

## Comparaison des données d'enquête et des sources administratives : données sur l'immigration et le travail et données démographiques de l'Étude longitudinale et internationale des adultes

James Hemeon, Meghan Fulford<sup>1</sup>

### Résumé

Les données administratives, selon leur source et leur objectif original, peuvent être considérées comme plus fiables au chapitre de l'information que les données recueillies dans le cadre d'une enquête. Elles ne nécessitent pas qu'un répondant soit présent et comprenne le libellé des questions, et elles ne sont pas limitées par la capacité du répondant à se rappeler des événements rétrospectivement. La présente communication compare certaines données d'enquête, comme des variables démographiques, tirées de l'Étude longitudinale et internationale des adultes (ELIA), à diverses sources administratives pour lesquelles des ententes de couplage avec l'ELIA sont en place. La correspondance entre les sources de données, et certains des facteurs qui pourraient l'affecter, sont analysés pour divers aspects de l'enquête.

Mots-clés : données administratives, longitudinales, enquête, couplage d'enregistrements, couplage de données

### 1. Introduction

Les organismes gouvernementaux recueillent et conservent des données à des fins administratives, comme l'imposition de la population ou des entreprises, ou la régulation du flux transfrontalier de biens et de personnes. Par suite de la demande croissante de données statistiques, ainsi que du fardeau de réponse et des coûts associés à la collecte, les sources administratives sont devenues des sources de données de plus en plus recherchées à des fins statistiques.

Le couplage des données d'enquête à des sources de données administratives représente une façon de recueillir des données sur des sujets délicats qu'une personne n'est peut-être pas à l'aise d'aborder avec un intervieweur, des données dont une personne peut avoir de la difficulté à se rappeler précisément, et/ou des données qui donneraient lieu à une interview très longue dans le cadre d'une enquête. Toutefois, lorsque l'on utilise les données administratives à des fins statistiques, on doit comprendre l'objectif original, la portée, les définitions de variables et les limites.

L'Étude longitudinale et internationale des adultes (ELIA) est une enquête sociale longitudinale canadienne menée par Statistique Canada, qui recueille une gamme variée de données socioéconomiques. L'ELIA comprend aussi des données de plusieurs sources administratives pour chaque année, depuis son lancement en 2011, ainsi que des données administratives historiques remontant aussi loin que 1980, et jusqu'en 2013, pour certains répondants. En raison de la variété et de la quantité de données, il existe quelques variables pour lesquelles l'ELIA utilise à la fois des données d'enquête et des données de sources administratives. Lorsqu'il existe des écarts entre les sources, l'ELIA ne corrige pas les données selon l'hypothèse qu'une des sources est correcte.

Dans le présent document, on comparera trois variables socioéconomiques qui existent à la fois dans les données de l'ELIA et dans les données administratives couplées à l'ELIA : année d'immigration (établissement), secteur d'emploi

---

<sup>1</sup> James Hemeon, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6, [James.Hemeon@Canada.ca](mailto:James.Hemeon@Canada.ca). Meghan Fulford, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario) Canada, K1A 0T6, [Meghan.Fulford@Canada.ca](mailto:Meghan.Fulford@Canada.ca).

et année du mariage. On mesurera la cohérence entre les sources et on tentera d'expliquer certaines des différences, afin de mieux comprendre les forces et les limites des sources, ainsi que la façon dont elles se complètent.

## **2. Source des données**

L'ELIA comprend des données démographiques sur les répondants, ainsi que des données sur la composition de la famille et du ménage, la prestation de soins, l'éducation et la formation, les compétences, la santé, le revenu et la richesse, la retraite et l'activité sur le marché du travail. L'échantillon a été sélectionné à partir d'une base de population constituée des logements des dix provinces au Canada qui ont répondu au Recensement de la population de 2011, et exclut les membres réguliers des Forces canadiennes, les personnes vivant en établissement et les personnes vivant dans des réserves et d'autres établissements autochtones.

L'ELIA compte actuellement deux bases de données d'enquête, appelées ELIA 2012 et ELIA 2014. L'ELIA 2012 comprend 23 926 répondants âgés de 15 ans et plus, dont 10 299 (43 %) ont été interviewés en 2011, et 13 627 (57 %) en 2012.

