



---

## **Normes et lignes directrices relatives à la confidentialité et à la qualité des données (version publique)**

Diffusion du Recensement de 2011

## Table des matières

	Page
1. Introduction .....	4
2. Règles relatives à la confidentialité (non-divulgateion).....	4
2.1 Suppression des régions pour les régions géographiques normalisées et non normalisées .	4
2.2 Règles d'agrégation minimales des codes postaux .....	5
2.3 Arrondissement aléatoire.....	6
2.4 Prévention de la divulgation pour les statistiques .....	7
2.4.1 Suppression de statistiques .....	7
2.4.2 Calculs statistiques spéciaux .....	7
3. Pratiques relatives à la confidentialité.....	8
3.1 Ajustement des chiffres de population et des logements pour des raisons de confidentialité	8
3.2 Prévention de la divulgation .....	9
3.3 Totalisations du Recensement de l'agriculture.....	9
4. Pratiques relatives à la qualité des données.....	10
4.1 Mesures de la qualité des données .....	10
4.1.1 Indicateurs de la qualité des données .....	10
4.1.1.1 Les réserves indiennes et les établissements indiens incomplètement dénombrés.....	10
4.1.1.2 Taux de réponse globaux .....	11
4.1.1.3 Indicateur d'erreur dans les chiffres de population et des logements.....	11
4.1.1.4 Indicateur de qualité des données sans objet (indicateur relatif à la qualité des données-échantillon [20 %]) .....	11
4.1.1.5 Indicateur d'ajustement du chiffre de population de 2006.....	11
4.2 Autres méthodes de suppression relatives à la qualité des données.....	14
4.3 Calcul de statistiques d'ordre .....	14
4.4 Règle sur la qualité des données pour la diffusion des données relatives à la population de 100 ans et plus.....	14

4.5 Règles relatives à la qualité des données pour la diffusion des données sur les couples de même sexe et de sexe opposé.....	15
5. Suppression des données – Autre.....	16
5.1 Fréquence de déclaration.....	16
5.2 Suppression des zéros .....	16
6. Meilleures pratiques .....	17
6.1 Signes conventionnels relatifs à la qualité et à la confidentialité dans les tableaux de données.....	17

## 1. Introduction

Les données du recensement diffusées font l'objet de divers processus automatisés et manuels visant à déterminer s'il y a lieu de les supprimer. Ces processus sont mis en œuvre principalement pour deux raisons : 1) pour garantir la non-divulgence de l'identité et des caractéristiques des répondants (notion représentée par le terme **confidentialité** dans la suite du présent document) et 2) pour restreindre la diffusion de données de qualité inacceptable (notion représentée par l'expression **qualité des données** dans la suite du présent document).

En outre, des règles de suppression de données peuvent être appliquées pour des motifs propres à des produits en particulier, habituellement liés au formatage. Le terme **produit** renvoie en général à des tableaux. Les données peuvent être soit modifiées dans le produit, soit supprimées complètement conformément aux règles de suppression appliquées. Le présent document résume les normes et lignes directrices relatives à la confidentialité et à la qualité des données devant être appliquées dans le cadre du Projet de diffusion du Recensement de 2011.

Les règles/directives suivantes (nouvelles ou mises à jour) ont été adoptées pour 2011 :

- Ajustement des chiffres de population et des logements au niveau de la circonscription électorale fédérale (consultez la section 3.1);
- Mise à jour à la règle en matière de diffusion des données relatives à la population âgée de 100 ans et plus (consultez la section 4.4);
- Adoption des signes conventionnels dans les tableaux de Statistique Canada pour les produits du Recensement de 2011 (consultez la section 6.1).

## 2. Règles relatives à la confidentialité (non-divulgence)

La présente section décrit les diverses règles utilisées pour garantir la confidentialité (ou non-divulgence) de l'identité et des caractéristiques des répondants. Toutes les données du recensement sont assujetties à des règles (de non-divulgence).

### 2.1 Suppression des régions pour les régions géographiques normalisées<sup>1</sup> et non normalisées

La suppression des régions a pour objet d'éliminer toutes les données sur les caractéristiques pour les régions géographiques dont la population est inférieure à un seuil donné.

---

<sup>1</sup>. Pour obtenir de plus amples renseignements sur les régions géographiques normalisées, consultez le Dictionnaire du recensement.

