

Mesure des erreurs de couverture du recensement canadien
COVERAGE ERROR MEASUREMENT OF THE CANADIAN CENSUS
MEDIDA DE LOS ERRORES DE COBERTURA DEL CENSO
CANADIENSE

Jean-René Boudreau

Volume 18, Number 2, Fall 1989

Sous-dénombrement et estimation de population

URI: <https://id.erudit.org/iderudit/010017ar>

DOI: <https://doi.org/10.7202/010017ar>

[See table of contents](#)

Publisher(s)

Association des démographes du Québec

ISSN

0380-1721 (print)

1705-1495 (digital)

[Explore this journal](#)

Cite this article

Boudreau, J.-R. (1989). Mesure des erreurs de couverture du recensement canadien. *Cahiers québécois de démographie*, 18(2), 265–284.
<https://doi.org/10.7202/010017ar>

Article abstract

Coverage errors are to some extent unavoidable in a Census of Population, and can have serious effects on counts, distributions, intercensal estimates and demographic projections. Consequently, it is important to determine the nature and the magnitude of coverage errors. Methods have been designed to achieve this goal. This paper describes the methods used by Statistics Canada to measure coverage errors in the Canadian Census.

Mesure des erreurs de couverture du recensement canadien

Jean-René BOUDREAU*

INTRODUCTION

Le but premier d'un recensement a toujours été et demeure le dénombrement de la population de chaque région géographique d'un pays. Il doit inventorier cette population et ses conditions de vie. C'est la seule source de renseignements qui permet aux citoyens d'obtenir une image complète et fiable de leur communauté et de leur pays. L'image enregistrée au jour du recensement résulte de la combinaison des réponses à diverses questions, par exemple sur l'âge, le sexe, l'éducation et l'emploi, et de la comparaison des changements survenus entre les recensements. Le recensement constitue une source précieuse de connaissances utiles à tous les paliers de l'administration publique ainsi qu'aux entreprises, aux associations et au public en général.

Dans un projet de grande envergure tel que le recensement du Canada, les résultats ne sont jamais parfaits. Les erreurs de couverture peuvent influencer sur l'exactitude des chiffres, c'est-à-dire sur la taille des univers de la population, des familles, des ménages et des logements. Il existe deux types d'erreurs de couverture : celles qui causent le sous-dénombrement et celles qui causent le surdénombrement. On parle de sous-dénombrement lorsqu'une unité faisant partie de l'univers visé par le recensement est omise lors du dénombrement. Quant au surdénombrement, il peut se produire dans deux cas :

* Statistique Canada, Division des méthodes d'enquêtes sociales, Ottawa.

premièrement, lorsqu'une unité faisant partie de l'univers visé est dénombrée plus d'une fois; et deuxièmement, lorsqu'une unité ne faisant pas partie de l'univers visé est dénombrée par erreur.

Ces erreurs peuvent se produire à la collecte ou au dépouillement d'un recensement. Ainsi, au cours du dépouillement, l'annulation d'inscriptions de personnes ou de ménages, la perte de questionnaires ou la création artificielle de personnes ou de ménages sont des erreurs de couverture. Cependant, en général, les erreurs de couverture résultent d'erreurs de collecte liées aux procédures, cartes ou manuels utilisés, ou commises par les recenseurs ou les répondants eux-mêmes. Ainsi, des logements peuvent ne pas être recensés parce qu'ils sont cachés ou semblent inhabitables. Des personnes sont omises lorsqu'on a, par erreur, classé leur logement comme inoccupé ou qu'elles n'ont pas été incluses dans le questionnaire en raison d'une mauvaise interprétation des instructions concernant les personnes à inclure. De plus, des logements (et par conséquent des ménages et des personnes) peuvent être omis ou comptés plus d'une fois en raison de l'utilisation de cartes imparfaites, ou parce qu'une portion de secteur de dénombrement (plus petite unité géographique utilisée durant la collecte) n'a pas été recensée ou qu'une mauvaise interprétation des limites des secteurs de dénombrement a entraîné un chevauchement.