L'ELIA 2014 comprend 19 178 répondants âgés de 15 ans et plus, dont 16 895 personnes qui ont été interviewées dans le cadre de l'ELIA 2012, 1 582 personnes qui faisaient partie de l'échantillon de l'ELIA 2012 et qui n'ont pas répondu, et 701 personnes qui se sont jointes au ménage d'un répondant de l'ELIA 2012 entre l'ELIA 2012 et l'ELIA 2014. Les interviews de l'ELIA 2014 ont eu lieu en 2014.

Au total, l'ELIA a obtenu au moins une réponse d'enquête pour 26 209 répondants âgés de 15 ans et plus.

## **3. Couplage**

Des couplages de données ont été effectués entre l'ELIA et les bases de données administratives suivantes : le fichier sur la famille T1 (FFT1), l'état et le sommaire de rémunération payée par les employeurs (fichier T4), le Registre des entreprises (RE), la Base de données sur les régimes de pension au Canada (RPAC) et la Base de données longitudinales sur les immigrants (BDIM).

Pour permettre le couplage avec ces bases de données, l'ELIA est couplée aux fichiers maître de l'impôt sur le revenu des particuliers T1, au moyen de méthodes de couplage déterministe et probabiliste, afin d'obtenir un numéro d'assurance sociale (NAS). Dans l'ELIA 2012, les NAS ont été couplés au moyen des fichiers maîtres T1 de 2011 et 2010, qui comprennent des données pour tous les déclarants fiscaux au Canada pour ces années respectives. Dans l'ELIA 2014, les NAS ont été couplés au moyen des fichiers maîtres T1 de 2013 et 2012. Un couplage a été effectué entre l'ELIA et un NAS (et des données administratives) pour 24 763 (94,5 %) des 26 209 répondants au total, y compris 23 081 répondants (96,5 %) à l'ELIA 2012, et 18 209 répondants (95,0 %) à l'ELIA 2014. Le taux de couplage des vagues précédentes s'est amélioré au fil du temps, le couplage avec le NAS étant effectué pour les répondants existants et les nouveaux répondants.

Une fois qu'un NAS est déterminé, il sert au couplage avec le FFT1, le fichier T4 et le fichier de la BDIM. Les données T4 servent ensuite à coupler le RE et la RPAC. Les données du FFT1 sont disponibles de 1982 à 2013, et les données T4, du RE et de la RPAC sont disponibles de 2000 à 2013. Les données de la BDIM sont des données longitudinales allant de 1980 à 2012.

Les échantillons utilisés pour le présent document varient selon l'analyse. L'ELIA 2012 et l'ELIA 2014 servent à comparer l'année d'immigration (par rapport à la BDIM), l'ELIA 2012 sert à comparer le secteur d'emploi (comparaison des fichiers T4/du RE de 2011 et 2012), et l'ELIA 2014 sert à comparer l'année du mariage actuel (comparaison avec le FFT1 de 1982 à 2013).

## 4. Résultats

### 4.1 Année d'immigration : ELIA et BDIM

La BDIM est une base de données qui combine des données couplées de l'immigration et des dossiers fiscaux (tirés principalement du FFT1). La base de données comprend les personnes qui ont obtenu le statut d'immigrant depuis 1980 jusqu'en 2012 (au moment du couplage avec l'ELIA 2014), et qui ont produit au moins une déclaration de revenu depuis 1982.

Parmi les 26 209 répondants de l'échantillon combiné de l'ELIA 2012 et l'ELIA 2014, 4 864 (18,6 %) ont déclaré être un immigrant reçu et ont fourni une année d'immigration. Parmi eux, 3 276 (67,4 %) ont été couplés avec succès au fichier de la BDIM. Au total, 1 258 répondants (25,9 %) n'ont pas pu être couplés à la BDIM, mais avaient déclaré une année d'immigration antérieure à 1980 ou postérieure à 2012, et ne faisaient par conséquent pas partie du champ de la BDIM. Les 330 qui restent (6,8 %) n'ont pas pu être couplés et ont déclaré une année d'établissement entre 1980 et 2012. Parmi ceux qui ont déclaré avoir immigré entre 1980 et 2012, 90,8 % ont été couplés à la BDIM.

