, on a découvert que la plupart des déclarants avant l'admission qui étaient des résidents permanents non couplés à un enregistrement de résident non permanent étaient probablement des immigrants ayant produit une déclaration de revenus lorsque cela n'était pas requis : 96 % de ces déclarants avant l'admission ont uniquement produit une déclaration de revenus pour l'année précédant leur année d'admission et 75 % n'ont déclaré aucun revenu (96 % n'avaient aucun revenu d'emploi). Comme le montre le graphique 4, la plupart de ces déclarants avant l'admission ont été admis au cours des premiers mois de l'année, avant la date d'échéance de la production d'une déclaration de revenus pour l'année précédente. Il semble que certains immigrants admis avant le mois de mai ont produit une déclaration de revenus pour l'année précédant leur année d'admission; ce qu'ils n'étaient pas tenus de faire.

Selon ces résultats, qu'il soit justifié de supprimer les enregistrements présentant les variables de déclarant avant l'admission `Prefiler_ind=1` et `TR_IND=0` des études sur les immigrants ayant une expérience avant leur admission dépend de l'analyse, puisque `TR_IND=0` signifie qu'aucun enregistrement ne figure dans le fichier de permis de résidents non permanents.

Graphique 4**Répartition des déclarants avant l'admission sans permis de résidence non permanente, par mois d'admission**

nombre d'immigrants



Source : Statistique Canada, Base de données longitudinales sur l'immigration, 2014.

Tous les immigrants ayant une expérience avant leur admission ne sont pas considérés comme étant des déclarants avant l'admission : 478 100 immigrants disposent d'enregistrements de résidents non permanents avec la variable `Prefiler_ind=0`. En fonction du sujet étudié, utiliser la variable `TR_IND=1` ou le nombre de permis de résidence temporaire (variable `NUMBER_ALL_PERMITS`) est plus approprié pour étudier les immigrants ayant une expérience avant leur admission. La variable `Prefiler_ind=0` indique qu'aucune déclaration de revenus n'a été produite avant l'admission, mais elle ne signifie pas que la personne n'a pas d'expérience au Canada avant son admission.

7.2.4 Indicateur de conjoint

La BDIM contient des variables permettant aux utilisateurs de données d'obtenir des renseignements sur l'état matrimonial et les conjoints. La section suivante présente les résultats d'une étude menée sur la BDIM de 2012. Aucune modification majeure n'a été apportée depuis aux codes d'état matrimonial ni à l'indicateur relatif à la famille.

L'**identificateur de conjoint** (`SP_IDI`) est dérivé des fichiers de données fiscales. Ce renseignement peut uniquement être dérivé lorsque le répondant déclare son conjoint ou sa conjointe lors de sa déclaration de revenus; cela entraîne une sous-estimation des couples par rapport à l'état matrimonial déclaré dans les fichiers de données fiscales. Il est également possible à partir du FFT1 d'obtenir l'état matrimonial au moment de la déclaration.

Avant 1991, la catégorie « **célibataire** » n'était pas proposée comme **état matrimonial** (`MSTCO`). La catégorie « **conjoint de fait** » est disponible depuis 1992 pour tous les ensembles de données (1982 à 2012). Depuis 1992, la proportion des enregistrements de la BDIM indiquant un état matrimonial « célibataire » varie entre 20 % et 30 %. La proportion de personnes « **séparées** » a diminué, passant de 30 % avant 1992 à 4 % par la suite. Des modifications de tendance n'ont pas influé sur les autres catégories d'état matrimonial.

Une analyse portant sur la répartition de l'état matrimonial (`MSTCO` des fichiers de données fiscales) et l'identificateur de conjoint (`SP_IDI`) a révélé des différences entre ces deux variables. Ces différences sont dues à l'absence de valeurs d'état matrimonial dans certains enregistrements. Dans une situation parfaite, les enregistrements de toutes les personnes mariées présenteraient des renseignements sur le conjoint et ceux de toutes les personnes célibataires n'en indiqueraient pas. L'analyse effectuée a indiqué que la qualité des données s'était améliorée après 1992, lorsque les états matrimoniaux distincts « conjoint de fait » et « célibataire » ont été introduits.

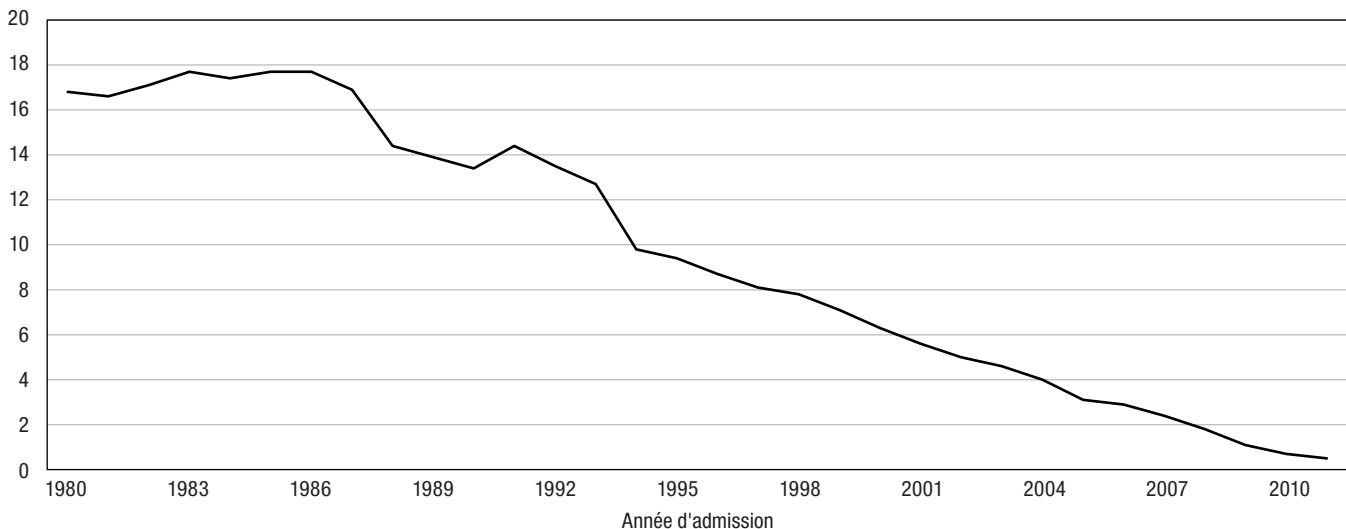
Lacunes d'indication de conjoint

À la suite d'un examen de l'historique longitudinal des immigrants sur la BDIM de 2012, on a relevé des cas où le conjoint ou conjoint de fait manquait (ou différait) pour une année donnée et où le même conjoint était déclaré deux ou trois ans plus tard. Le graphique 5 fournit un résumé de ces lacunes.

Graphique 5

Proportion des cas présentant un identificateur de conjoint (SP__IDI) incohérent, par année d'admission

pourcentage



Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2012.

La plupart des immigrants du fichier ont un conjoint ou non entre 1980 et 2012 selon les fichiers IMDB_T1FF. Il faut souligner qu'aucun état matrimonial (et aucun renseignement sur le conjoint) n'est disponible pour 1,2 million d'immigrants sur environ 6 millions d'immigrants.

7.3 Imputation des variables relatives à la scolarité

Un problème de qualité des données a en outre été relevé quant aux variables relatives au niveau de scolarité et aux années d'études. Une proportion non négligeable de personnes n'ayant pas déclaré leur niveau de scolarité ou d'années d'études ont reçu le code « 0 » soit « aucun » plutôt que « manquant » pour les variables **EDUCATION_QUALIFICATIONS** et **YEARS_OF_SCHOOLING**. Ce problème est devenu courant depuis 2011. En 2011, 35 % des immigrants ont indiqué ne pas avoir suivi d'études, par rapport à environ 10 % dans les années 1990.

Ce problème a été résolu en imputant des variables de scolarité à partir des moyennes des valeurs des variables de scolarité de 2008 à 2010, pour refléter les variables de scolarité de l'année la plus récente. Pour cette imputation, on a utilisé des variables telles que l'âge à l'admission, le groupement de catégories d'immigration 2, la profession envisagée, le sexe et le pays de dernière résidence. La méthodologie d'imputation du plus proche voisin a été employée. La variable d'imputation de la scolarité **Education_imputation_ind** (0 : non; 1 : oui), disponible dans le FRPN, a été créée pour déterminer les enregistrements au sein desquels les variables de scolarité ont été imputées.

7.4 Couverture

7.4.1 Couverture du Fichier intégré des résidents permanents et non permanents (FRPN)

Le Fichier intégré des résidents permanents et non permanents (FRPN) de 2014 contient plus de 7,1 millions d'enregistrements (tableau 7); parmi ceux-ci, près de 5,9 millions (82,5 %) sont couplés à au moins un fichier de données fiscales. Les données sur l'immigration des non-déclarants sont incluses dans un fichier intitulé PNRF_NONFILER_2014, alors que les données sur l'immigration des déclarants est intitulé PNRF_2014. Le tableau suivant présente la répartition des enregistrements en fonction de leur présence dans les différents fichiers d'immigration et de données fiscales. Environ 1,5 million d'enregistrements portent sur des immigrants qui étaient des résidents temporaires avant de devenir résidents permanents; près de 1,4 million de ces enregistrements sont couplés à au moins un fichier de données fiscales. Voir l'annexe B pour des chiffres de répartition détaillés par année d'admission.

Tableau 7
Couverture des résidents permanents

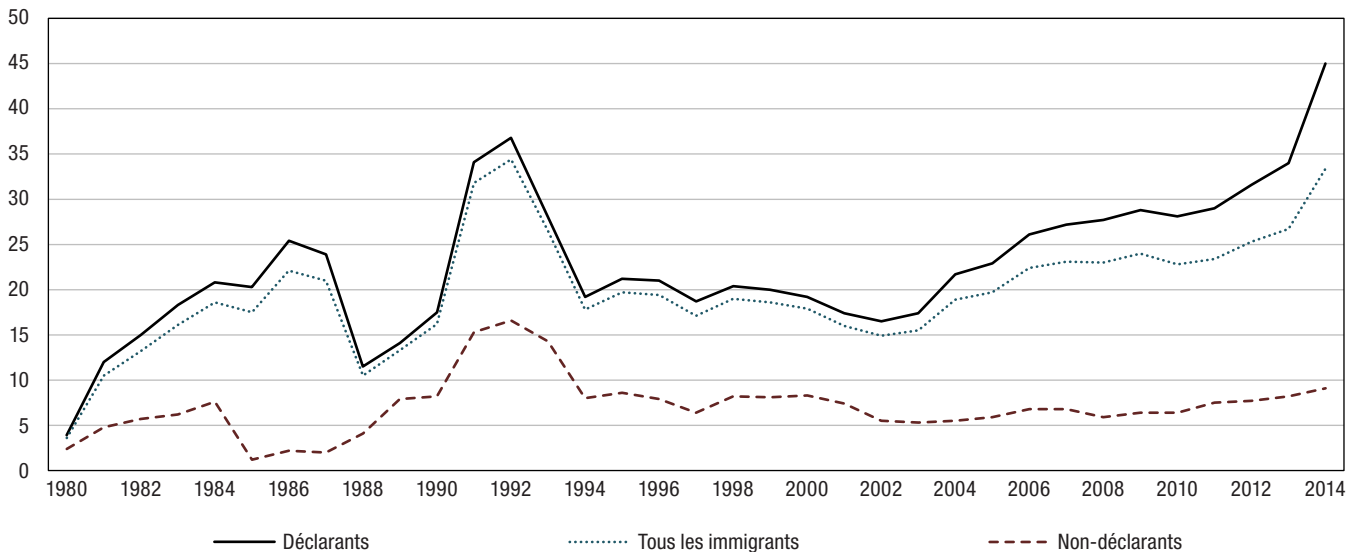
	Résidents permanents	Résidents permanents ayant un permis de résidence non permanente		Nombre de déclarants
		nombre		
Total des déclarants	4 515 500	1 392 900		5 908 400
Total des non-déclarants	1 159 000	91 700		1 250 700
Total	5 674 500	1 484 600		7 159 100
		pourcentage		
Pourcentage de déclarants	79,6	93,8		82,5

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

Les données relatives aux immigrants possédant des permis de résidence non permanente sont disponibles. La proportion d'immigrants ayant une expérience avant l'admission varie en fonction de l'année d'admission (graphique 6); elle va de 4 % en 1980 à 33 % en 2014. Comme il est mentionné dans la section 2.1.2, les permis de résidence non permanente délivrés avant 1980 ne sont pas disponibles dans la BDIM. Par conséquent, la proportion d'immigrants ayant une expérience avant l'admission au début des années 1980 est sous-représentée. La proportion des immigrants déclarants possédant une expérience avant leur admission (ligne pleine) est supérieure à la proportion globale des immigrants possédant une expérience avant leur admission (ligne pointillée), car le taux de couplage de ces immigrants est supérieur à celui des immigrants sans expérience avant leur admission.