Le seuil de population défini pour toutes les régions normalisées<sup>1</sup> ou les regroupements de régions normalisées, sauf les îlots, les côtés d'îlot et les régions définies selon le code postal, est de 40. Par conséquent, aucune caractéristique ou donnée totalisée n'est diffusée pour les régions comptant moins de 40 habitants.

Dans le cas des régions définies selon le code postal à six caractères (région de tri d'acheminement – unité de distribution locale [RTA-UDL]), des régions géocodées et des régions normalisées constituées d'îlots, de côtés d'îlot ou d'UDL, le seuil de population est établi à 100 personnes. Par conséquent, aucune caractéristique ou donnée totalisée ne doit être diffusée si la population totale de la région est inférieure à 100 habitants. En général, la population des îlots et des côtés d'îlot urbains (un côté d'une rue, situé entre deux intersections consécutives) est trop faible pour respecter les seuils de population établis. Les données relatives aux regroupements d'îlots ou de côtés d'îlot dont la population est supérieure au seuil fixé peuvent être extraites à l'aide d'un tableau personnalisé.

Ces seuils de population sont appliqués aux données du Recensement de 2011 et à celles de tous les recensements antérieurs.

Veuillez consulter la section 2.2 Règles d'agrégation minimales des codes postaux pour les règles supplémentaires applicables aux données relatives aux codes postaux.

## **2.2 Règles d'agrégation minimales des codes postaux**

En plus des règles de confidentialité concernant la diffusion des données du recensement auprès des ménages assorties des codes postaux, les règles suivantes s'appliquent aux codes postaux. Ces règles s'inscrivent dans le cadre de la clause 03.01 (n) de l'Entente de licence commerciale concernant l'utilisation des codes postaux à des fins autres que la correspondance conclue entre Statistique Canada et la Société canadienne des postes.

- Toutes les demandes doivent comprendre des lots d'au moins deux codes postaux. Seuls les codes postaux dont le second caractère est un zéro (codes postaux ruraux) font exception.
- On doit attribuer aux groupes de codes postaux un numéro ou une classification unique (p. ex., K1A 0T6, 0T7, 0T8 = Région personnalisée 1); selon les modalités du contrat mentionné ci-dessus, les clients ne peuvent pas recevoir de listes de codes postaux, seul le nom mentionné dans la requête du client peut être utilisé.
- Toutes les autres règles de confidentialité pour les extractions personnalisées s'appliquent, conformément à la section 2.1.

En outre, l'avertissement qui suit doit être inclus dans toutes les demandes de codes postaux personnalisés :

Avis de non-responsabilité concernant la validation du code postal : Statistique Canada ne fait aucune déclaration et n'offre aucune garantie concernant l'exactitude des données relatives aux codes postaux<sup>MO</sup> soumis à Statistique Canada.

Ces règles s'appliquent aussi aux demandes de codes postaux historiques.

### **2.3 Arrondissement aléatoire**

Tous les chiffres présentés dans les totalisations du recensement font l'objet d'un arrondissement aléatoire, qui consiste à transformer les chiffres bruts en chiffres arrondis aléatoirement. Cette mesure réduit les risques de divulgation de l'identité des répondants dans les totalisations.

Tous les chiffres sont arrondis à un multiple de 5, ce qui signifie que tous les chiffres se termineront par un 0 ou par un 5. L'algorithme d'arrondissement aléatoire utilisé permet de contrôler les résultats et d'arrondir le chiffre des unités selon une fréquence prédéterminée comme le montre le tableau 2.1 suivant. Prenez note que les chiffres se terminant par un 0 ou par un 5 ne sont pas modifiés.

Tableau 2.1 Fréquence de l'arrondissement aléatoire

Valeurs de l'unité	Arrondies à un chiffre se terminant par 0	Arrondies à un chiffre se terminant par 5
1	4 fois sur 5	1 fois sur 5
2	3 fois sur 5	2 fois sur 5
3	2 fois sur 5	3 fois sur 5
4	1 fois sur 5	4 fois sur 5
5	Jamais	Toujours
6	1 fois sur 5	4 fois sur 5
7	2 fois sur 5	3 fois sur 5
8	3 fois sur 5	2 fois sur 5
9	4 fois sur 5	1 fois sur 5
0	Toujours	Jamais

L'algorithme d'arrondissement aléatoire se sert d'une valeur de départ aléatoire pour déclencher le processus d'arrondissement pour les tableaux. Lorsque ces routines sont appliquées, il est possible que le même chiffre dans le même tableau soit arrondi vers le haut dans une exécution et vers le bas dans une autre en raison de la méthode utilisée pour amorcer le processus.