Supposons que T représente le nombre total d'unités dans l'univers visé et que C correspond au chiffre du recensement pour cet univers. Dans ce cas, l'erreur ou le biais qui résulte de l'utilisation de C au lieu de T est

$$B=T-C$$

Il s'agit de l'erreur de couverture nette. Supposons que M représente le sous-dénombrement total, c'est-à-dire le nombre total d'unités de l'univers visé qu'on a omis de dénombrer, et que E représente le nombre total d'unités distinctes de l'univers visé qui sont dénombrées. Dans ce cas :

$$T=M+E$$

et nous pouvons écrire que

$$B=(M+E)-C=M-(C-E),$$

où le deuxième terme de l'équation ci-dessus,

$$O=C-E$$

correspond au surdénombrement. Cette erreur découle non seulement du dénombrement répété des mêmes unités mais aussi du dénombrement d'unités qui ne font pas partie de l'univers visé.

Ces erreurs sont souvent exprimées en termes relatifs, c'est-à-dire en tant que proportion du nombre total d'unités de l'univers visé. Nous définissons donc :

- le taux de sous-dénombrement : $TM = M/T - M/(C+M-O)$
- le taux de surdénombrement T : $TO = O/T = O/(C+M-O)$
- le taux d'erreur de couverture nette : $TB = TM-TO$.

Nous nous proposons d'expliquer la méthodologie employée par Statistique Canada afin de déterminer les erreurs de couverture du recensement canadien pour l'univers des personnes. Précisons au préalable que cette population comprend :

- tous les citoyens canadiens et immigrants reçus qui ont un lieu habituel de résidence au Canada le jour du recensement;
- tous les citoyens canadiens et immigrants reçus qui sont à l'étranger, en poste sur une base militaire ou attachés à une mission diplomatique le jour du recensement;
- tous les citoyens canadiens et immigrants reçus qui se trouvent en mer ou au port, à bord de navires marchands immatriculés au Canada ou de vaisseaux de guerre canadiens, le jour du recensement.

SOUS-DÉNOMBREMENT

Le sous-dénombrement de la population est généralement considéré comme l'une des plus importantes sources d'erreur qui touchent les données du recensement. Il introduit un biais vers le bas dans la mesure où les données publiées par le recensement sous-estiment les vrais totaux de la population. Il peut également fausser les distributions de la population estimées à partir des données du recensement lorsque les personnes recensées et les personnes non recensées ne possèdent pas les mêmes caractéristiques.

Pour estimer ce biais, Statistique Canada a choisi d'utiliser la méthode de micro-appariement entre les données du recensement et celles d'une source externe fiable, soit la contre-vérification des dossiers (CVD). Cette méthode s'explique

sommairement de la façon suivante. Un échantillon des personnes qui devaient être recensées est choisi à partir de sources indépendantes du recensement. Des opérations de dépistage, de recherche et de suivi permettent de classer de façon définitive chaque personne échantillonnée dans les catégories suivantes : recensée, non recensée, décédée et émigrée. Ces résultats sont ensuite pondérés en fonction de la taille de la population. (Le Yukon et les Territoires du Nord-Ouest ne sont pas inclus dans l'enquête à cause de la difficulté d'y choisir un échantillon représentatif de la population, dont au moins le tiers est composée de migrants entrants ou sortants intercensitaires.)

Sélection de l'échantillon

L'ensemble des personnes qui doivent être recensées est, sauf exceptions, l'union des quatre ensembles ou bases de personnes suivants, appelés bases de sélection : a) la totalité des personnes recensées au recensement antérieur; b) la totalité des naissances intercensitaires; c) la totalité des immigrants reçus arrivés durant la période intercensitaire; d) la totalité des personnes non recensées au recensement antérieur qui appartiennent à la couverture du recensement. La base de sélection (d) est conceptuelle : il est évident que nous ne possédons pas la liste de ces personnes. Les exceptions auxquelles nous avons fait allusion sont toutes les personnes décédées ou émigrées durant la période intercensitaire.