Graphique 6
Répartition des immigrants avec permis de résidence non permanente, par année d'admission

pourcentage



Source : Statistique Canada, Base de données longitudinales sur l'immigration, 2014.

7.4.2 Taille du Fichier sur la famille T1 et couverture par année

Les fichiers de données fiscales sont disponibles à partir de 1982 pour les résidents permanents couplés. Certains résidents permanents étaient des résidents non permanents avant leur admission. Le tableau 8 fournit des détails sur la répartition des résidents permanents couplés, avec ou sans permis de résidence non permanente avant leur admission, par année d'imposition. Au moins un fichier de données fiscales est disponible pour 79,6 % de résidents permanents sans permis de résidence non permanente avant leur admission et pour 93,8 % de résidents permanents qui étaient résidents non permanents avant leur admission. Le fait que les résidents permanents possédant des permis temporaires avant leur admission présentent un taux plus élevé de déclarations de revenus que les résidents permanents sans permis avant leur admission peut s'expliquer par une exigence relative au processus de demande de résidence permanente pour les résidents non permanents. Les résidents non permanents qui demandent la résidence permanente doivent en effet respecter l'obligation de déclaration de revenus au Canada. Le nombre de déclarants dans le fichier IMDB_T1FF augmente à mesure des années, puisque la taille de la population dans le champ d'enquête augmente.

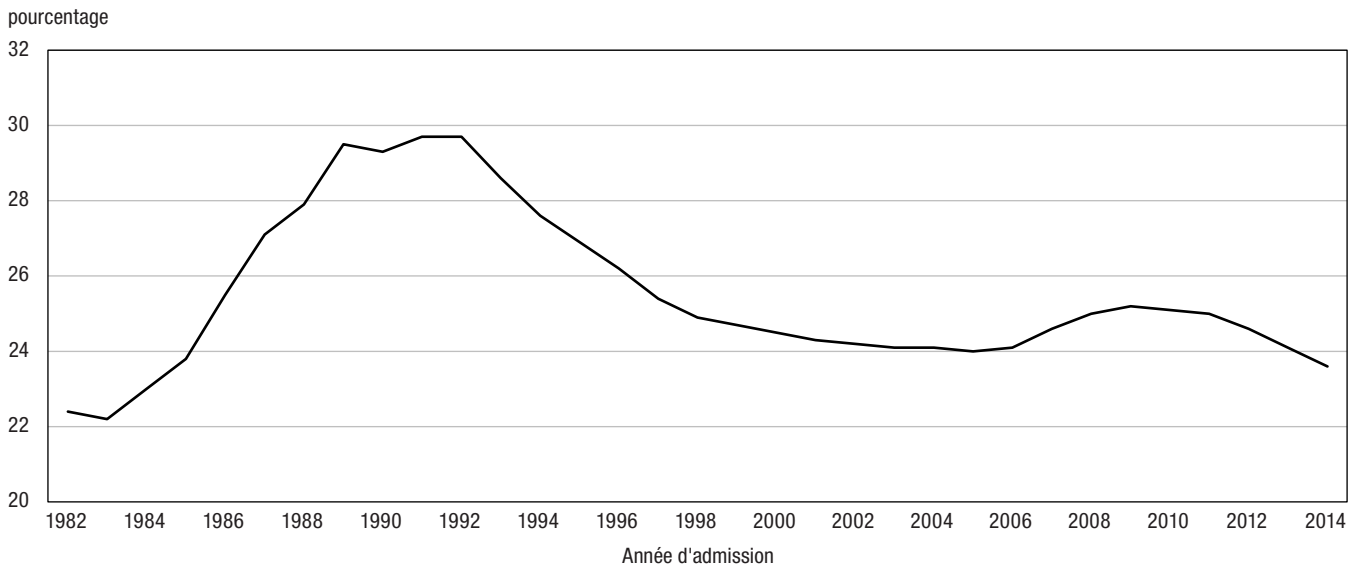
Tableau 8
Résidents permanents et non permanents par année d'imposition

	Résidents permanents	Résidents permanents ayant un permis de résidence non permanente	Nombre de déclarants
	nombre		
1982	183 200	52 900	236 100
1983	218 200	62 400	280 600
1984	255 700	76 200	331 900
1985	289 700	90 700	380 400
1986	346 700	118 800	465 500
1987	404 700	150 600	555 300
1988	493 600	191 000	684 600
1989	603 600	252 800	856 400
1990	721 900	298 800	1 020 700
1991	818 000	346 400	1 164 400
1992	922 800	389 100	1 311 900
1993	1 064 000	427 100	1 491 100
1994	1 184 700	452 000	1 636 700
1995	1 296 500	476 900	1 773 400
1996	1 404 300	497 300	1 901 600
1997	1 518 900	517 500	2 036 400
1998	1 619 300	537 400	2 156 700
1999	1 744 000	572 100	2 316 100
2000	1 888 600	611 200	2 499 800
2001	2 047 500	659 000	2 706 500
2002	2 180 100	695 900	2 876 000
2003	2 304 400	732 400	3 036 800
2004	2 438 000	773 300	3 211 300
2005	2 557 800	806 200	3 364 000
2006	2 710 700	860 900	3 571 600
2007	2 837 600	923 800	3 761 400
2008	2 965 900	989 300	3 955 200
2009	3 084 300	1 040 200	4 124 500
2010	3 212 400	1 077 600	4 290 000
2011	3 344 200	1 112 500	4 456 700
2012	3 463 000	1 130 800	4 593 800
2013	3 595 000	1 138 900	4 733 900
2014	3 701 100	1 140 300	4 841 400
Total des déclarants	4 515 500	1 392 900	5 908 400
Total des non-déclarants	1 159 000	91 700	1 250 700
		pourcentage	
Pourcentage de déclarants	79,6	93,8	82,5

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

Le couplage d'un immigrant est possible sans enregistrement de FFT1, puisque divers fichiers auxiliaires permettent de coupler les immigrants et les résidents non permanents. On ne s'attend pas, par exemple, à ce que les enfants produisent de déclarations de revenus, mais ils peuvent être couplés au Fichier de contrôle des couplages (voir la section 2.3 pour une description du FCC). Le taux de couplage des immigrants admis à partir de 1980 était de 89 %, alors qu'au moins un FFT1 était disponible pour 82,5 % de ces immigrants.

Le graphique 7 montre que la proportion des résidents permanents qui étaient résidents non permanents avant leur admission, par année, varie de 22,2 %, plus bas niveau correspondant à l'année d'imposition de 1983, à 29,7 %, niveau le plus élevé correspondant aux années d'imposition de 1991 et 1992. Depuis la fin des années 1990, cette proportion est demeurée stable, à environ 25 %.

Graphique 7**Pourcentage de résidents permanents qui étaient résidents non permanents avant leur admission, par année d'imposition**

Source : Statistique Canada, Base de données longitudinales sur l'immigration, 2014.

Un immigrant qui produit une déclaration de revenus pour une année donnée n'en produira pas nécessairement l'année suivante. Si une personne A, par exemple, a été admise en 1983, elle peut figurer dans les fichiers de données fiscales de 1984 à 1999, mais pas dans le fichier de 2000, et y figurer ensuite à nouveau de 2001 à 2013. Ainsi, des fichiers de données fiscales étaient disponibles tous les ans pour 34 % des déclarants de la cohorte de 1980. Une émigration externe, un décès ou un retard de déclaration peut expliquer, en outre, que des immigrants déclarants aient pu ne plus produire de déclaration de revenus ou ne pas en produire certaines années.

La plupart des immigrants déclarent leurs revenus pour la première fois l'année de leur admission ou l'année précédente ou suivante. Ainsi, sur les 250 800 immigrants admis en 2006, 101 530 (40,5 %) ont produit une déclaration de revenus pour la première fois en 2006, alors que 16 010 (6,3 %) l'ont fait en 2007 et 3 135 (1,25 %) en 2014.

7.5 Évaluation de la qualité du Fichier intégré des résidents permanents et non permanents (FRPN)

Le contenu du FRPN de 2014 a été évalué. Alors que la collecte des données fiscales et relatives à l'admission est obligatoire pour les personnes dans le champ d'enquête, ces dernières peuvent ne pas remplir tous les champs. Ces champs ont pu être laissés vides parce que les renseignements en question n'étaient pas jugés de haute importance (champs facultatifs), parce que le répondant ne connaissait pas la réponse ou pour d'autres raisons, que les utilisateurs des données ne peuvent connaître (p. ex. refus) (McLeish, 2011). L'absence de réponse à certaines questions peut présenter des problèmes lorsque la BDIM est utilisée à des fins statistiques, notamment dans les cas suivants :

1. Si l'utilisateur de la base de données souhaite produire un échantillon fondé sur les caractéristiques pour lesquelles des enregistrements manquent, une erreur de couverture peut survenir (c.-à-d. les personnes incluses dans la base d'échantillon peuvent ne pas être représentatives de la population cible).
2. S'il s'agit d'une absence de données informatives (p. ex. l'absence du renseignement n'est pas un cas aléatoire; l'absence de réponse indique ce que la réponse aurait été), toute analyse utilisant ces variables serait biaisée.

On a évalué la présence de variables manquantes et de valeurs non valides. Les chiffres présentés dans cette section sont arrondis. Les valeurs non valides sont soit incohérentes soit elles ne figurent pas dans les tableaux de métadonnées à la disposition des utilisateurs (voir la composante sur l'immigration du dictionnaire de données

en annexe). La plupart des problèmes de qualité figurant dans le tableau 9 concernent des données recueillies dans les années 1980 et 1990. Il convient de souligner que des valeurs semblant valides peuvent également être erronées.

La variable de **numéro d'identification du cas** (CASE_ID) présente des taux de réponse aux éléments se situant généralement dans la frange supérieure des 90 % (généralement plus de 99 %). Cependant, pour certaines années d'admission, le taux de réponse diminue significativement (jusqu'à 80 % en 1991 et 1992). Par conséquent, toute analyse fondée sur cette variable pour toutes les années d'admission sous-représentera ces années, pour lesquelles le manque de réponse à l'élément est plus élevé (p. ex. 1986, 1987, 1990, 1991, 1992, 1993). Aucune détection des valeurs non valides n'a eu lieu pour la variable de numéro d'identification du cas (CASE_ID).