## 2.4 Prévention de la divulgation pour les statistiques

Les statistiques (moyenne, erreur-type, somme, médiane, centile, ratio ou pourcentage) ne sont pas arrondies aléatoirement. Toutefois, l'inclusion des statistiques dans les tableaux où figurent également les chiffres à partir desquels elles ont été calculées peut se solder par la divulgation de l'identité des répondants. Pour éviter ce problème, on applique des méthodes de suppression pour les statistiques ou des calculs statistiques spéciaux.

### 2.4.1 Suppression de statistiques

Dans le cas de toutes les variables quantitatives, les statistiques sont supprimées si le nombre d'enregistrements réel à partir duquel elles ont été calculées est inférieur à 4.

### 2.4.2 Calculs statistiques spéciaux

- 1) La valeur prise par la statistique n'est jamais arrondie, sauf dans le cas des fréquences.
- 2) Toutes les statistiques fondées sur un classement par ordre des valeurs (médianes, centiles) sont calculées de la façon habituelle et ne sont jamais arrondies. Le minimum ou le maximum d'une statistique n'est jamais diffusé.

- 3) Lorsqu'on précise une somme pour l'âge, le programme multiplie alors la moyenne non arrondie du groupe en question par la fréquence arrondie. Autrement, si une somme spécifiée est pour une variable autre que l'âge, le programme arrondit la somme réelle.

Lorsqu'il faut faire une division (moyennes, pourcentages, ratios, etc.), le programme doit appliquer la règle énoncée au point 3) tant au numérateur qu'au dénominateur avant d'effectuer la division.

**Nota :** Les statistiques fondées sur un classement par ordre des valeurs, telles que la médiane et les centiles, sont toujours calculées au moyen d'interpolations linéaires. Ces statistiques ne sont donc pas fiables dans le cas de cellules comportant de faibles valeurs. C'est pourquoi aucune autre règle relative à la confidentialité ne leur est appliquée.

**Nota :** La moyenne d'un âge n'est pas modifiée par l'arrondissement, parce que le numérateur correspond au produit de la moyenne réelle multipliée par la fréquence arrondie, et que le dénominateur correspond à la fréquence arrondie. Les deux fréquences s'annulent l'une l'autre de sorte que la moyenne réelle n'est pas modifiée.

### **3. Pratiques relatives à la confidentialité**

#### **3.1 Ajustement des chiffres de population et des logements pour des raisons de confidentialité**

Il est possible que les chiffres de population relatifs aux îlots de diffusion comptant peu d'habitants soient ajustés afin de mieux protéger le caractère confidentiel des données. De fait, tous les chiffres de population inférieurs à 15, correspondant aux îlots de diffusion, doivent être arrondis à un multiple de 5. Cet ajustement est cependant contrôlé de sorte que les agrégations (totaux) des chiffres de population des aires de diffusion (AD) se situent toujours à plus ou moins 5 de la valeur réelle. Le contrôle sera encore plus serré à l'échelon des subdivisions de recensement (SDR) et des circonscriptions électorales fédérales (CEF). De fait, bien qu'ils demeurent toujours à plus ou moins 5 de la valeur réelle, les chiffres de population rajustés et les valeurs réelles concorderont pour le plus grand nombre de subdivisions de recensement et de circonscriptions électorales fédérales possible. Enfin, les chiffres de population rajustés et les valeurs réelles concorderont pour toutes les divisions de recensement, ce qui signifie que les chiffres de population des divisions de recensement (DR) demeurent inchangés.



## 3.2 Prévention de la divulgation

Il faut également tenir compte des risques de divulgation directe ou par recoupements au moment de déterminer le contenu des produits. L'évaluation du risque nécessite la prise en considération d'un certain nombre de facteurs. Les données détaillées relatives aux diverses variables, le recoupement des variables et l'échelon géographique selon lequel les données sont réparties ont tous une incidence sur le niveau de risque. Par exemple, il se peut que la production de tableaux indiquant le nombre de personnes que compte le logement et la tranche de valeur détaillée de l'âge selon diverses caractéristiques des membres des ménages ne pose pas de risque de divulgation dans le cas de grandes régions géographiques, mais que ce risque augmente pour les unités géographiques de niveau inférieur.