Un échantillon de personnes est tiré dans chacune des bases de sélection. L'échantillon de la base de sélection (d) est essentiellement l'ensemble des personnes non recensées déterminé par la CVD du recensement précédent. Pour les trois autres bases de sélection, on utilise l'allocation de Neyman (Cochran, 1977 : 99), afin de déterminer la répartition de l'échantillon entre les strates qui optimise la précision de l'estimation globale. Cependant, on incorpore un facteur additionnel à cette technique pour tenir compte de l'émigration, des décès et de la non-réponse (des unités pour lesquelles il n'est pas possible de déterminer si elles sont recensées ou non recensées).

L'échantillonnage est effectué de façon indépendante à l'intérieur de chaque base de sélection. Le plan de sondage varie d'une base de sélection à l'autre selon la nature de la liste utilisée.

La base du recensement est d'abord stratifiée par province du lieu habituel de résidence au recensement antérieur. L'allocation de l'échantillon entre les dix provinces est établie par un compromis entre la précision de l'estimation nationale et la précision des estimations provinciales. Les tailles d'échantillon minimales permettant d'atteindre un critère de précision minimale sont déterminées pour les petites provinces. On répartit le reste de l'échantillon dans les autres provinces en vue d'optimiser la précision de l'estimation globale. À l'intérieur de chaque province, les strates sont définies selon la méthode d'énumération employée au recensement précédent (retour par la poste, reprise et interview) et la taille des régions urbanisées. On utilise l'allocation de Neyman pour attribuer l'échantillon dans ces strates. À l'intérieur de chaque strate, un échantillon à deux degrés est sélectionné. Au premier degré, un échantillon de secteurs de dénombrement (SD), qui comprend en moyenne 200 ménages, est choisi proportionnellement à la taille de la population des SD. Puis, à l'intérieur de chaque SD choisi, un échantillon de personnes systématique et de taille fixe est sélectionné.

La base des naissances est stratifiée par province du lieu habituel de résidence de la mère et par année de naissance. L'allocation de l'échantillon entre les années de naissance se fait de façon proportionnelle au nombre de personnes. Après cette attribution, l'allocation de l'échantillon entre les provinces se fait encore par un compromis entre la précision des estimations nationales et la précision des estimations provinciales. On utilise l'échantillonnage systématique dans chacune des strates.

La base des immigrants reçus est stratifiée par année d'arrivée au Canada. L'allocation est établie proportionnellement au nombre de personnes dans chacune des strates.

On utilise la méthode des répliques (Wolter, 1985) pour obtenir une estimation de l'erreur échantillonnale des estimations produites par l'étude. Quatre répliques sont utilisées. Pour les bases de sélection (a), (b) et (c), elles sont déterminées par le choix de quatre positions de départ différentes pour toutes les sélections systématiques à l'intérieur des strates. Une attribution au hasard des répliques est effectuée à l'échantillon de la base de sélection (d). Chaque réplique contient donc le quart de l'échantillon global. Ordinairement, chaque traitement effectué aux données doit respecter l'intégrité des

répliques pour qu'elles soient considérées comme quatre échantillons indépendants de la même population.

Opérations de dépistage et de recherche

Le but des diverses opérations de la CVD est de classer chaque personne choisie dans une des catégories suivantes : recensée, non recensée, décédée, émigrée.

L'ensemble des opérations nécessaires pour arriver à ce résultat se divise en deux étapes : le dépistage et la recherche. Puisque les adresses obtenues au moment de la sélection de l'échantillon ne sont généralement pas exactes (par exemple parce qu'il y a eu déménagement durant la période intercensitaire), il faut d'abord entreprendre un dépistage afin d'établir l'adresse de chaque personne choisie au jour du recensement présent. Le système de dépistage consiste en une suite d'opérations qui varient d'une base de sélection à l'autre et comprennent :

a) un jumelage avec un fichier administratif, qui permet d'obtenir une mise à jour de l'adresse de la personne choisie; cette opération doit être faite avant le recensement;

b) un appariement manuel intégré tôt dans les opérations de dépouillement, qui consiste en une recherche dans les questionnaires du recensement visant à déterminer si les personnes choisies sont dénombrées à l'adresse obtenue à l'étape (a);

c) des dépistages par suivis téléphoniques faits aux bureaux régionaux;

d) des recherches dans divers dossiers administratifs.