La variable d'âge à l'admission **Landing_age** est définie comme non valide lorsqu'elle dépasse 99, même s'il est possible dans certains cas que ces valeurs soient exactes. Il convient de souligner que, selon les valeurs de cette variable, le nombre d'immigrants admis après l'âge de 99 ans était bien plus élevé entre 1986 et 1994 que pour les autres années d'admission. Il pourrait s'agir d'un problème de saisie de données.

Dans le FRPN de 2014, dix enregistrements présentaient une **année de naissance** antérieure à 1880 et cinq, une année d'admission antérieure à l'année de naissance. La valeur de **sexe** manque pour 570 immigrants, principalement pour des personnes admises dans les années 1980.

Les variables relatives au pays présentent également des problèmes de qualité. Le **pays de naissance** manque dans certains enregistrements pour pratiquement toutes les années d'admission. Des valeurs manquent, par exemple, pour plus de 100 enregistrements dans chacune des années de 1985 à 1993. Le **pays de citoyenneté** manque pour moins de 10 enregistrements par année d'admission pour la plupart des années (à l'exception de 2004, 2005 et 2006, pour lesquelles plus de 40 enregistrements ne contiennent pas d'année d'admission). Le **pays de résidence** manque pour de nombreux enregistrements d'admission pour 2013 (cette valeur manque ainsi pour 920 enregistrements, soit 0,5 % des admissions de cette année-là) et pour 2014 (cette valeur manque pour 3500 enregistrements, soit 2 % des admissions de cette année-là). Cela peut s'expliquer par le fait que ce champ est devenu facultatif en 2013.

Les **variables de scolarité** après imputation (voir la section 6.3) présentent plus de 150 valeurs manquantes par année d'admission de 1980 à 1984; cela équivaut à un taux de valeurs manquantes par année d'admission de moins de 0,5 %.

Le pourcentage de réponses valides pour les **variables de profession** se situe entre 98 % et 99,9 % pour toutes les années d'admission.

La plupart des valeurs manquantes pour les variables de statut dans la famille **Family_Status**, de langue maternelle **Mother_Tongue**, de langue officielle **Official_Language**, de point de service **Point_of_Service**, de point d'entrée **Point_of_Entry**, d'immigrants sélectionnés par le Québec **CSQ_IND** et de province de destination **Destination_Province** concernent les enregistrements des années d'admission antérieures à 1999. La variable de langue maternelle **Mother_Tongue** manque pour 450 enregistrements relatifs aux admissions de 2011.

Les variables d'**année et de mois de décès** manquent pour certaines personnes reconnues comme décédées (Death_Indicator=1). La valeur « 9999 » a été attribuée à la variable d'année de décès Death_Year et la valeur « 99 » à celle du mois de décès Death_Month dans les cas où l'année et le mois de décès étaient inconnus.

Tableau 9
Évaluation de la qualité du Fichier intégré des résidents permanents et non permanents

Variables du FRPN	Réponses valides		Champs vides		Réponses non valides	
	nombre	pourcentage	nombre	pourcentage	nombre	pourcentage
Case_ID	5 737 500	97,11	170 900	2,89	0	0,00
Landing_age	5 906 500	99,97	10	0,00	1 900	0,03
Birth_Year	5 908 400	100,00	0	0,00	15	0,00
Gender	5 907 900	99,99	570	0,01	0	0,00
Country_Birth	5 906 100	99,96	2 400	0,04	0	0,00
Country_Citizenship	5 908 000	99,99	430	0,01	0	0,00
Country_Residence	5 903 200	99,91	5 200	0,09	0	0,00
Country_residence_world_area	5 903 200	99,91	5 270	0,09	0	0,00
Education_Qualification	5 906 800	99,97	1 700	0,03	0	0,00
Level_of_Education	5 907 000	99,98	1 400	0,02	0	0,00
Years_of_Schooling	5 906 800	99,97	1 600	0,03	0	0,00
Landing_age_6_groups	5 908 400	100,00	10	0,00	0	0,00
Landing_age_9_groups	5 908 400	100,00	10	0,00	0	0,00
Occupation_ID	5 904 000	99,92	4 500	0,08	8	0,00
NOC5-NOC2	5 858 300	99,15	50 200	0,85	0	0,00
LM_Intention	5 859 100	99,16	49 400	0,84	0	0,00
Family_Status	5 906 100	99,96	2 300	0,04	0	0,00
Family_Status_rollup	5 906 100	99,96	2 300	0,04	0	0,00
Marital_status	5 907 800	99,99	670	0,01	0	0,00
Marital_status_rollup	5 907 800	99,99	670	0,01	0	0,00
Mother_Tongue	5 906 100	99,96	2 300	0,04	0	0,00
Official_Language	5 906 300	99,96	2 100	0,04	0	0,00
Point_of_Service	5 906 900	99,97	1 500	0,03	0	0,00
Point_of_Entry	5 907 200	99,98	1 200	0,02	0	0,00
Special_Program	1 131 400	19,15	4 777 100	80,85	0	0,00
CSQ_ind	5 908 300	100,00	180	0,00	0	0,00
Destination_CD	5 820 500	98,51	87 900	1,49	0	0,00
Destination_CMA	5 820 500	98,51	87 900	1,49	0	0,00
Destination_CSD	5 820 500	98,51	87 900	1,49	0	0,00
Destination_Province	5 906 300	99,96	2 100	0,04	0	0,00
Permits and NPR-specific variables	1 345 000	100,00	0	0,00	0	0,00
Death_Year and Death_Month	208 900	99,51	1 030	0,49	0	0,00

Notes : FRPN : Fichier intégré des résidents permanents et non permanents. RNP : résidents non permanents. Seules les variables présentant des valeurs manquantes ou non valides ont été incluses dans ce tableau. Tous les chiffres sont arrondis.

Source : Statistique Canada Base de données longitudinales sur l'immigration.

8 Comparabilité

8.1 Variation historique de la couverture

Au cours des années, la couverture et le contenu de la BDIM ont évolué. Les fichiers IMDB_T1FF initiaux comprenaient uniquement des données sur les immigrants admis au Canada à partir de 1980. Depuis la parution de la BDIM de 2013, les résidents non permanents déclarants ont été ajoutés aux fichiers IMDB_T1FF pour les années d'imposition de 1982 et ultérieures. Du fait de ce changement, on peut désormais disposer de renseignements sur les permis de résidence temporaire des immigrants ayant une expérience au Canada avant leur admission.

En 2012, la BDIM a fait l'objet d'une refonte. La couverture de la BDIM a été modifiée pour inclure à la base de données des immigrants ayant obtenu leur statut d'immigrant reçu à partir de 1980 et ayant produit au moins une déclaration de revenus depuis 1982, qu'ils en aient produit une ou non après leur admission. La BDIM incluait initialement uniquement des personnes ayant obtenu le statut d'immigrant reçu à partir de 1980 et ayant produit au moins une déclaration de revenus après avoir été reçues comme immigrants. Avant ce cycle, la BDIM incluait jusqu'aux seize premières années des fichiers de données fiscales relatifs à un résident permanent donné (Dryburgh, 2004). Cette limite du nombre de fichiers de données fiscales pour une personne donnée ne s'applique plus.

Les données fiscales incluses dans la BDIM provenaient à l'origine principalement des formulaires T1 et seulement quelques variables fiscales importantes au niveau personnel étaient conservées. Dans la BDIM de 2006 et les BDIM suivantes, on a utilisé les fichiers FFT1 relatifs à 1982 et aux années ultérieures; cela a produit un taux de couplage initial de 80 %. À partir de là, la BDIM a exclu les fichiers de données fiscales de 1980 et 1981, puisque les renseignements pour ces années ne figurent pas dans le FFT1.

On utilisait initialement le Système de soutien des opérations des bureaux locaux (SSOBL) pour rassembler les données d'immigration incluses dans la BDIM. Pour l'année d'immigration de 2013 et les suivantes, le système mondial de gestion des cas (SMGC) sera utilisé. IRCC a, par conséquent, cessé de fournir certaines variables. Ces anciennes variables seront disponibles dans le fichier PNRF_extra; elles figurent dans la composante sur l'immigration du dictionnaire des données de la BDIM.

8.2 Changements méthodologiques

La méthodologie employée pour effectuer le couplage d'enregistrements a fait l'objet de changements au cours des années.

Le taux de couplage initial de la BDIM était de 55 % pour la BDIM de 1995 (Langlois et Dougherty, 1997). Les méthodes et outils employés pour le couplage d'enregistrements ont depuis évolué; ce qui explique l'amélioration des taux de couplage au cours des années.

À la fin des années 2000, le taux de couplage était d'environ 81 %. Pour la BDIM de 2012, les renseignements sur les personnes à charge ont servi pour procéder au couplage d'enregistrements; cela a permis de coupler une plus grande proportion d'enfants immigrants. Ces renseignements figuraient dans le fichier de la PFCE. Il convient de noter que l'ajout des enfants n'améliore pas le taux de déclarants. Grâce aux changements méthodologiques, le taux de couplage pour la BDIM de 2014 a atteint 89 %.

8.3 Changements du contenu de la base de données historiques

Veillez-vous reporter aux dictionnaires de la BDIM (composantes sur l'immigration et les impôts) pour une description complète du contenu des fichiers. Certaines importantes modifications récentes du contenu de la BDIM sont énumérées ci-après.

En 2012, la BDIM de 2009 a fait l'objet d'un remaniement; un indicateur a en effet été ajouté pour détecter les valeurs aberrantes dans les fichiers FFT1 créés. La variable d'identificateur de conjoint (SP_IDI) a été introduite dans la BDIM de 2010, permettant de recenser les immigrants ayant un conjoint immigrant. Les années et mois de décès ont été ajoutés à la BDIM de 2013; cela a permis de recenser les immigrants admis au Canada à partir

de 1980 et décédés depuis. À la suite de l'ajout à la BDIM des données relatives aux résidents non permanents, certains détails sur les permis de résidence temporaire (type, dates d'entrée en vigueur, etc.) sont désormais disponibles depuis la BDIM de 2013.

8.4 Comparabilité avec d'autres sources des données sur l'immigration

La BDIM est l'un des nombreux programmes statistiques pouvant servir à produire des estimations relatives à la population immigrante. Dans certains cas, ces estimations différeront du fait de plusieurs facteurs, comme la couverture et des limites liées au type de données (données administratives, données d'enquête, données de recensement). Certains de ces programmes statistiques et différences relatives à la BDIM sont décrits dans la présente section. On utilise la BDIM de 2013 pour procéder à des comparaisons.

8.4.1 Banque de données administratives longitudinales (DAL)

La Banque de données administratives longitudinales¹¹ (DAL) est un échantillon longitudinal de 20 % des déclarants canadiens. Elle est couplée à la BDIM pour inclure un échantillon de 20 % des enregistrements de la BDIM et pour ajouter des variables propres aux immigrants, comme l'année d'admission, la catégorie d'immigration et l'état matrimonial à l'admission. Elle contient des renseignements sur les particuliers et sur les familles de recensement. Elle est utile aux analyses longitudinales, qui comparent le revenu et la mobilité des immigrants à ceux des déclarants canadiens. Toute analyse comparant des immigrants déclarants à la population de déclarants canadiens devrait se servir de cet ensemble de données.