La méthode la plus largement utilisée pour prévenir la divulgation consiste à définir le contenu qui convient à un niveau géographique donné. Il est aussi possible de définir des seuils de population croissants ou de procéder à des suppressions manuelles, au besoin. Étant donné que ces méthodes sont liées aux exigences propres des divers produits, elles ne sont pas appliquées par les systèmes de suppression automatisés.

## 3.3 Totalisations du Recensement de l'agriculture

Les données du Recensement de l'agriculture et du Recensement de la population sont appariées à l'aide des données d'identification géographique, du nom, de l'âge et du sexe des exploitants agricoles. On effectue une imputation afin de tenir compte des cas de non-appariement. Il est possible d'obtenir des données sur tous les membres des ménages dont un des membres est un exploitant agricole.

Les données du Recensement de l'agriculture comprennent le type de ferme, la vente de produits agricoles, la superficie des terres de culture et le bétail, tandis que celles du Recensement de la population fournissent des données qui comprennent l'âge, le sexe, la structure de la famille, l'état matrimonial et la langue des membres des familles et des ménages. Des produits normalisés et des totalisations personnalisées dont l'élaboration est prévue sont offerts à l'échelle provinciale seulement. Les tableaux multidimensionnels sont communs et comprennent habituellement deux ou trois dimensions.

Les mesures de confidentialité comprennent l'arrondissement aléatoire des « chiffres » (p. ex., le nombre d'exploitants, le nombre de familles de recensement) et, dans quelques cas précis, une suppression des valeurs inférieures à un seuil précisé. Les moyennes sont fournies sans modification. Les totalisations sont toutes vérifiées à l'interne par le personnel du Recensement de l'agriculture avant leur diffusion.

Pour un résumé des règles de confidentialité concernant le Recensement de 2006 et les recensements antérieurs, veuillez consulter le document des [Normes et lignes directrices relatives à la confidentialité et à la qualité des données de 2006 \(version publique\)](#).

## **4. Pratiques relatives à la qualité des données**

La section ci-après décrit les méthodes utilisées pour restreindre la diffusion de données du recensement de qualité inacceptable.

### **4.1 Mesures de la qualité des données**

#### **4.1.1 Indicateurs de la qualité des données**

Des indicateurs de la qualité des données sont liés à toutes les régions géographiques normalisées du lieu de résidence sur lesquelles des données sont diffusées. Dans l'environnement de base de données du recensement, les indicateurs de la qualité des données consistent en un champ numérique à cinq chiffres. Dans les produits électroniques parcourus à l'aide de Beyond 20/20, ces indicateurs sont affichés en utilisant un code numérique à cinq chiffres (exemple : 0 2 1 0 0). Dans le site Web du recensement, les utilisateurs peuvent reconnaître les régions incomplètement dénombrées à l'aide de signes conventionnels.

##### **4.1.1.1 Les réserves indiennes et les établissements indiens incomplètement dénombrés**

En 2011 et dans les recensements antérieurs, le dénombrement n'a pas été complété dans certaines réserves indiennes et certains établissements indiens, suite à la non-participation. Les règles de qualité des données exigent que l'on identifie ces réserves indiennes et ces établissements incomplètement dénombrés, et qu'on les retire des produits. De plus, les régions géographiques à niveaux supérieurs comprenant ces régions doivent être identifiées dans les produits. Ces indicateurs de niveaux supérieurs sont inclus automatiquement dans les tableaux de sortie en format Beyond 20/20, CSV et fichiers plats. Pour une liste des réserves indiennes et des établissements indiens incomplètement dénombrés, les utilisateurs peuvent consulter la section du matériel de référence sur le site Web du recensement.

Bien que les données du recensement ne soient pas disponibles pour les réserves indiennes et les établissements indiens incomplètement dénombrés, les régions elles-mêmes sont comprises dans les hiérarchies géographiques normalisées de la base de données du recensement.