Les cas résolus à l'opération (b) ci-dessus sont classés comme recensés et ne font pas l'objet d'autres recherches. Sinon, dès que le dépistage par voie de suivis téléphoniques permet d'identifier une adresse possible de la personne choisie, on effectue une recherche (après les opérations de dépouillement du recensement) dans les documents du recensement au Bureau central, à Ottawa, afin de déterminer si cette personne est recensée. De plus, on examine les registres de décès dans le cas des personnes choisies qui sont vraisemblablement décédées, afin de s'assurer de la validité de cette information. Toutefois, aucune vérification des personnes dépistées comme «ayant émigré avant le jour du recensement» ne peut se faire, car il n'existe pas de registre de

l'émigration au Canada. Une personne n'est classée dans cette catégorie que si la source d'information est jugée assez fiable.

Malgré ces mesures, une certaine proportion des personnes choisies dans l'échantillon ne sont pas dépistées et, par conséquent, ne peuvent pas être classées dans l'une des quatre catégories. Nous les appellerons «non-répondants» ou «non-dépistés».

Pondération

Le facteur de pondération (poids) initial d'une personne choisie des bases de sélection (a), (b) et (c) est l'inverse de la probabilité de sélection. Pour la base de sélection (d), ce facteur correspond au poids final de cette même personne à la CVD antérieure.

Comme nous l'avons dit précédemment, il est impossible de classer toutes les personnes choisies dans l'une des catégories : recensée, non recensée, décédée ou émigrée. Il va de soi que si l'on estimait le nombre total de personnes appartenant à l'une de ces catégories en ne sommant que les facteurs de pondération correspondants, on aboutirait à une sous-estimation de ce total. Il faut donc transférer les poids des non-répondants, c'est-à-dire des personnes non dépistées, aux poids des répondants afin de compenser ce biais.

Deux types d'ajustement sont appliqués aux facteurs de pondération : un ajustement pour tenir compte de la non-réponse et une post-stratification des bases de sélection.

En ce qui concerne l'ajustement pour tenir compte de la non-réponse, on forme des groupes de réponses dans le but de minimiser les biais. Ces groupes sont censés contenir des personnes choisies ayant sensiblement les mêmes caractéristiques. Les ajustements sont faits de façon indépendante à l'intérieur de chaque groupe de réponses. Les variables utilisées pour la formation de ces groupes sont :

a) base du recensement : la province de résidence et la taille des régions urbanisées, ainsi que le sexe et l'âge au recensement antérieur et la réplique;

b) base des naissances : la province de naissance, l'année de naissance et la réplique;

c) base des immigrants : la province de naissance, l'année d'immigration et la réplique;

d) base des personnes non recensées : la réplique.

En plus, deux variables codées aux opérations de dépistage sont utilisées pour chacune des bases. La première variable résulte de la mise à jour des adresses (jumelage avec un fichier administratif), la deuxième de l'appariement manuel (voir la section sur les opérations de dépistage et de recherche).

Un groupe de réponses est jugé valide si le nombre de personnes choisies à l'intérieur de ce groupe excède un minimum déterminé ou si le taux de non-réponse est inférieur à un pourcentage donné à l'avance. Si ces deux facteurs ne sont pas respectés, on fusionne un ou plusieurs groupes de façon à ce que le nouveau groupe satisfasse aux conditions de validité. La plupart des regroupements sont faits de façon automatique.