La DAL contient un moins grand nombre de variables relatives à l'immigration que la BDIM. Les renseignements portant sur la période précédant l'admission, par exemple (comme le nombre de permis de travail et d'études), ne figurent pas dans la DAL. Les renseignements sur l'admission, notamment la profession envisagée et la province de destination, ne figurent pas non plus dans la DAL.

Le tableau 10 présente le revenu total moyen et médian (XTIRC) pour l'année d'imposition de 2012 des immigrants admis entre 1982 et 2013, par sexe, afin d'illustrer le degré de comparaison des estimations produites à partir de ces bases de données. Le revenu total moyen et médian par sexe, comme nous nous y attendons, est similaire pour les deux sources de données. Les différences peuvent s'expliquer par le fait que la DAL est un échantillon de 20 % de la population canadienne et que la BDIM est un recensement des immigrants déclarants couplés admis au Canada depuis 1980. Les chiffres de population diffèrent, mais ni l'une ni l'autre de ces sources ne devrait être utilisée pour des chiffres de population; la DAL étant un échantillon et la BDIM étant limitée aux immigrants déclarants. La population de la DAL est estimée en multipliant les enregistrements par un facteur de pondération de 5.

Tableau 10
Comparabilité du revenu total de 2012 entre la DAL et la BDIM pour les immigrants admis entre 1982 et 2013

	Hommes			Femmes			Total		
	Population nombre	Moyenne dollars	Médiane	Population nombre	Moyenne dollars	Médiane	Population nombre	Moyenne dollars	Médiane
Particulier									
BDIM	2 776 700	41 900	29 400	2 906 000	28 700	20 600	5 682 690	35 000	24 200
DAL	2 686 300	41 700	29 200	2 803 100	28 700	20 500	5 489 390	34 900	24 100
Famille									
BDIM	...	73 700	56 300	...	69 900	51 400	...	71 700	53 700
DAL	...	73 800	56 500	...	69 700	51 500	...	71 700	53 900

... n'ayant pas lieu de figurer

Note : BDIM : Base de données longitudinales sur l'immigration; DAL : Banque de données administratives longitudinales.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2013 et Banque de données administratives longitudinales de 2013.

11. Pour de plus amples détails sur la DAL, reportez-vous à la description figurant sur le [site Web](#) de Statistique Canada.

Dans le tableau 11, la comparabilité a été limitée au revenu total pour 2012 des immigrants admis en 2011. Les différences estimées observées entre la BDIM et la DAL pour ce groupe sont supérieures à celles observées pour la population immigrante admise pour toute année entre 1980 et 2013. Cela peut s'expliquer par le fait que la population d'intérêt est moins importante et plus précise. Les estimations de la DAL sont dérivées des enregistrements inclus dans l'échantillon de 20 % des immigrants admis en 2011. Ces enregistrements ne correspondent pas toujours au 20 % de la population précise de la BDIM. Ils constituent probablement une proportion plus réduite de la population précise de la BDIM, puisque l'échantillon n'a pas été prélevé pour être représentatif de cette population particulière. Les estimations de la BDIM sont dérivées de la population immigrante couplée admise en 2011 et ayant produit une déclaration de revenus en 2012. Ainsi, les estimations de la DAL peuvent prendre des valeurs légèrement différentes de celles de la BDIM lors de l'examen de sous-ensembles de population.

Tableau 11
Comparabilité du revenu total moyen et médian pour 2012 des immigrants admis en 2011

	Hommes		Femmes		Total	
	Moyenne	Médiane	Moyenne	Médiane	Moyenne	Médiane
	dollars					
Particulier						
BDIM	30 100	22 400	18 900	14 100	24 300	17 800
DAL	29 500	22 100	18 700	13 900	23 900	17 500
Famille						
BDIM	49 900	39 300	48 200	37 100	49 000	38 200
DAL	49 300	39 000	48 000	36 600	48 600	37 800

Note : BDIM : Base de données longitudinales sur l'immigration; DAL : Banque de données administratives longitudinales.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2013 et Banque de données administratives longitudinales de 2013.

8.4.2 Recensement

Le questionnaire détaillé du recensement et l'Enquête nationale auprès des ménages (ENM) de 2011 permettent de recueillir des données relatives aux immigrants. Ces données sont recueillies pour une portion de la population (reportez-vous à la description du programme du recensement pour connaître la proportion exacte; cette valeur a varié au cours du temps). On y recueille ainsi le lieu de naissance, le lieu de naissance des parents, le statut d'immigrant, l'année d'immigration, l'âge à l'immigration et la citoyenneté. Depuis le Recensement de 2016, la catégorie d'immigration est également disponible. Le recensement recueille des données sur les Canadiens des première, deuxième et plus anciennes générations, alors que la BDIM recueille uniquement des données sur les nouveaux arrivants et leur famille. Le recensement comprend également des données sur les minorités visibles, la scolarité, le logement et la langue pour l'année de recensement; à moins que l'année d'admission ne soit également une année de recensement, ce dernier ne présente pas ces renseignements pour l'année d'admission. Le recensement ne permet pas d'effectuer d'études longitudinales sur la situation économique ou la mobilité à long terme des immigrants. De plus amples détails sur le programme du recensement figurent sur le [site Web](#) de Statistique Canada.

L'Enquête nationale auprès des ménages (ENM) de 2011 a permis d'estimer que plus de 4,6 millions d'immigrants vivant au Canada en 2011 avaient été admis entre 1981 et 2011 (source : tableau CANSIM 99-010-X2011026). Le tableau 12 compare les estimations de la population immigrante par décennie d'admission selon l'ENM et le FRPN. Le FRPN de 2013 ne devrait pas être utilisé pour estimer les chiffres de population, même après suppression des enregistrements des personnes décédées. Le faire contribuerait à une surestimation de la population immigrante vivant au Canada admise entre 1981 et 2011, car le FRPN ne tient pas compte de l'émigration. De plus, le FRPN est un sous-ensemble de la population immigrante, puisque seuls les déclarants y sont inclus. Cela peut expliquer les chiffres de population plus faibles pour le FRPN que pour l'ENM pour la cohorte d'immigrants la plus récente (de 2001 à 2011). Le nombre d'immigrants au tableau 12 prend en considération le nombre de décès en utilisant la variable `Death_indicator` (décrit à la section 7.2.2).

Tableau 12

Comparabilité des chiffres de population entre la Base de données longitudinales sur l'immigration et l'Enquête nationale auprès des ménages

Décennie d'admission	Chiffres de l'ENM	Chiffres du FRPN pour 2013
	nombre	
1981 à 1990	949 890	1 052 650
1991 à 2000	1 539 055	1 896 235
2001 à 2011	2 154 985	2 120 290
Total	4 643 930	5 069 175

Note : BDIM : Base de données longitudinales sur l'immigration; ENM : Enquête nationale auprès des ménages; FRPN : Fichier intégré des résidents permanents et non permanents.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2013 et Enquête nationale auprès des ménages de 2011.

8.4.3 Enquête longitudinale auprès des immigrants du Canada (ELIC)

L'Enquête longitudinale auprès des immigrants du Canada (ELIC) vise à fournir des renseignements sur la manière dont les nouveaux immigrants s'adaptent à la vie au Canada et à comprendre les facteurs pouvant faciliter ou empêcher cette adaptation. L'ELIC est conçue pour examiner les quatre premières années de l'établissement. On a recueilli en trois vagues les données relatives à des immigrants de 15 ans et plus admis au Canada entre le 1^{er} octobre 2000 et le 30 septembre 2001. L'ELIC permet d'étudier les compétences linguistiques, le logement, la scolarité, la reconnaissance des diplômés étrangers, l'emploi, la santé, les valeurs et les attitudes, l'établissement et l'utilisation des réseaux sociaux, le revenu et les perceptions relatives à l'établissement au Canada. La BDIM contient des caractéristiques comme la scolarité et la langue uniquement à l'admission, alors que l'ELIC permet une évaluation de l'évolution au cours du temps. De plus amples détails sur l'ELIC figurent sur le [site Web](#) de Statistique Canada.

Selon l'ELIC, il a été estimé que 164 200 immigrants âgés de 15 ans et plus ont été admis au Canada entre le 1^{er} octobre 2000 et le 30 septembre 2001. L'estimation de cette même population est 156 670 selon la BDIM de 2013 en utilisant des calculs basés sur le FRPN (tableau 13). Une partie de la différence est due à la fois à l'exclusion des non-déclarants de l'estimation du FRPN et à l'émigration n'ont capturée au sein de la BDIM. Une autre partie s'explique par le fait que l'ELIC est une enquête introduisant des estimations de variance. Comme le montre le tableau 14, les proportions de la couverture selon la tranche d'âge varient en fonction de la tranche d'âge, même si la population de l'ELIC est en âge de déclarer des revenus. Il convient de noter que l'âge dans l'ELIC est celui environ six mois après l'admission, alors que l'âge dans la BDIM est celui à l'admission. De plus, le calcul des estimations de l'ELIC est fondé sur les facteurs de pondération de la première vague, conçus pour estimer le nombre d'immigrants dans cette cohorte vivant toujours au Canada six mois après leur admission. La proportion inférieure d'immigrants âgés de 65 ans et plus pourrait être due à une proportion inférieure de déclarants pour cette tranche d'âge. Le nombre d'immigrants âgés de 15 à 24 ans plus élevé dans la BDIM que dans l'ELIC s'explique probablement par la non-prise en compte de l'émigration.

Tableau 13

Répartition par sexe : comparaison de la Base de données longitudinales sur l'immigration et de l'Enquête longitudinale auprès des immigrants du Canada

	ELIC de 2003		FRPN de 2013	
	nombre	pourcentage	nombre	pourcentage
Hommes	81 550	49,7	77 640	49,6
Femmes	82 650	50,3	78 830	50,4
Total	164 200	100,0	156 470	100,0

Note : ELIC : Enquête longitudinale auprès des immigrants du Canada; FRPN : Fichier intégré des résidents permanents et non permanents.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2013 et Enquête longitudinale auprès des immigrants du Canada, 1^{re} vague, 2003.

Tableau 14

Répartition par tranche d'âge : comparaison de la Base de données longitudinales sur l'immigration et de l'Enquête longitudinale auprès des immigrants du Canada

Tranche d'âge	ELIC de 2003		FRPN de 2013	
	nombre	pourcentage	nombre	pourcentage
15 à 24 ans	26 700	16,3	28 000	17,9
25 à 34 ans	65 500	39,9	63 100	40,3
35 à 49 ans	54 000	32,9	49 000	31,3
50 à 64 ans	12 900	7,8	12 300	7,8
65 ans ou plus	5 100	3,1	4 100	2,6
Total	164 200	100,0	156 500	100,0

Note : ELIC : Enquête longitudinale auprès des immigrants du Canada; FRPN : Fichier intégré des résidents permanents et non permanents.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2013 et Enquête longitudinale auprès des immigrants du Canada, 1^{re} vague, 2003.

9 Nouvelles analyses possibles grâce à la BDIM

La BDIM a été créée pour permettre d'effectuer une analyse de sujets reliés à l'immigration; cette section présente un aperçu des analyses possibles grâce aux renseignements supplémentaires désormais disponibles dans la BDIM. Comme le décrit le présent rapport, le contenu de la BDIM a évolué; cela a accru ses capacités analytiques. Trouvez ci-après des exemples d'analyses que la BDIM permet d'effectuer.