#### **4.1.1.2 Taux de réponse globaux**

On détermine les taux de réponse globaux pour chacune des régions géographiques du recensement. Ces régions sont identifiées dans la base de données selon le taux de non-réponse. Les régions qui présentent un taux de non-réponse égal ou supérieur à 25 % sont supprimées des totalisations. Les régions géographiques qui présentent un taux de non-réponse égal ou supérieur à 5 %, mais inférieur à 25 % ont été regroupées en deux catégories et sont marquées d'un indicateur distinct selon que le taux de non-réponse se situe entre 5 % et 10 %, ou entre 10 % et 25 %. Ces régions géographiques ne sont pas supprimées, mais sont marquées d'un indicateur dans les totalisations. On fournit un indicateur numérique avec l'identificateur de région pour indiquer des données de faible qualité dans les produits électroniques.

#### **4.1.1.3 Indicateur d'erreur dans les chiffres de population et des logements**

À la suite de la diffusion des chiffres de population et des logements, des erreurs sont occasionnellement relevées dans les données. Il est impossible d'apporter des changements aux données du recensement présentées. Les utilisateurs peuvent toutefois obtenir les modifications aux chiffres de population et des logements touchant les subdivisions de recensement et d'autres niveaux géographiques en consultant la section consacrée au Recensement de 2011, de 2006 ou de 2001 dans le site Web de Statistique Canada à l'adresse [www.statcan.gc.ca](http://www.statcan.gc.ca).

#### **4.1.1.4 Indicateur de qualité des données sans objet (indicateur relatif à la qualité des données-échantillon [20 %])**

Le quatrième code numérique de l'indicateur de qualité des données de cinq chiffres dans la base de données ne s'applique pas pour le Recensement de 2011 et est automatiquement remis à zéro pour chaque région géographique. La valeur résidant dans la base de données est un paramètre fictif pour des raisons d'ordre historique; en 2006 et pour les recensements antérieurs, les cinq chiffres s'appliquaient, le quatrième chiffre était l'indicateur relatif à la qualité des données-échantillon (20 %).

#### **4.1.1.5 Indicateur d'ajustement du chiffre de population de 2006**

Les utilisateurs qui souhaitent comparer les données du Recensement de 2011 avec celles des recensements antérieurs doivent tenir compte des modifications qui peuvent être apportées aux limites des régions géographiques d'un recensement à un autre. Afin de faciliter les comparaisons, les chiffres du Recensement de 2006 ont été ajustés au besoin en fonction des modifications apportées aux limites entre le Recensement de 2006 et celui de 2011. Les chiffres de 2006 qui ont été corrigés, ainsi que les régions créées depuis 2006, telles que les nouvelles municipalités constituées (subdivisions de recensement) et les nouvelles localités désignées,

sont eux aussi marqués d'un indicateur. La plupart de ces indicateurs ont toutefois pour objet de signaler des modifications de limites.

Le tableau 4.1 ci-après décrit la zone d'indicateur de la qualité des données et son contenu. Prenez note qu'un zéro dans l'un ou l'autre des cinq chiffres est la valeur implicite pour l'indicateur correspondant et signifie qu'aucune mesure relative à la qualité des données n'est requise.

**Tableau 4.1 Indicateurs de la qualité des données – Recensement de 2011**

<b>Caractère numérique</b>	<b>Description</b>	<b>Indicateur</b>	<b>Description des indicateurs</b>
1 <sup>er</sup> (0XXXX)	Indicateur de dénombrement partiel	0	Valeur implicite.
		1	Réserve indienne ou établissement indien partiellement dénombré (supprimées).
		2	Ne comprend pas les données du recensement pour une ou plusieurs réserves indiennes ou établissements indiens partiellement dénombrés.
2 <sup>e</sup> (X0XXX)	Indicateur relatif à la qualité des données	0	Valeur implicite.
		1	Indice de la qualité des données indiquant, pour le questionnaire du recensement abrégé (données intégrales [100 %]), un taux global de non-réponse supérieur ou égal à 5 %, mais inférieur à 10 %.
		2	Indice de la qualité des données indiquant, pour le questionnaire du recensement abrégé (données intégrales [100 %]), un taux global de non-réponse supérieur ou égal à 10 %, mais inférieur à 25 %.
		3	Indice de la qualité des données indiquant, pour le questionnaire du recensement abrégé (données intégrales [100 %]), un taux global de non-réponse supérieur ou égal à 25 % (supprimées).
3 <sup>e</sup> (XX0XX)	Indicateur d'erreur dans les chiffres de population et des logements	0	Valeur implicite.