Les ajustements se font en trois étapes parce que trois types de non-réponses sont définis. Le premier type (non identifiable) regroupe toutes les personnes choisies pour lesquelles on n'a pas assez d'informations pour entreprendre quelque recherche ou dépistage que ce soit. Le deuxième type (non dépisté) regroupe toutes les personnes choisies pour lesquelles les opérations de dépistage ne donnent rien. Finalement, le troisième type (non classifiable) regroupe toutes les personnes choisies pour lesquelles les recherches de dépistage permettent d'établir qu'elles ne sont ni décédées, ni émigrées, mais pour lesquelles il est impossible d'établir de façon précise la catégorie finale.

Dans un premier temps, un poids nul est attribué aux personnes choisies non identifiables et on ajuste les poids des autres personnes choisies de façon à compenser la perte de masse (à l'intérieur de chaque groupe de réponses). Dans la seconde étape, on accomplit le même processus, mais cette fois en attribuant un poids de zéro aux personnes choisies non dépistées. Finalement, un dernier ajustement est fait pour les personnes choisies non classifiables. On ne modifie cependant pas dans cet ajustement les poids des personnes décédées ou émigrées. Il convient de noter également qu'il faut laisser nul le poids des personnes choisies non identifiables lors de l'ajustement des personnes choisies non dépistées. Il en est de même pour les personnes choisies non identifiables et non dépistées lors de l'ajustement des personnes choisies non classifiables.

L'objectif du deuxième type d'ajustement, c'est-à-dire la post-stratification des bases de sélection, est d'obtenir la concordance, à l'intérieur de groupes spécifiques (appelés groupes de pondération), entre les comptes estimés tirés de l'échantillon et les comptes de la population obtenus à partir

des bases de sélection. La post-stratification permet également d'obtenir une erreur quadratique moyenne plus petite. Deux ajustements indépendants sont faits : l'un tient compte de l'appartenance aux répliques, tandis que le deuxième ignore les répliques.

Les variables qui définissent ces groupes de pondération sont :

a) base du recensement : la province de résidence et la taille des régions urbanisées, ainsi que l'âge, le sexe et l'état matrimonial au recensement antérieur et la réplique;

b) base des naissances : la province de résidence de la mère, l'année de naissance et la réplique;

c) base des immigrants : l'année d'immigration, l'âge, le sexe et la réplique;

d) base des personnes non recensées : la réplique.

Un groupe de pondération doit avoir un nombre minimal de personnes choisies. Une stratégie de regroupement est opérée, en majeure partie de façon automatique, pour que chaque groupe de pondération satisfasse à cette contrainte. Pour ce qui est du deuxième ajustement indépendant, on redéfinit les groupes de pondération sans tenir compte des répliques. La stratégie de regroupement est la même que celle de la première post-stratification.

Les deux types d'ajustement suivent essentiellement la même procédure. Il s'agit, premièrement, de faire le compte de la population tel qu'il apparaît à la lecture des bases de sélection, et ce pour chacun des groupes de pondération; deuxièmement, d'estimer le compte de la population pour chacun des groupes de pondération en sommant les poids des personnes choisies le constituant (en utilisant les poids obtenus suite aux ajustements visant à tenir compte de la non-réponse); et, troisièmement, d'ajuster le poids de chaque personne choisie par un facteur constant à l'intérieur de chaque groupe de pondération de manière à ce que la somme des poids concorde avec les comptes obtenus des bases de sélection au niveau des groupes de pondération.

Estimation

Notons O1 l'estimation d'une caractéristique consécutive à la première post-stratification et O2 l'estimation de cette même caractéristique consécutive à la deuxième post-stratification.

De même, notons V1 et V2 leur variance associée. Il est noté (Cochran, 1977 : 166) que O2 est moins sujet au biais que O1 : c'est la raison du choix de O2 pour les estimations des totaux et proportions. D'autre part, les variances V1 et V2 sont généralement du même ordre de grandeur (Wolter, 1985 : 24). Puisqu'on peut trouver une estimation non biaisée de V1 en utilisant les poids de la première post-stratification, nous avons décidé d'utiliser cette estimation de V1 comme mesure de dispersion pour O2.