9.1 Possibilités analytiques découlant des données relatives aux résidents non permanents

L'ajout des données relatives aux résidents non permanents élargit la portée de l'analyse que permet actuellement la BDIM. Les renseignements de nombre et de type de permis obtenus avant l'admission peuvent permettre d'établir le profil avant leur admission des immigrants ayant une expérience avant leur admission. De plus, les données relatives aux résidents non permanents peuvent être utilisées pour recenser les résidents permanents ayant ou non une expérience au Canada avant leur admission. En comparant ces populations (avec ou sans expérience au Canada avant l'admission), il est possible d'évaluer les répercussions de l'expérience au Canada avant l'admission sur la situation économique et sur les tendances de mobilité des immigrants. Le profil socioéconomique particulier au moment de la délivrance du permis de résidence temporaire est également désormais disponible et permet d'évaluer la situation économique et la mobilité avant l'admission. L'évolution de la profession envisagée, du niveau de compétence et du niveau de scolarité en fonction de la délivrance de permis de résidence temporaire est également disponible.

9.2 Possibilités analytiques grâce aux données relatives aux décès

L'ajout au FRPN des variables d'indicateur, de mois et d'année de décès permet d'estimer la proportion d'enregistrements inclus dans la BDIM qui concernent des immigrants décédés. Ces variables complètent la variable d'année de décès (YOD) des fichiers de données fiscales. La variable YOD est uniquement présente dans les cas où un formulaire T1 a été rempli de façon posthume au nom de la personne décédée, alors que l'année et le mois du décès figurent dans tout enregistrement couplé à l'ensemble de données sur la mortalité, quel que soit le profil de déclaration de revenus. De nouvelles analyses possibles peuvent inclure l'évaluation de profils économiques avant le décès et l'étude des attentes relatives à la vie après l'admission, par catégorie d'immigration et profil économique.

10 Résumé

La BDIM est un ensemble de données combinant des enregistrements sur l'immigration et les impôts à des fins d'analyses socioéconomiques et relatives à la mobilité des immigrants (avec ou sans expérience avant l'admission) admis au Canada depuis 1980. La BDIM permet d'effectuer des analyses concernant les déclarants. C'est un point important, car les déclarants peuvent présenter un profil différent des non-déclarants. Le présent rapport technique a été élaboré pour fournir une description complète de la qualité des données de la BDIM ainsi que des modifications récemment apportées à cette base de données.

Comment accéder à la BDIM :

Comme le décrit la section 6, plusieurs produits sont à la disposition des chercheurs. Ils sont accessibles sur le site Web de Statistique Canada en sélectionnant « Diversité ethnique et immigration » comme sujet, puis [Base de données longitudinales sur l'immigration \(BDIM\)](#) sous Produits vedettes.

Annexe

A) Liens vers les principaux documents et pages Web relatifs à la BDIM

Dictionnaires (composantes sur les impôts et l'immigration) :

Disponible aux utilisateurs de données ou sur demande en communiquant avec Statistique Canada par courriel à l'adresse STATCAN.infostats-infostats.STATCAN@canada.ca)

Tableaux CANSIM de la BDIM :

<http://www5.statcan.gc.ca/COR-COR/objList?tgtObjType=ARRAY&srcObjType=SDDS&srcObjId=5057&lang=fra>

Historique de la BDIM :

http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getInstanceList&Id=7196

Publications de la BDIM dans Le Quotidien :

<http://www5.statcan.gc.ca/COR-COR/objList?tgtObjType=DAILYART&srcObjType=SDDS&srcObjId=5057&lang=fra>

Analyses fondées sur la BDIM :

<http://www5.statcan.gc.ca/COR-COR/objList?tgtObjType=STUDIES&srcObjType=SDDS&srcObjId=5057&lang=fra>

Indice des prix à la consommation (62-001-X) :

<http://www5.statcan.gc.ca/olc-cel/olc.action?lang=fr&ObjId=62-001-X&ObjType=2>

Description des estimations annuelles du revenu des familles de recensement et des particuliers (Fichier sur la famille T1) :

http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=4105&lang=fr&db=imdb&adm=8&dis=2.

B) Couverture

La BDIM de 2014 a été utilisée pour produire ces chiffres. Les déclarants sont des immigrants couplés ayant produit au moins une déclaration de revenus depuis 1982.

Tableau 15
Répartition des déclarants et des non-déclarants par année d'admission

Année d'admission	Déclarants ¹				Non-déclarants				Total			
	Immigrants	RP	RP RNP	Décès	Immigrants	RP	RP RNP	Décès	Immigrants	RP	RP RNP	Déclarants pourcentage
	nombre											
1980	114 200	109 700	4 500	15 600	28 900	28 200	700	1 700	143 100	137 900	5 200	79,8
1981	101 500	89 300	12 200	13 900	27 000	25 700	1 300	1 400	128 600	115 000	13 500	78,9
1982	98 100	83 400	14 700	12 500	23 000	21 700	1 300	1 300	121 100	105 100	16 000	81,0
1983	72 800	59 500	13 300	10 100	16 200	15 200	1 000	1 100	89 000	74 700	14 300	81,8
1984	73 500	58 200	15 300	9 300	14 500	13 500	1 100	1 000	88 000	71 700	16 400	83,5
1985	71 400	56 900	14 500	8 200	12 600	12 400	150	700	83 900	69 300	14 650	85,1
1986	84 800	63 300	21 500	8 300	13 900	13 700	300	600	98 800	77 000	21 800	85,8
1987	131 400	100 100	31 400	10 300	19 700	19 400	400	700	151 200	119 500	31 800	86,9
1988	138 800	122 900	15 900	9 700	22 000	21 100	900	600	160 700	144 000	16 800	86,4
1989	165 500	142 200	23 300	10 000	25 200	23 200	2 000	600	190 700	165 400	25 300	86,8
1990	186 100	153 500	32 600	10 600	29 300	26 900	2 400	500	215 400	180 400	35 000	86,4
1991	203 100	133 900	69 200	11 600	28 800	24 400	4 400	600	231 800	158 300	73 600	87,6
1992	223 800	141 500	82 300	11 600	30 100	25 100	5 000	500	253 900	166 600	87 300	88,1
1993	226 900	163 300	63 600	10 900	28 700	24 600	4 100	500	255 700	187 900	67 700	88,7
1994	196 000	158 400	37 600	8 800	27 600	25 300	2 200	400	223 600	183 700	39 800	87,7
1995	186 600	147 100	39 500	7 100	25 500	23 400	2 200	400	212 100	170 500	41 700	88,0
1996	196 300	155 000	41 300	6 000	29 000	26 700	2 300	300	225 300	181 700	43 600	87,1
1997	187 400	152 300	35 100	5 000	28 000	26 200	1 800	200	215 400	178 500	36 900	87,0
1998	154 100	122 700	31 400	3 600	19 600	18 000	1 600	200	173 700	140 700	33 000	88,7
1999	167 200	133 700	33 500	3 400	22 100	20 300	1 800	160	189 400	154 000	35 300	88,3
2000	199 000	160 600	38 300	3 400	27 800	25 500	2 300	200	226 700	186 100	40 600	87,8
2001	216 100	178 600	37 500	3 500	33 700	31 200	2 500	180	249 800	209 800	40 000	86,5
2002	193 800	161 800	32 000	3 200	34 400	32 500	1 900	180	228 200	194 300	33 900	84,9
2003	184 700	152 500	32 200	2 600	35 800	34 000	1 900	150	220 500	186 500	34 100	83,8
2004	194 800	152 500	42 300	2 100	40 300	38 100	2 200	130	235 100	190 600	44 500	82,9
2005	212 700	164 000	48 700	1 700	48 800	45 800	2 900	80	261 400	209 800	51 600	81,4
2006	203 800	150 600	53 100	1 800	47 000	43 800	3 200	110	250 800	194 400	56 300	81,3
2007	188 900	137 600	51 300	1 400	47 000	43 800	3 200	90	235 900	181 400	54 500	80,1
2008	193 600	140 000	53 600	1 200	52 700	49 700	3 100	90	246 300	189 700	56 700	78,6
2009	196 600	139 800	56 700	1 000	54 800	51 300	3 500	90	251 300	191 100	60 200	78,2
2010	210 800	151 400	59 300	700	69 000	64 600	4 400	70	279 700	216 000	63 700	75,4
2011	183 700	130 500	53 200	500	64 100	59 300	4 800	40	247 800	189 800	58 000	74,1
2012	189 000	129 300	59 700	400	67 800	62 600	5 200	10	256 900	191 900	64 900	73,6
2013	185 800	122 700	63 100	300	72 200	66 200	5 900	10	258 000	188 900	69 000	72,0
2014	175 700	96 700	79 000	40	83 600	76 000	7 600	10	259 200	172 700	86 600	67,8
Total	5 908 500	4 515 000	1 392 900	210 340	1 250 700	1 159 000	91 700	14 900	7 159 200	5 674 000	1 484 600	82,5

1. Les déclarants sont des immigrants couplés ayant produit au moins une déclaration de revenus depuis 1982.

Notes : RP : résident permanent; RNP : résident non permanent. Tous les chiffres sont arrondis.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

Tableau 16
Proportion de déclarants couplés par tranche d'âge à l'admission, sexe et décennie d'admission

Sexe et cohortes	Âge à l'admission						Total
	0 à 14 ans	15 à 24 ans	25 à 34 ans	35 à 44 ans	45 à 64 ans	65 ans ou plus	
	pourcentage						
Cohortes de 1980 à 1989							
Hommes	77,8	91,9	93,3	91,4	81,0	57,2	86,5
Femmes	76,1	84,3	87,4	87,7	75,4	55,2	81,2
Total	77,0	88,0	90,4	89,5	77,9	56,1	83,8
Cohortes de 1990 à 1999							
Hommes	78,5	92,9	92,8	92,1	88,3	74,2	88,5
Femmes	77,0	91,0	91,2	91,0	85,6	73,7	87,1
Total	77,7	91,9	92,0	91,6	86,8	73,9	87,8
Cohortes de 2000 à 2009							
Hommes	43,6	95,0	92,4	92,9	92,8	88,6	81,8
Femmes	42,6	94,7	93,3	93,4	92,6	87,3	83,1
Total	43,1	94,8	92,9	93,2	92,7	88,0	82,5
Cohortes de 2010 à 2014							
Hommes	4,4	82,8	93,2	90,5	86,7	82,7	71,3
Femmes	4,6	84,8	93,1	91,1	86,1	81,5	73,8
Total	4,5	83,9	93,1	90,8	86,4	82,1	72,6

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

C) Analyses précédentes

Depuis sa création, la BDIM a permis de produire plusieurs analyses. Trouvez ci-après un résumé de certaines des études de Statistique Canada fondées sur la BDIM.

Ces dernières années, plusieurs publications du *Quotidien* se sont basées sur la BDIM. Les sujets discutés ont inclus l'évolution de la répartition régionale des nouveaux immigrants au Canada, le revenu et la mobilité des immigrants, les immigrants dans les régions et les immigrants qui quittent le Canada. Ces articles sont accessibles sur le [site Web](#) de Statistique Canada.

Les séries de publications *L'emploi et le revenu en perspective (75-001-X)* et de [documents de recherche de la Direction des études analytiques](#) ont proposé plusieurs documents fondés sur la BDIM. Parmi les sujets couverts, on peut citer le revenu des immigrants qui entreprennent des études postsecondaires au Canada et les avantages en matière de revenu des immigrants admis ayant été auparavant résidents temporaires au Canada.