Tous les répondants de l'ELIA ont été couplés à la BDIM, peu importe le statut d'immigrant déclaré dans l'enquête. Le fichier de la BDIM et de l'ELIA couplées comprend 3 446 répondants. Au total, 3 278 répondants (95,1 %) couplés ont déclaré une année d'établissement pendant l'interview. Les 170 qui restent (4,9 %) n'ont pas déclaré être un immigrant reçu dans le cadre de l'enquête.

Parmi les raisons pour lesquelles certains répondants figurent comme immigrants reçus dans une source de données, mais pas dans l'autre, figurent les erreurs de réponse au moment de la collecte, ou les erreurs de traitement dans le couplage des données. Par ailleurs, un répondant qui a une année d'immigration antérieure à 1980 ou postérieure à 2012, ou qui n'a pas produit au moins une déclaration de revenu depuis 1982, ne figurera pas dans la BDIM.

Après avoir pris en compte les différences entre les sources de données, nous examinerons maintenant la cohérence. La majorité (77,9 %) des gens qui ont déclaré une année d'établissement entre 1980 et 2012 dans l'ELIA et qui ont été couplés à la BDIM ont déclaré une année qui est égale à leur année d'établissement dans la BDIM, et 88,9 % ont déclaré une année se situant à plus ou moins un an de leur année d'établissement dans la BDIM (voir le tableau 4.1.1).

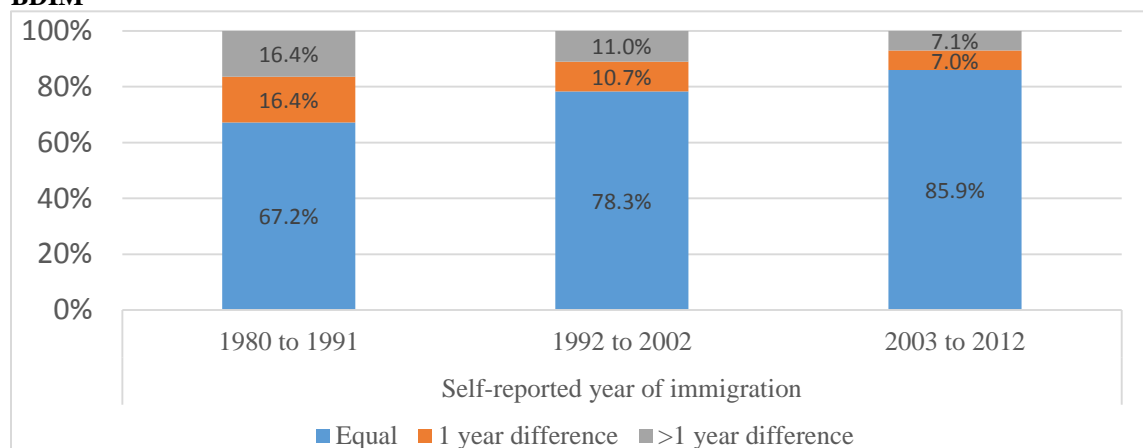
**Tableau 4.1.1 : Cohérence entre l'année d'immigration de l'ELIA et l'année d'immigration de la BDIM**

Cohérence	%	% cumulatif
Égale	77,9 %	77,9 %
Différence d'un an	11,0 %	88,9 %
Différence de 2 à 5 ans	9,6 %	98,4 %
Différence de >5 ans	1,6 %	100,0 %

**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2012 et 2014; Statistique Canada, Base de données longitudinales sur les immigrants, 1980-2012.

La différence entre les sources quant à l'année d'établissement peut être le résultat d'un biais de remémoration. Même si les répondants peuvent avoir de la difficulté à se rappeler avec précision d'événements qui remontent à loin, les données administratives ne comportent pas cette limite. Lorsque l'on compare l'année d'établissement de l'ELIA et celle de la BDIM, on note une plus grande cohérence entre les sources pour les personnes qui ont déclaré avoir immigré plus récemment (voir le tableau 4.1.2). Parmi celles qui ont immigré entre 2003 et 2012, 85,9 % ont déclaré une année égale à l'année figurant dans la BDIM. Si l'on remonte dans le temps, le pourcentage ayant déclaré une année égale à l'année figurant dans la BDIM diminue. Lorsque l'année d'établissement se situe entre 1980 et 1991, le pourcentage déclarant une année égale à celle de la BDIM diminue pour passer à 67,2 %.