Tableau 4.1 Indicateurs de la qualité des données – Recensement de 2011 (suite)

Caractère numérique	Description	Indicateur	Description des indicateurs
		1	Il y a une erreur dans les chiffres de population et des logements de 2011 pour cette région. Pour obtenir de plus amples renseignements, veuillez consulter la section des données sur les chiffres de population et des logements du fichier « Notes ».
		2	En 2006, les chiffres de population et/ou des logements sont erronés pour cette subdivision de recensement. Étant donné qu'il n'est pas possible d'apporter des changements aux données de 2006 présentées dans ces tableaux, les données de 2006 doivent être utilisées avec prudence. Pour obtenir de plus amples renseignements, veuillez consulter la section des données sur les chiffres de population et des logements du fichier « Notes ».
		3	Les chiffres de population et/ou des logements de 2011 et de 2006 pour cette région sont erronés. Étant donné qu'il n'est pas possible d'apporter des changements aux données du recensement présentées dans ces tableaux, ces données doivent être utilisées avec prudence. Pour obtenir de plus amples renseignements, veuillez consulter la section des données sur les chiffres de population et des logements du fichier « Notes ».
4 <sup>e</sup> (XXXOX)	Sans objet	0	Valeur implicite.
5 <sup>e</sup> (XXXX0)	Indicateur de rajustement des chiffres de population de 2006	0	Valeur implicite.
		1	Chiffre de 2006 rajusté; la plupart de ces rajustements sont le résultat de modifications des limites.

**Nota** : Les indicateurs de la qualité des données ne s'appliquent pas aux chiffres de population et des logements.

Veillez consulter le document des [Normes et lignes directrices relatives à la confidentialité et à la qualité des données de 2006 \(version publique\)](#) pour le tableau des indicateurs pour les années de recensement antérieures.

## **4.2 Autres méthodes de suppression relatives à la qualité des données**

Les méthodes de suppression mentionnées jusqu'à maintenant sont suffisantes pour supprimer les régions pour lesquelles la qualité des données est inacceptable et pour signaler les données de qualité inférieure dans la plupart des produits de données du recensement. Toutefois, le secteur qui établit les spécifications ou le secteur chargé de la production peut demander que des règles de suppression additionnelles en raison de la qualité des données soient appliquées pour certains produits : par exemple, en augmentant les seuils de population appliqués ou en supprimant des répartitions ou des cellules. Il s'agit de règles de suppression qui s'appliquent à des produits en particulier; elles ne font donc pas partie des systèmes de suppression automatisés. Dans tous les cas, il est nécessaire d'utiliser un processus manuel.

## **4.3 Calcul de statistiques d'ordre**

Les variables, dont les valeurs sont des nombres entiers, une médiane (ou autre quantile) est calculée au moyen d'interpolations linéaires afin de donner une décimale à la variable, même si la variable est un nombre entier. Ceci est effectué afin de donner une idée de la position relative de l'enregistrement de la médiane parmi les enregistrements de même valeur (la médiane). Une valeur de 23,46 veut dire que l'enregistrement du milieu (la médiane) démontre Âge = 23 et 46 % des enregistrements avec Âge = 23 se trouvent à la gauche du milieu. Par exemple, l'ensemble suivant {23, 23, 23, 23, **23**, 23, 23, 23, 23} (la médiane est en caractère gras) donne 23,11 comme médiane calculée/déclarée.

## **4.4 Règle sur la qualité des données pour la diffusion des données relatives à la population de 100 ans et plus**

Les données sur la population âgée de 100 ans et plus ne peuvent être diffusées en année d'âge. Pour les demandes personnalisées qui exigent une ventilation plus détaillée que celle des produits de données normalisés, dans lesquels on groupe la population âgée de 100 ans et plus, la seule ventilation détaillée possible est comme suit, et ne peut être fournie que pour « Canada » :

**Population totale âgée de 100 ans et plus**

**100 ans à 104 ans**

**105 ans à 109 ans**

**110 ans et plus**

**4.5 Règles relatives à la qualité des données pour la diffusion des données sur les couples de même sexe et de sexe opposé**

Les questionnaires du Recensement de 2011 et de l'Enquête nationale auprès des ménages de 2011 ont utilisé, pour la première fois, une réponse précise sur les liens entre les membres du ménage afin de déterminer le nombre de couples mariés de même sexe. L'analyse des données sur les couples mariés de même sexe a montré qu'une surestimation de ce type de familles et d'état matrimonial a pu survenir. Le Recensement de 2011 montre un total de 64 575 couples de même sexe au Canada, dont 21 015 sont des couples mariés. À l'échelle nationale, l'écart de la surestimation de ces deux chiffres varie entre 0 et 4 500.