D) Pratiques exemplaires et conseils pour les analystes

D.1 Conseils de programmation

Cette section offre des renseignements relatifs à la programmation pour les personnes souhaitant mieux comprendre la structure de programmation utilisée pour accéder aux données des fichiers de la BDIM. Toute personne peut utiliser sa propre programmation. Il existe deux types de fichiers au sein de la BDIM : les fichiers de données annuelles de la BDIM et les données sur l'immigration (pour de plus amples détails sur les fichiers constituant la BDIM, reportez-vous à la section 3). Le nom des variables relatives à l'impôt dans la BDIM consiste en trois parties : 1) le sigle indiqué dans le dictionnaire sur les données fiscales de la BDIM, 2) le niveau d'agrégation (I ou F) et 3) l'année (l'extension de l'année à quatre chiffres existe dans la plupart des cas, mais pas tous).

Exemple : La variable de revenus d'intérêts et d'investissements au niveau individuel pour 2014 serait intitulée INVI_I2014.

Les observations contenues dans les fichiers de la BDIM sont triées en fonction d'une variable, **IMDB_id** (qui ne comprend pas d'extension d'année), permettant aux utilisateurs de maintenir un lien entre les années. L'accès aux données est rendu possible par le langage de programmation SAS. Un exemple de programme SAS permettant d'accéder aux données de la BDIM est fourni ci-dessous. Les exemples ci-dessous permettent d'effectuer la tâche suivante :

« obtenir le nombre de **bénéficiaires d'aide sociale (AS)** parmi les immigrants admis entre 2000 et 2005, vivant en Ontario entre 2010 et 2012 et pour lesquels aucun revenu ne figurait sur leur feuillet T4, par sexe et année (de 2010 à 2012) ».

Nous encourageons les chercheurs utilisant nouvellement la BDIM à passer cet exemple de programme SAS en revue. L'exemple comprend généralement trois composantes :

1. la configuration de la bibliothèque : les attributions de bibliothèque sur les deux premières lignes sont les emplacements des fichiers d'entrée (première ligne) et de sortie (deuxième ligne);
2. les étapes permettant de générer un ensemble de données de travail :
 - a. Les fichiers d'entrée étant au format SAS, ils sont accessibles au moyen d'une instruction **SET** ou **MERGE**.
 - b. Ce programme vise à obtenir le nombre de **bénéficiaires d'aide sociale (AS)** parmi les immigrants :
 - i. admis entre 2000 et 2005;
 - ii. vivant en Ontario de 2010 à 2012;
 - iii. dont le feuillet T4 n'indiquait aucun revenu;
 et génère le nombre de bénéficiaires d'aide sociale, par **sexe et année** (dans ce cas, de 2010 à 2012);
3. l'ensemble de données utilisé pour produire le nombre de bénéficiaires d'aide sociale. Cette partie, qui commence par « proc freq », produit les chiffres souhaités, tels qu'ils sont précisés dans le reste du document. À la fin du programme, quatre tableaux sont générés à partir du fichier de données créé.

Il est généralement recommandé que les programmes utilisent des variables figurant dans le FRPN plutôt que dans les fichiers de données fiscales annuelles, par souci de cohérence. L'exemple de programme utilise, par exemple, la variable de groupement par sexe **GENDER_ROLLUP**, variable figurant dans le **FRPN** (PNRF), plutôt que **SXCO_I&year**, variable figurant dans le fichier **IMDB_T1FF** annuel. Dans ce programme, seules les personnes ayant rempli une déclaration chaque année entre 2000 et 2002 sont sélectionnées.

Lors de toute programmation dans SAS, il est important de distinguer les valeurs manquantes des zéros présents dans les champs numériques. Dans SAS, la plupart des opérations mathématiques effectuées avec des valeurs manquantes produiront des valeurs manquantes. Dans la BDIM, pour les années au cours desquelles une personne est présente, les variables numériques qui ne s'appliquent pas à cette personne prennent la valeur zéro « 0 ». Par exemple, si une personne sans conjoint a déclaré un revenu en 2010, la valeur de la variable **RRSPSI2010** (contributions au REER d'un conjoint) devrait être zéro « 0 ». Si cette personne n'avait pas produit de déclaration de revenus en 2010, la valeur serait manquante.

Exemple de programme pour la BDIM

*** Exemple de programme SAS pour la BDIM**

libname source1 '\\f8prod05\cic\1.Database'; * emplacement des fichiers de la BDIM ;

libname Out '\\f8prod05\cic\3.workarea'; * répertoire de l'utilisateur ;

*** L'objectif de ce programme est d'utiliser la BDIM pour obtenir le nombre de personnes recevant des prestations d'aide sociale, vivant en Ontario, pour lesquelles aucun revenu d'emploi ne figurait sur leur feuillet T4, selon le sexe et l'année (dans ce cas, de 2010 à 2012). Les données relatives aux provinces et aux revenus proviennent des fichiers annuels de la BDIM, alors que la variable relative au sexe provient du fichier **PNRF_2014** (FRPN). ;**

* La première étape est de créer un fichier de données contenant toute l'information nécessaire à la création de nos tableaux. Ce fichier est nommé SAOnt et est enregistré dans le répertoire de sortie. L'identifiant longitudinal (IMDB_ID) est utilisé pour fusionner les ensembles de données annuels de la BDIM. ;

```
data out.SAOnt;
merge
sourcel.imdb_tlff_2010_outliers(where=(prco_i2010 = 5) in=a keep=imdb_id prco_i2010 saspyf2010 t4e__i2010)

sourcel.imdb_tlff_2011_outliers(where=(prco_i2011 = 5) in=b keep=imdb_id prco_i2011 saspyf2011 t4e__i2011)

sourcel.imdb_tlff_2012_outliers(where=(prco_i2012 = 5) in=c keep=imdb_id prco_i2012 saspyf2012 t4e__i2012)

sourcel.pnrf_2014(keep=imdb_id gender_rollup landing_year immigration_category);

by IMDB_id ;

If a and b and c and (landing_year>=2000 and landing_year<=2005);
```

* La personne doit avoir produit une déclaration de revenus pour chacune des trois années et avoir été admise entre 2000 et 2005 (population d'intérêt).;

* On crée une variable dichotomique identifiant les bénéficiaires d'aide sociale pour chaque année. Cela produit trois variables :

```
flag_sa2010, flag_sa2011 et flag_sa2012, prenant la valeur 1 ou 0.;
If (t4e__i2010=0 and saspyf2010>0) then flag_sa2010 = 1 ;
else flag_sa2010 = 0 ;
if (t4e__i2011=0 and saspyf2011>0) then flag_sa2011 = 1 ;
else flag_sa2011 = 0 ;
if (t4e__i2012=0 and saspyf2012>0) then flag_sa2012 = 1 ;
else flag_sa2012 = 0 ;
run;
```

* La procédure 'freq' de SAS permet de produire nos tableaux. On doit ensuite s'assurer que les règles relatives à la confidentialité sont respectées. ;

```
proc freq data = out.SAOnt;
tables immigration_category*flag_sa2010*flag_sa2011*flag_sa2012
gender_rollup*flag_sa2010*flag_sa2011*flag_sa2012 /missing ;
run;
* Fin de l'exemple de programmation.;
```

D.2 Création d'une cohorte

Avant de commencer l'analyse, il est essentiel de déterminer la cohorte d'intérêt. Cette cohorte peut être limitée par l'année d'admission, des variables géographiques ou toute autre variable d'intérêt (p. ex. la catégorie d'admission ou le sexe), selon les besoins du chercheur. Une cohorte unique clairement définie doit être choisie pour assurer la comparabilité. Un chercheur peut s'intéresser, par exemple, aux femmes admises en 2000 et vivant au sein d'une famille ayant reçu l'aide sociale en 2001 (tableau 17). Une question d'étude relative à cette cohorte pourrait être « Quelle portion de cette cohorte a reçu l'aide sociale au cours des deux années suivantes (2002 et 2003)? » Il convient de noter que l'Agence du revenu du Canada (ARC) exige que le conjoint recevant le revenu net le plus élevé déclare le versement de l'aide sociale. Par conséquent, il est conseillé de générer la mesure de l'aide sociale (SASPY_F), y compris pour les particuliers, à partir des données au niveau familial.

Tableau 17

Exemple : Femmes admises en 2000 ayant reçu l'aide sociale (SASPY_F) en 2001

IMDB_ID	Année d'admission	Sexe	SASPY_F2001	SASPY_F2002	SASPY_F2003
			dollars		
IM583	2000	Femme	20 500	19 000	14 000
IM145	2000	Femme	3 000	0	0
IM548	2000	Femme	11 500	13 800	0
IM798	2000	Femme	16 000	18 000	8 000
IM961	2000	Femme	10 000	0	0
IM967	2000	Femme	9 500	0	0
IM110	2000	Femme	5 000	2 000	1 000
IM125	2000	Femme	1 000	0	200

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration (BDIM).

D.3 Calcul des taux de rétention

L'un des principaux avantages de la BDIM est qu'elle comprend des variables géographiques permettant d'étudier la mobilité et la rétention. Aucune autre base de données ne présente un niveau comparable de détails annuels relatifs aux déclarants, en particulier pour des régions plus réduites. Disposer de mises à jour annuelles au niveau des provinces, des divisions de recensement, des régions métropolitaines de recensement, des agglomérations de recensement, des subdivisions de recensement et des secteurs de recensement permet une vaste gamme d'analyses.

Il est possible d'étudier les trajectoires de mobilité individuelle en attribuant simplement des indicateurs aux changements de code postal et de calculer des tendances de mobilité en étudiant les déménagements à des niveaux géographiques donnés. La mobilité au niveau des subdivisions de recensement (changements de subdivision de recensement d'une année à l'autre) et la mobilité à l'échelle provinciale (changement de province d'une année à l'autre) varie de manière significative en fonction de plusieurs caractéristiques d'immigrant, comme l'âge et la catégorie d'admission. Ces données géographiques sont dérivées du code postal (variable PSCO dans la BDIM au niveau individuel et familial). Le code postal est un code alphanumérique à six caractères définissant le lieu de livraison du courrier adressé aux clients du service postal au Canada. Voir la section 3.4.1 pour une description des variables géographiques.

Dans l'exemple ci-dessous (tableau 18), le chercheur s'intéresse à la mobilité jusqu'en 2002. Les personnes IM798, IM961, IM967 et IM110 peuvent être exclues de l'étude sur la mobilité du fait de l'absence de données (ou de fichiers).

Tableau 18

Exemple : Mobilité jusqu'en 2002 des immigrants admis en 2000

IMDB_ID	Année d'admission	Province de destination	PRCO 2000	PRCO 2001	PRCO 2002
IM583	2000	C.-B.	C.-B.	C.-B.	C.-B.
IM145	2000	Alb.	Alb.	Sask.	Sask.
IM548	2000	Alb.	Ont.	Ont.	Ont.
IM798	2000	Ont.	..	Ont.	Ont.
IM961	2000	N.-B.	N.-B.	N.-B.	..
IM967	2000	Ont.	..	Alb.	Ont.
IM110	2000	..	Qc	..	Qc

.. indisponible pour une période de référence précise

Note : PRCO : province de résidence.

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration (BDIM).

Alors que la mobilité au niveau individuel est relativement simple, le taux de rétention des immigrants dans un territoire peut être calculé de diverses manières. La façon de calculer cette rétention est une décision analytique fondée sur les besoins particuliers du chercheur. Il est relativement simple de définir le nombre de personnes restant : il s'agit du nombre de personnes produisant une déclaration de revenus dans le territoire d'intérêt à un moment donné. Il est nécessaire de prendre une décision quant à la cohorte initiale d'admission pour laquelle le chiffre de rétention est calculé (le dénominateur du taux de rétention).