**Figure 4.1.2 : Comparaison entre les sources : année d'immigration de l'ELIA et année d'immigration de la BDIM**



**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2012 et 2014; Statistique Canada, Base de données longitudinales sur les immigrants, 1980-2012.

Cela laisse supposer que la source administrative n'est pas sujette à un biais de remémoration qui pourrait exister pour les répondants qui, dans leur interview, ont tenté de se rappeler un événement remontant aussi loin que 1980.

#### 4.2 Secteur d'emploi : ELIA 2012 et Registre des entreprises (2011-2012)

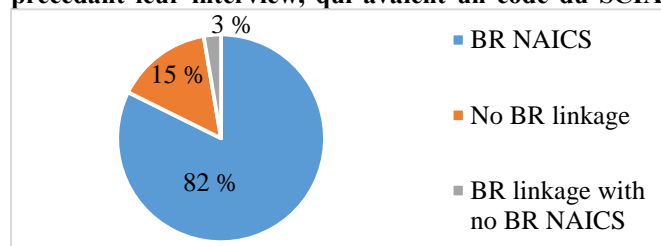
Parmi les 23 926 répondants à l'ELIA 2012, 19 582 (81,8 %) ont reçu un code du Système de classification des industries de l'Amérique du Nord (SCIAN) attribué à partir des données de l'ELIA. Parmi eux, 15 168 (77,5 %) ont aussi un code du SCIAN du Registre des entreprises (RE), qui est le résultat du couplage avec le T4 comportant la rémunération la plus élevée pour l'année fiscale correspondant à leur année d'interview de l'ELIA 2012 (2011 ou 2012). Au total, 504 répondants ayant un code du SCIAN de l'ELIA (2,6 %) ont été couplés à une observation du RE comportant un code du SCIAN en blanc, ce qui peut se produire si l'entreprise associée au T4 du répondant n'a pas fourni une description suffisante pour qu'un SCIAN puisse être attribué. Au total, 3 936 répondants ayant un code du SCIAN de l'ELIA (20,1 %) n'ont pas été couplés au RE. Toutefois, on doit tenir compte du fait que, dans l'ELIA 2012, les données sur le secteur d'emploi ont été recueillies auprès des répondants qui ont déclaré avoir travaillé au cours des cinq ans précédant l'interview, les répondants n'ayant pas travaillé l'année de leur interview étant moins susceptibles d'avoir reçu un T4 pour cette année-là.

Au total, 4 344 répondants (18,1 %) n'ont pas de code du SCIAN dans l'ELIA; 429 (9,9 %) répondants n'ayant pas de code du SCIAN dans l'ELIA ont un code du SCIAN dans le RE, et 345 (80,4 %) d'entre eux ont indiqué ne pas avoir travaillé au cours des 12 mois précédant leur interview (dont 341 au cours des cinq ans précédant l'interview). Il existe par conséquent un écart entre les sources : les répondants ont indiqué ne pas avoir travaillé, mais ont été couplés à un T4.

Les 3 915 répondants qui restent n'ayant pas de code du SCIAN de l'ELIA (90,1 %) n'ont pas non plus de code du SCIAN du RE. Parmi eux, 3 862 (98,6 %) ont indiqué ne pas avoir travaillé au cours des 12 mois précédant leur interview (dont 3 834 cours des cinq années précédant l'interview). Cela est cohérent, les personnes n'ayant pas travaillé étant moins susceptibles de recevoir un T4.

Si l'on tient compte uniquement des 18 142 répondants qui ont déclaré avoir travaillé dans un emploi rémunéré au cours des 12 mois précédant leur interview, 18 005 (99,2 %) ont reçu un code du SCIAN de l'ELIA. Parmi ceux ayant un code du SCIAN de l'ELIA, 14 822 (82,3 %) avaient aussi un code du SCIAN du RE, 2 699 (14,9 %) n'étaient pas couplés au RE, et 484 (2,7 %) étaient couplés à une observation du RE sans code du SCIAN (voir la figure 4.2.1). Parmi les 14 906 répondants ayant un code du SCIAN du RE, 14 822 (99,4 %) avaient aussi un code du SCIAN de l'ELIA.