Les niveaux géographiques tels que le Canada, les provinces, les territoires et les régions métropolitaines de recensement (RMR) affichent des chiffres généralement plus élevés, donc on s'attend à ce que la surestimation potentielle soit relativement petite ; toutefois, il faut quand même interpréter les données avec prudence.

À un niveau géographique moins élevé, la même surestimation potentielle peut être relativement importante, alors non seulement faut-il interpréter les données avec prudence, mais, certaines règles de suppression limitent leur publication. Ces règles s'appliquent aux données du Recensement de 2011 et aux données de l'Enquête nationale auprès des ménages de 2011.

Premièrement, la ventilation des données sur les couples de même sexe ou les couples de sexe opposé selon la situation conjugale, c.-à-d. s'ils sont mariés ou vivant en union libre, ne doit pas être diffusée pour des régions géographiques autres que le Canada, les provinces, les territoires et les RMR.

Deuxièmement, les données qui identifient les couples de même sexe ou les couples de sexe opposé (au total, couples mariés ou vivant en union libre) pour toutes les régions où la population compte moins de 5 000 habitants (tel qu'établi lors du Recensement de 2011 pour les ménages privés) ne doivent pas être diffusées.

En résumé,

- Toutes les données pour les couples de même sexe ou les couples de sexe opposé pour le Canada, les provinces, les territoires et les régions métropolitaines de recensement (RMR) peuvent être diffusées, toutefois elles doivent être interprétées avec prudence.
- Les données sur les couples de même sexe ou les couples de sexe opposé peuvent être diffusées pour d'autres régions géographiques si leur population compte 5 000 habitants ou plus, à condition que la ventilation selon la situation conjugale (couples mariés ou vivant en union libre) ne soit pas incluse.
- Aucune données identifiant les couples de même sexe ou les couples de sexe opposé ne peuvent être diffusées pour les régions comptant une population de moins de 5 000 habitants.

## **5. Suppression des données – Autre**

Des règles de suppression de données peuvent être appliquées pour des raisons propres à des produits en particulier, par exemple, la taille des produits ou les contraintes liées au support de diffusion utilisé.

### **5.1 Fréquence de déclaration**

On se sert de la fréquence de déclaration pour ordonner ou classer les données sur les caractéristiques (variables) selon la taille dans les produits. On peut ainsi choisir d'inclure dans un produit uniquement les « n » catégories les plus importantes d'une caractéristique.

### **5.2 Suppression des zéros**

La suppression des zéros consiste à retirer les enregistrements dans lesquels tous les chiffres sont des zéros. Cette mesure sert à réduire la taille d'un produit en supprimant toutes les lignes de la matrice de sortie renfermant des zéros.



## 6. Meilleures pratiques

### 6.1 Signes conventionnels relatifs à la qualité et à la confidentialité dans les tableaux de données

Nouveau en 2011 : trois [Signes conventionnels dans les tableaux](#) de Statistique Canada seront inclus dans les produits normalisés personnalisés du recensement. Le tableau ci-dessous illustre chaque nouveau signe conventionnel adopté en 2011 et sa description ainsi que le signe utilisé dans les produits du Recensement de 2006 (format HTML et B20/20).

#### Nouveaux signes conventionnels relatifs à la qualité et à la confidentialité

Nouveaux signes conventionnels	Description	Signes conventionnels utilisés pour les produits du Recensement de 2006 (HTML)	Signes conventionnels utilisés pour les produits du Recensement de 2006 (B20/20)
..	indisponible pour une période de référence précise	N	- (tiret)
...	n'ayant pas lieu de figurer	N	- (tiret)
x	confidentiel en vertu des dispositions de la <i>Loi sur la statistique</i>	0 (zéro)	0 (zéro)

Il est important de noter que ces signes sont utilisés **seulement pour les produits du Recensement de 2011**. Aucun produit historique (Recensement de 2006 et avant) ni aucune donnée historique dans les produits du Recensement de 2011 ne contiendront ces signes.