Les taux provinciaux publiés dans *Le Quotidien* indiquent la proportion des immigrants déclarants résidant dans la province où ils ont été admis (définie comme la province de destination envisagée) à un moment donné. Pour une cohorte donnée (p. ex. l'année d'admission) et une année d'imposition donnée (ou un certain nombre d'années depuis l'admission), le dénominateur est le nombre de déclarants correspondant à la province d'admission sélectionnée. Le numérateur est le nombre de déclarants correspondant à la province d'admission sélectionnée qui résident également dans cette province.

En utilisant le tableau CANSIM 054-0003 pour calculer les taux de rétention de la cohorte de 2011 trois ans après l'admission, un chercheur choisirait, par exemple, toutes les provinces d'admission (p. ex. la province de destination envisagée), toutes les provinces de résidence, l'année d'admission = 2011 et l'année de référence = 2014. Ce tableau ressemblerait au suivant :

Tableau 19
Province de résidence en 2014 et province d'admission, cohorte de 2011

Province d'admission	Province de résidence					
	Total – Province de résidence	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec
	nombre d'immigrants					
Total – Province d'admission	174 740	405	330	1 365	880	31 505
Terre-Neuve-et-Labrador	515	325	0	5	0	5
Île-du-Prince-Édouard	1 245	0	265	25	10	30
Nouvelle-Écosse	1 460	10	5	1 080	10	25
Nouveau-Brunswick	1 340	0	10	35	750	55
Québec	36 275	10	10	35	15	30 200
Ontario	69 135	35	25	115	70	875
Manitoba	11 190	0	0	15	0	55
Saskatchewan	6 360	0	0	0	0	20
Alberta	21 940	10	0	20	0	95
Colombie-Britannique	25 000	5	0	30	5	140
Autre	280	0	0	0	0	0

Province d'admission	Province de résidence					
	Ontario	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Autre lieu de résidence
	nombre d'immigrants					
Total – Province d'admission	70 590	9 695	6 120	26 965	26 390	500
Terre-Neuve-et-Labrador	75	5	0	60	30	0
Île-du-Prince-Édouard	560	0	0	50	295	0
Nouvelle-Écosse	185	0	5	90	30	10
Nouveau-Brunswick	275	0	10	80	120	0
Québec	3 255	40	75	1 190	1 400	45
Ontario	63 145	275	335	2 815	1 325	115
Manitoba	645	9 170	80	825	380	10
Saskatchewan	295	45	5 370	445	165	10
Alberta	810	65	140	20 170	590	35
Colombie-Britannique	1 330	85	100	1 200	22 030	70
Autre	15	0	0	35	20	200

Source : Statistique Canada, tableau CANSIM 054-0003.

Les résultats pour la Nouvelle-Écosse donnent des renseignements à ce sujet. Au total, 1 460 personnes ont été admises en Nouvelle-Écosse en 2011 et ont produit une déclaration de revenus en 2014. Pour 1 080 de ces personnes, la Nouvelle-Écosse figurait comme province de résidence en 2014. Le taux de rétention sur trois ans pour la Nouvelle-Écosse serait 1 080/1 460, soit environ 74 %. Le tableau CANSIM fournit également des renseignements sur les migrants secondaires¹² : 1 365 personnes admises en 2011 résidaient en Nouvelle-Écosse en 2014; parmi ces dernières, 1 080 envisageaient de résider en Nouvelle-Écosse et 285 avait indiqué une autre province de destination.

12. Personnes ayant déménagé au sein du Canada après avoir atteint leur destination initiale au Canada.

La définition susmentionnée de rétention part du principe que le nombre de déclarants pour la province donnée de destination envisagée est la population totale pouvant y rester au cours d'une année (c.-à-d., si les 1 460 personnes ayant envisagé de résider en Nouvelle-Écosse y avaient produit une déclaration de revenus en 2014, la province aurait enregistré un taux de rétention de 100 %). Cette méthode ne tient pas compte du comportement de déclaration de revenus tardif ou sporadique. Alors que la population totale visée pour la Nouvelle-Écosse pour l'année d'imposition de 2014 comprend 1 460 personnes, 1 425 immigrants ayant envisagé de s'installer en Nouvelle-Écosse ont produit une déclaration de revenus pour l'année d'imposition de 2013.

Une option est une approche purement longitudinale, selon laquelle on sélectionne une seule cohorte d'admission (selon la province de destination envisagée, la province de déclaration de revenus initiale, ou les deux) et le taux de rétention est calculé comme la portion de cette cohorte déclarant toujours ses revenus dans cette province. Lorsque la province de déclaration de revenus initiale est utilisée pour définir la cohorte d'admission, il est recommandé que la première déclaration de revenus ait lieu l'année d'admission des immigrants (année d'admission = année d'imposition), afin d'exclure les personnes arrivées ailleurs et ayant ensuite migré dans cette région avant de déclarer leurs revenus pour la première fois. Il est possible d'ajouter une autre restriction, si le chercheur s'intéresse à la population dont le lieu de destination correspond au lieu géographique de la première déclaration de revenus.

Étant donné qu'une portion de chaque cohorte ne déclare pas de revenus pour l'année d'admission, il peut être nécessaire d'accroître la taille de la population pour une région en définissant la cohorte d'admission comme toute personne ayant déclaré ses revenus pour la première fois dans la région au cours des deux années suivant son admission (c.-à-d., `first_tax_year = landing_year` ou `landing_year+1`). Il n'est pas recommandé d'inclure à la « cohorte d'admission » les personnes dont la première déclaration de revenus a eu lieu plusieurs années après leur admission, car il est possible que ces personnes soient d'abord arrivées ailleurs sans toutefois avoir déclaré de revenus. Il est également suggéré d'exclure les déclarants intermittents de ces analyses, car leur lieu de résidence n'est pas connu pour les années pour lesquelles aucune donnée fiscale n'existe. La rétention calculée de cette manière indiquera une diminution graduelle du nombre; cette diminution est due aux immigrants cessant de déclarer leurs revenus, émigrant ailleurs ou décédant.

Si un chercheur s'intéresse aux migrations secondaires vers une région, il est possible de l'évaluer en supprimant les personnes de la cohorte d'admission définie du nombre total d'immigrants déclarant des revenus dans la région au moment souhaité. Toutefois, à nouveau, il est suggéré de limiter ces analyses aux personnes ayant déclaré leurs revenus pour la première fois au cours de cette période (année 0 ou année 1), afin de ne pas prendre des déclarants tardifs pour des migrants dans la région. Si la cohorte d'admission est limitée aux immigrants dont la destination correspond au lieu de la première déclaration de revenus, il est suggéré de faire une autre distinction entre les migrants secondaires ayant produit leur première déclaration de revenus ailleurs (puis dans la région d'intérêt) et les immigrants ayant produit leur première déclaration de revenus dans la région d'intérêt, puis ayant été admis dans une autre région (ou lorsque toute donnée manque quant à leur destination envisagée).

Le tableau suivant présente un exemple d'une approche longitudinale de rétention provinciale utilisant des données fictives, selon diverses définitions de la cohorte d'admission initiale.

Tableau 20

Nombre d'immigrants déclarants au sein de la population en question résidant en Colombie-Britannique et taux de rétention associé, par année depuis l'admission

Années depuis l'admission	Déclarants ayant pour la première fois déclaré des revenus en C.-B. au cours de l'année 0		Déclarants ayant pour la première fois déclaré des revenus en C.-B. au cours de l'année 0 ou 1		Déclarants ayant pour la première fois déclaré des revenus en C.-B. au cours de l'année 0 ou 1 et pour lesquels la destination envisagée était la C.-B.	
	nombre	Taux de rétention pourcentage	nombre	Taux de rétention pourcentage	nombre	Taux de rétention pourcentage
0	20 000	100	20 000	...	17 500	...
1	18 000	90	25 000	100	19 000	100
2	17 000	85	23 000	92	18 000	95
3	16 500	83	22 000	88	17 500	92

... n'ayant pas lieu de figurer

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration.

Dans l'exemple ci-dessus, il est possible de calculer la rétention en Colombie-Britannique selon trois définitions de la population; le taux de rétention sur trois ans varie alors en fonction de la définition adoptée. Il est important que toutes les personnes de l'échantillon déclarent des revenus à chaque moment d'intérêt.

Enfin, les analystes doivent procéder avec précaution lors de l'étude de divisions géographiques de faible niveau sur une longue période, car les frontières des agglomérations et des régions métropolitaines de recensement évoluent et des subdivisions de recensement sont abandonnées et ajoutées. Lorsque cela est possible, les analystes devraient exécuter le programme de Fichier de conversion des codes postaux (FCCP+) pour uniformiser les codes postaux en divisions géographiques constantes.

D.4 Calcul des trajectoires de revenu au cours du temps

Comme pour le taux de rétention, calculer l'évolution des revenus d'emploi d'une année à l'autre (ou, d'ailleurs, toute variable économique) nécessite des données consécutives. Si un chercheur, par exemple, souhaite comparer le revenu d'emploi médian de la cohorte de femmes âgées de 24 ans à 54 ans en 2000, un an après leur admission et cinq ans après leur admission (tableau 21), les enregistrements présentant des fichiers FFT1 manquants devraient être supprimés de l'analyse. La décision de supprimer ces enregistrements se baserait sur le souhait d'évaluer le revenu médian de la cohorte par rapport au revenu médian des déclarants de la cohorte.

Tableau 21

Revenu d'emploi médian de la cohorte de femmes âgées de 24 à 54 ans en 2000, 1 an et 5 ans après l'admission

IMDB_ID	Année d'admission	Âge à l'admission	Sexe	Revenu d'emploi 2001	Revenu d'emploi 2005
				dollars	
IM583	2000	34	Femme	20 500	49 000
IM145	2000	53	Femme	..	56 000
IM548	2000	29	Femme	11 500	33 800
IM798	2000	31	Femme	36 000	0
IM961	2000	42	Femme	10 000	..
IM967	2000	40	Femme
IM110	2000	35	Femme	0	59 000

.. indisponible pour une période de référence précise

Source : Statistique Canada, exemple provenant de la base de données longitudinales sur l'immigration.

Des précautions doivent être prises lors du calcul du revenu pour la « première année au Canada », car il peut ne pas représenter une année d'imposition complète. Une personne admise, par exemple, en novembre 2013 et ayant déclaré des revenus pour 2013 ne déclarerait que deux mois de revenus en 2013. Une pratique exemplaire est d'utiliser la première année de revenu complète (année d'admission +1, voir le tableau 19). Une exception concerne les déclarants précoces, ayant déclaré des revenus au Canada avant leur admission ainsi que pour l'année d'imposition, car ils sont plus susceptibles de déclarer des revenus pour une année entière.

Il est également important d'étudier le revenu au cours du temps en dollars constants. Par conséquent, il est nécessaire d'apporter des ajustements à l'aide de l'Indice des prix au Canada (IPC) (voir l'annexe D.7). Cet ajustement a été apporté dans les tableaux CANSIM de la BDIM.

D.5 Arrondissement des données

Statistique Canada s'engage à respecter la vie privée des Canadiens. Par conséquent, tous les tableaux produits à partir des fichiers IMDB_TIFF font l'objet d'un arrondissement. L'intérêt de cet arrondissement est de veiller à ce qu'aucune cellule réduite ne soit divulguée pouvant révéler des renseignements sur des personnes données ou de petits groupes de personnes. En général, les macros prélèveront un ensemble de données d'entrée non arrondies de diverses statistiques (nombres, moyennes, médianes, etc.) et créeront un ensemble de données arrondi.