**Figure 4.2.1 Proportion des répondants de l'ELIA 2012 ayant travaillé au cours de la période de 12 mois précédant leur interview, qui avaient un code du SCIAN de l'ELIA, avec/sans un code du SCIAN du RE**



**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2012; Statistique Canada, Registre des entreprises, 2011-2012

Le code du SCIAN tiré des données de l'ELIA est un code à 4 chiffres, et le code du SCIAN tiré du RE est un code à 6 chiffres, qui fournit des détails additionnels sur le secteur d'emploi. Pour comparer les sources, on a mesuré la cohérence au niveau à 4 chiffres pour les 14 822 répondants qui avaient à la fois un code du SCIAN de l'ELIA et du RE. Au moment de la comparaison des deux sources, 43 % avait le même code du SCIAN au niveau à 4 chiffres, 57 %, au niveau à 3 chiffres, et 64 %, au niveau le plus élevé (2 chiffres) (voir le tableau 4.2.2). Cela peut probablement être expliqué par les différences entre les sources de données et le codage : le code du SCIAN de l'ELIA est donné par Statistique Canada à partir des renseignements fournis par le répondant, au moyen du même processus que celui utilisé pour d'autres enquêtes de Statistique Canada. Le code du SCIAN du RE, quant à lui, est attribué par la Sous-division du Registre des entreprises de Statistique Canada, à partir de plusieurs sources de données, y compris la description fournie par l'entreprise au moment de son enregistrement auprès de l'Agence du revenu du Canada, et les déclarations de revenu de l'entreprise (en date de 2011). Le code du SCIAN est aussi mis à jour par les utilisateurs du Registre des entreprises (responsables des profils d'entreprise et spécialistes des enquêtes auprès des entreprises) à Statistique Canada, ainsi qu'à partir de la rétroaction obtenue dans le cadre des enquêtes auprès des entreprises. Outre ces différences, il se peut que la description de son employeur fournie par un répondant de l'ELIA (employé) diffère de celle fournie par le propriétaire de l'entreprise (au moment de la constitution de l'entreprise en société). Par ailleurs, il arrive parfois que les entreprises modifient leur activité principale (SCIAN) au fil du temps et que ces changements ne soient pas pris en compte au moment des couplages.

Outre les sources de données et le codage, il existe des différences conceptuelles mineures entre les sources. Le code du SCIAN de l'ELIA est tiré de la description fournie par le répondant de son emploi actuel; toutefois, si un répondant occupait plus d'un emploi, l'emploi pour lequel il a travaillé le plus grand nombre d'heures a servi à attribuer le code du SCIAN. Si un répondant a travaillé à plus d'un emploi, mais le même nombre d'heures dans tous les emplois, on a utilisé l'emploi comportant la rémunération la plus élevée. Par conséquent, certains répondants peuvent avoir reçu un code du SCIAN de l'ELIA pour un emploi dans lequel ils ont travaillé le plus grand nombre d'heures, mais avoir occupé un autre emploi comportant la rémunération la plus forte. Dans ces cas, si l'on comparait l'ELIA et le T4 correspondant à la rémunération la plus forte, on ne comparerait pas les codes du SCIAN du même emploi.

**Tableau 4.2.2. Cohérence entre le code du SCIAN à 4 chiffres de l'ELIA et le code du SCIAN du RE**

Cohérence	%	% cumulatif
4 chiffres	43 %	43 %
3 chiffres	14 %	57 %
2 chiffres	7 %	64 %
Aucun appariement	36 %	100 %

**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2012; Statistique Canada, Registre des entreprises, 2011-2012

### 4.3 Année du mariage : ELIA 2014 et FFT1 (1982-2013)

Dans le cas des répondants qui ont déclaré être « mariés » ou « séparés » au moment de leur interview, l'ELIA recueille l'année pendant laquelle le mariage actuel du répondant a commencé. L'année du mariage actuel peut aussi être calculée au moyen de la variable de l'état matrimonial dans le FFT1, de 1982 à 2011. En fusionnant l'état matrimonial de chaque année du FFT1 en un ensemble de données, un historique de l'état matrimonial peut être créé. Pour la présente analyse, on a utilisé les personnes ayant déclaré être « mariées » dans l'ELIA.

Lorsque l'on compare les données sur le mariage actuel de l'ELIA aux données traitées du FFT1 (que l'on appellera « données brutes »), pour 6 371 répondants de l'ELIA 2014 qui ont déclaré être mariés après 1982 dans les données de l'ELIA, 63,9 % avait la même année de mariage dérivée du FFT1. Pour les 3 484 répondants de l'ELIA 2014 qui ont déclaré être mariés en 1982 ou avant dans les données de l'ELIA, 60,2 % avaient une année de mariage correspondant à « 1982 ou avant » tirées des données du FFT1. La série de données du FFT1 commence en 1982; par conséquent, si un répondant est marié au moment de la production de sa déclaration de revenu de 1982, il sera considéré comme ayant été marié en 1982 ou toute année auparavant.

Parmi les considérations figure le fait que les données du FFT1 sont utilisées à des fins différentes de celles prévues, les données fiscales servant à calculer l'état matrimonial. Certaines limites de la variable de l'état matrimonial dans les données du FFT1 ont été notées, ce qui laisse supposer qu'avec une « épuration » raisonnable de l'historique du mariage du FFT1, il y aurait davantage de cohérence avec les données de l'ELIA. Tout d'abord, pour certaines années, des données étaient manquantes, et des valeurs de l'état matrimonial étaient improbables ou impossibles. Par exemple, il y avait des anomalies pour une année, les personnes qui étaient « mariées » étant devenues « vivant en union libre » ou « séparées » pour une année (ce qui est possible, mais peu probable), étant devenues « célibataires, jamais mariées », (ce qui est impossible) pour une année, ou ayant une année de données manquantes dans les FFT1, mais étant revenues au statut de « mariées ». Ces valeurs ont été modifiées selon la valeur qui précédait/suivait (voir l'« étape 1 » dans le tableau 4.3.1 et le tableau 4.3.2).

Il y avait aussi des répondants comptant plusieurs années de données manquantes dans le FFT1. Par exemple, une série de données du FFT1 d'un répondant pouvait commencer en 1996, date à partir de laquelle il avait déclaré être « marié ». Comme les données antérieures étaient manquantes, on ne peut pas présumer qu'il s'agissait de l'année de mariage. Afin de permettre une comparaison plus utile avec les données de l'ELIA, nous avons procédé aux étapes suivantes. Tout d'abord, la répartition de l'âge du mariage actuel a été analysée au moyen des données existantes de l'ELIA et du FFT1, et le mode (23 ans) a été choisi. Les données manquantes, lorsqu'elles précèdent la production d'une déclaration de revenu dans laquelle le répondant est « marié », ont par conséquent été imputées rétrospectivement à un répondant âgé de 23 ans (voir l'« étape 2 » dans le tableau 4.3.1 et le tableau 4.3.2).

Enfin, les valeurs pour « vivant en union libre », « célibataire, jamais mariée » ont été modifiées à « mariée » si elles suivaient immédiatement une année pendant laquelle une personne était « mariée » (voir l'« étape 3 » du tableau 4.3.1 et du tableau 4.3.2). Par exemple, si une personne est devenue « célibataire, jamais mariée » pendant plusieurs années suivant immédiatement une année pendant laquelle elle déclarait être « mariée », ces valeurs « célibataire, jamais mariée » ont été modifiées à « mariée ».

Dans le cas des mariages autodéclarés postérieurs à 1982 (6 230 répondants), les étapes 1 à 3 ont augmenté la cohérence entre l'ELIA et l'année du mariage actuel du FFT1, la proportion étant passée de 62,5 % (« données brutes ») ayant la même année dans les deux sources, à 68,0 %, 69,1 % et 69,5 %, respectivement (voir le tableau 4.3.1). Dans le cas des mariages autodéclarés en 1982 et avant (3 450 répondants), les étapes 1 à 3 augmentent la capacité de prédire, à partir des données du FFT1, si le mariage s'est produit en 1982 ou avant, la proportion passaient de 57,3 % (« données brutes ») à 70,4 %, 90,4 % et 91,0 % respectivement (voir le tableau 4.3.2). Au total, 175 répondants qui ont déclaré être mariés n'avaient pas de données fiscales et n'ont par conséquent pas pu être analysés.