Les règles d'arrondissement sont confidentielles, mais les macros d'arrondissement sont à la disposition de tout chercheur. Des documents décrivant comment utiliser ces macros sont disponibles. Ces macros s'appliquent aux tableaux de données de sortie de tous les chercheurs, à toutes les demandes de données externes et à tous les tableaux CANSIM publiés.

D.6 Détection de valeurs aberrantes

La variable de valeur aberrante OUTLIER_IND a été créée pour détecter les valeurs aberrantes au sein du fichier FFT1 (voir la section 5.5). Elle devrait être utilisée pour supprimer les données aberrantes de tout calcul (p. ex. moyenne, médiane ou régression) utilisant des données fiscales. Les valeurs aberrantes varient d'une année à l'autre; ce qui signifie que les données relatives à une personne peuvent être détectées comme valeur aberrante pour une année, mais pas pour une année ultérieure.

Le tableau suivant présente la répartition des valeurs aberrantes dans les fichiers de données fiscales à partir de 1982, par type de résident, pour la BDIM de 2014. Moins de 0,1 % des enregistrements ont été relevés comme étant des valeurs aberrantes par année d'imposition. La proportion de valeurs aberrantes a augmenté de 1995 à 1996, du fait de mises à jour de la méthode de détection des valeurs aberrantes appliquée aux fichiers de données fiscales à partir de l'année d'imposition de 1997.

Tableau 22
Répartition des déclarants avec des valeurs aberrantes par année d'imposition

	RP	RP avec permis de RNP	Total	
	nombre		pourcentage	
1982	10	10	20	0,01
1983	30	0	30	0,01
1984	40	10	50	0,01
1985	30	0	40	0,01
1986	50	10	50	0,01
1987	60	10	70	0,01
1988	70	20	90	0,01
1989	70	20	90	0,01
1990	60	20	80	0,01
1991	70	20	100	0,01
1992	150	40	180	0,01
1993	150	60	210	0,01
1994	60	20	90	0,01
1995	160	70	230	0,01
1996	340	190	530	0,03
1997	360	220	580	0,03
1998	470	250	720	0,03
1999	390	230	620	0,03
2000	480	260	740	0,03
2001	440	260	700	0,03
2002	510	260	780	0,03
2003	470	240	710	0,02
2004	600	290	890	0,03
2005	660	340	1 000	0,03
2006	610	310	920	0,03
2007	590	340	930	0,02
2008	620	380	1 000	0,03
2009	710	370	1 080	0,03
2010	720	430	1 150	0,03
2011	730	430	1 150	0,03
2012	730	420	1 150	0,02
2013	790	460	1 250	0,03
2014	920	440	1 350	0,02

Note : RP : résident permanent; RNP : résident non permanent.

Source : Statistique Canada, Base de données longitudinales sur l'immigration de 2014.

D.7 Ajustement du revenu selon l'Indice des prix à la consommation (IPC)

Afin de tenir compte du coût de la vie, il est nécessaire d'ajuster tous les revenus en fonction de l'Indice des prix à la consommation (IPC) du Canada. « L'Indice des prix à la consommation (IPC) est un indicateur de la variation des prix à la consommation payés par les Canadiens. Pour l'établir, on compare au fil du temps le coût d'un panier fixe de biens et services achetés par les consommateurs. Puisque le panier comprend des biens et services d'une quantité et d'une qualité constante ou équivalente, l'indice reflète uniquement la fluctuation pure des prix¹³. » Les facteurs de correction pour 2104 sont présentés au tableau 23. Pour convertir les données en dollars constants d'une année donnée (année de référence), les utilisateurs de données doivent multiplier les valeurs en dollars de toutes les années, à l'exception de l'année de référence, par un facteur de correction propre à une année. Pour obtenir ces facteurs de correction, les utilisateurs de données doivent diviser l'IPC de l'année de référence par l'IPC de l'année étudiée. L'année de référence dans le tableau 23 est 2014.

Tableau 23
Facteur de correction de l'Indice des prix à la consommation de 2014

Année	L'ajustement de l'Indice des prix à la consommation de 2014 est égal à 125,2 divisé par	
	nombre	
1982		54,9
1983		58,1
1984		60,6
1985		63,0
1986		65,6
1987		68,5
1988		71,2
1989		74,8
1990		78,4
1991		82,8
1992		84,0
1993		85,6
1994		85,7
1995		87,6
1996		88,9
1997		90,4
1998		91,3
1999		92,9
2000		95,4
2001		97,8
2002		100,0
2003		102,8
2004		104,7
2005		107,0
2006		109,1
2007		111,5
2008		114,1
2009		114,4
2010		116,5
2011		119,9
2012		121,7
2013		122,8
2014		125,2

Source : Statistique Canada, tableau CANSIM 326-0021.

D.8 Calcul de mesures de revenu clés

Les tableaux CANSIM de la BDIM contiennent plusieurs mesures du revenu. Le tableau 22 décrit les variables du FFT1 incluses dans ces calculs.

13. http://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=2301.

Tableau 24
Description des mesures de revenu dans la Base de données longitudinales sur l'immigration

Mesure	Composantes	Formule
Revenu d'emploi	Revenu des feuillets T4; autre revenu d'emploi	$T4E_i + OEI_i$
Revenu de travail indépendant		
Depuis 1988	Revenu provenant d'un travail indépendant comme des revenus d'entreprise, de profession libérale, de commissions, d'agriculture ou de pêche; revenu de société en commandite	$SEI_i + LTPI_i$
Avant 1988	Revenu provenant d'un travail indépendant comme des revenus d'entreprise, de profession libérale, de commissions, d'agriculture ou de pêche	SEI_i
Revenu de placements	Revenus d'intérêts et d'investissements; dividendes; gains et pertes en capital, net imposable*	$INVI_i + XDIV_i + CLKGLi^*$
1982 à 1987	Gains et pertes en capital, net imposable	$2 \times CLKGLi$
1988 et 1989	Gains et pertes en capital, net imposable	$(3/2) \times CLKGLi$
1990 à 1999	Gains et pertes en capital, net imposable	$(4/3) \times CLKGLi$
2000	Gains et pertes en capital, net imposable	$(100/64,58) \times CLKGLi$
Depuis 2000	Gains et pertes en capital, net imposable	$2 \times CLKGLi$
Indemnités d'assurance-emploi	Indemnités d'assurance-emploi	$EINS_i$
Allocations sociales	Allocations sociales (utiliser le niveau familial)	$SASPYf$
Revenu total	Somme de toutes les mesures décrites ci-dessus	

Source : Statistique Canada, Base de données longitudinales sur l'immigration, traitement de tableau CANSIM.

Il convient de noter que toutes les valeurs aberrantes sont supprimées de ces calculs ($Outlier_ind=1$), que l'on utilise la variable de province de résidence à la fin de l'année ($PRCO_$) pour déterminer la province et que tous les revenus sont corrigés en fonction de l'Indice des prix à la consommation (IPC) de l'année la plus récente pour laquelle le FFT1 est disponible. « Moyenne, avec revenu » désigne le revenu moyen des immigrants déclarants ayant un revenu du type donné. « Médiane, avec revenu » désigne le revenu médian des immigrants déclarants ayant un revenu du type donné.

Bibliographie

- A. Carpentier et G. Pinsonneault. 1994. *Representativeness Study of Immigrants Included in the Immigrant Data Bank (IMDB Project)*. Ministère des Affaires internationales, de l'Immigration et des Communautés culturelles. Gouvernement du Québec.
- A. Diaz-Papkovich, *Evaluating the quality of the Permanent Resident Linkage*. Document non publié. Ottawa, Statistique Canada, 2016a.
- A. Diaz-Papkovich, *IMDB Linkage Highlights 2016*. Document non publié. Ottawa, Statistique Canada, 2016b.
- Commission de l'immigration et du statut de réfugié du Canada, *Section de la protection des réfugiés*, 2015. Disponible à l'adresse <http://www.irb-cisr.gc.ca/Fra/RefClaDem/Pages/RpdSpr.aspx> (site consulté le 13 janvier 2016).
- C. Langlois et C. Dougherty, *La Base de données longitudinale sur l'immigration (BDIM): Une introduction*. Recueil de la Conférence 1997 sur l'immigration, l'emploi et l'économie de Citoyenneté et Immigration Canada, 1997.
- Dryburgh, H., *La Banque de données administratives longitudinales (DAL) et la Banque de données longitudinales sur les immigrants (BDIM) : création de la DAL_BDIM - un document technique* (n° 89-612-XIE au catalogue) Ottawa, Statistique Canada, 2004. Disponible à l'adresse <http://publications.gc.ca/Collection/Statcan/89-612-X/89-612-XIE2003001.pdf> <http://publications.gc.ca/collections/Collection/Statcan/89-612-X/89-612-XIF2003001.pdf> (site consulté le 13 mars 2017).
- Gouvernement du Canada, *Déterminer son admissibilité – Citoyenneté*, 2016. Disponible à l'adresse <http://www.cic.gc.ca/francais/citoyennete/devenir-admissibilite.asp> (site consulté le 20 juillet 2016).
- Immigration, Réfugiés et Citoyenneté Canada, *Rapport annuel au Parlement sur l'immigration*, 2015. Disponible à l'adresse <http://www.cic.gc.ca/francais/ressources/publications/rapport-annuel-2015/index.asp> (site consulté le 10 juin 2016).
- Immigration, Réfugiés et Citoyenneté Canada, *Rapport sur les plans et priorités de 2015-2016*, 2016. Disponible à l'adresse <http://www.cic.gc.ca/francais/ressources/publications/rpp/2015-2016/index.asp> (site consulté le 10 juin 2016).
- J. Badets, et C. Langlois. 2000. « Les défis de l'utilisation de données administratives à l'appui de la recherche en matière de politiques: l'exemple de la base de données longitudinale sur l'immigration (IMDB) » Au *Symposium 99 - Combiner des données de sources différentes*, 1999. Symposium international de Statistique Canada: Recueil, produit n° 11-522-XPE au catalogue de Statistique Canada.
- M. Rotermann, C. Sanmartin, R. Trudeau, and H. St-Jean, « Couplage des données du Recensement de 2006 et des données hospitalières au Canada » *Rapports sur la santé*. Vol. 26, n° 10 (n° 82-003-X au catalogue) Ottawa, Statistique Canada, 2015. Disponible à l'adresse <http://www.statcan.gc.ca/pub/82-003-x/2015010/article/14228-fra.pdf> (site consulté le 13 mars 2017).
- S.B. Dusetzina, S. Tyree, A.M. Meyer, A. Meyer, L. Green et W.R. Carpenter, *Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]*. Préparé par l'Université de la Caroline du nord à Chapel Hill sous le contrat n° 290-2010-000141. AHRQ Publication n° 14-EHC033-EF. Rockville, MD: Agency for Healthcare Research and Quality, 2014 Disponible à l'adresse <http://www.ncbi.nlm.nih.gov/books/NBK253312/> (site consulté le 14 mars 2017).
- S. McLeish, 2008 *IMDB Landing File: Data Quality Working Paper*. Document non publié. Ottawa, Statistique Canada, 2011.

Statistique Canada, 150 ans d'immigration au Canada. Mégatendances canadiennes (n° 11-630-X au catalogue) Ottawa, Statistique Canada, 2016. Disponible à l'adresse <http://www.statcan.gc.ca/pub/11-630-x/11-630-x2016006-fra.htm> (site consulté le 14 mars 2017).

W.E. Winkler, « Record linkage » *Sample Surveys: Design, Methods and Applications* 29A, 2009, p. 351–380.