**Tableau 4.3.1. Cohérence entre l'année du mariage de l'ELIA et celle dérivée du FFT1, pour les personnes qui ont déclaré un mariage postérieur à 1982**

FFT1	Données brutes	Étape 1	Étape 2	Étape 3
Égale	62,5 %	68,0 %	69,1 %	69,5 %
Différence de 1 an	8,3 %	8,9 %	9,4 %	9,5 %
Différence de 2 ans	3,6 %	3,2 %	4,3 %	4,3 %
Différence de 3 à 5 ans	6,5 %	5,2 %	6,0 %	6,0 %
Différence de 6 à 10 ans	7,5 %	6,0 %	4,5 %	4,3 %
Différence de 10 ans et plus	11,2 %	8,2 %	3,5 %	3,3 %
FFT1 de 1982 ou avant	0,4 %	0,5 %	3,2 %	3,2 %

**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2014; Statistique Canada, Fichier sur la famille T1, 1982-2013

**Tableau 4.3.2. Cohérence entre l'année de mariage de l'ELIA et celle dérivée du FFT1, pour les personnes qui ont déclaré un mariage en 1982 ou avant**

FFT1	Données brutes	Étape 1	Étape 2	Étape 3
1982 ou avant	57,3 %	70,4 %	90,4 %	91,0 %
FFT1 postérieur à 1982	42,7 %	29,6 %	9,6 %	9,0 %

**Source :** Statistique Canada, Étude longitudinale et internationale des adultes, 2014; Statistique Canada, Fichier sur la famille T1, 1982-2013

## 5. Conclusion

Il n'est pas courant pour les programmes d'enquête de disposer des mêmes données tirées d'une enquête et d'une source administrative. Le présent document est le premier à effectuer une telle analyse concernant les données de l'ELIA, analyse qui montre que les renseignements obtenus à partir des dossiers administratifs sont habituellement, mais pas toujours, les mêmes que ceux fournis par les répondants d'enquête.

L'ELIA comporte des données d'enquête pour les années qui ne figurent pas dans les sources administratives couplées, et ses concepts et définitions peuvent être contrôlés par une équipe d'enquête spécialisée. Les données administratives, par définition, n'ont pas été recueillies à des fins statistiques, et on doit accorder une attention particulière pour s'assurer que les concepts sont bien compris avant de remplacer une source par une autre.

L'année d'immigration, lorsque l'on compare les données d'enquête et les données administratives, était la même pour de nombreux répondants. Pour cette comparaison, les concepts étaient les mêmes, même si les années de couverture étaient limitées dans le cas de la source administrative. Pour ce qui est du secteur d'emploi (SCIAN), les données d'enquête et les données administratives étaient les mêmes dans certains cas; toutefois, de nombreuses comportaient un code différent du SCIAN dans chaque source, ce qui peut être attribuable aux différences de concept et de codage. Dans le cas de l'année du mariage actuel, les données d'enquête et les données administratives étaient assez différentes, ces dernières ayant servi à calculer l'année de mariage à partir d'une série de données fiscales. Toutefois, lorsque l'on a fait certaines hypothèses pour combler les années manquantes de données, ainsi que pour « épurer » les valeurs impossibles ou improbables, la cohérence a été grandement améliorée.

Lorsque l'on a analysé des concepts comparables d'années d'immigration entre les sources, la cohérence entre les données d'enquête et les données administratives était plus grande pour les événements s'étant produits plus

récemment. Cela laisse supposer que les données administratives peuvent constituer une source plus précise de données pour les événements qui se sont produits il y a longtemps, à condition que les concepts et la qualité des données administratives demeurent constants au fil du temps.

En comparant les deux sources, on peut tirer parti de leurs forces, et les sources peuvent se compléter afin de fournir des données plus complètes que dans l'une ou l'autre source séparément.