Remarques

Les résultats de la contre-vérification des dossiers, quoique instructifs, ne peuvent pas nécessairement être acceptés comme exacts. Même si l'étude est menée très soigneusement, sa méthodologie présente des lacunes certaines. Les bases de sondage ne sont pas exhaustives. Rappelons que l'étude ne couvre pas le Yukon et les Territoires-du-Nord-Ouest; donc, les personnes qui ont quitté ces territoires pour l'une des provinces durant la période intercensitaire ne peuvent pas être sélectionnées par l'étude. Sont aussi exclus les Canadiens qui rentrent au pays après avoir résidé à l'étranger pendant un certain temps.

Nous ne sommes pas sûr que la tendance des personnes choisies non dépistées à se dénombrer soit la même que celle des personnes choisies qui ont été classées. Ces deux groupes de personnes de l'échantillon ont-ils le même comportement quant au dénombrement ? Pour qu'un biais important ne soit pas créé dans les estimations, il faut que dans l'échantillon le pourcentage de personnes choisies non dépistées soit très inférieur au taux de personnes non recensées. Si cette condition n'est pas respectée, il y a lieu de considérer les résultats de la contre-vérification des dossiers avec une certaine réserve.

La méthodologie de l'enquête induit un biais positif dans les estimations de personnes non recensées du fait que les erreurs de classification ne sont pas symétriques. Il n'y a pas d'erreur de classification lorsque le résultat est « recensé » parce que ces personnes ont été trouvées, justement, dans les documents du recensement. Toutefois, les personnes classées « non recensées » ne sont pas trouvées à leur adresse dépistée; si l'adresse est incorrecte, il est possible qu'elles soient dénombrées ailleurs au Canada.

On peut comparer quelques résultats de l'étude avec d'autres sources de données pour évaluer sa performance. Par exemple, la CVD permet d'estimer le nombre de personnes recensées dans les dix provinces; une comparaison de cette estimation avec le compte du recensement peut être envisagée. La CVD estime également le nombre de décès survenus durant la période intercensitaire; ce résultat peut être vérifié à l'aide des registres de l'état civil. De même, le nombre de personnes émigrées durant la période intercensitaire estimé par la CVD peut être comparé aux estimations démographiques du nombre d'émigrés. Enfin, on peut vérifier le nombre de personnes qui n'ont pas été recensées parce que leur logement a été classé comme inoccupé en utilisant les résultats d'une étude spéciale sur les logements inoccupés, prévue par le programme de mesure de l'erreur de couverture du recensement de 1986.

Quelques résultats pour 1986

Les chiffres présentés ci-dessous (tableaux 1 à 3) témoignent de l'ampleur des différentes opérations et donnent les paramètres utilisés dans les opérations. En 1986, la taille de l'échantillon de la CVD s'établit à 36 378; 32 200, 1776, 1341 et 1061 personnes sont respectivement choisies dans les bases de sélection (a), (b), (c) et (d).

On a vu que le nombre minimal de personnes choisies utilisé pour la construction des groupes de réponses a été fixé à 20 et que le pourcentage maximal de non-réponses est de 30 (voir la section sur la pondération). Avant la sélection, on dédouble les personnes âgées de 15 à 19 ans au recensement de 1981 pour augmenter leur probabilité de sélection d'un facteur 2, dans le but de rendre le plan de sondage plus efficace. En effet, ce groupe d'âge, c'est-à-dire les personnes âgées de 20 à 24 ans au recensement de 1986, a le taux de sous-dénombrement le plus élevé. La stratégie du regroupement des groupes de réponses (voir la section sur la pondération) s'énonce comme suit : la discrimination sur le sexe est la première oubliée; viennent ensuite l'âge et le code de l'opération de jumelage avec un fichier administratif (voir la section sur les opérations de dépistage et de recherche). Si toutes ces règles sont vaines, on regroupe des strates (taille des régions urbanisées, année de naissance ou d'immigration selon le cas) pour satisfaire à la condition de taille énoncée ci-dessus. Pour ce qui

TABLEAU 1
*Contre-vérification des dossiers (CVD). Nombre de cas
dans chacune des catégories finales selon la base^a*

Classification des personnes choisies	Base du recensement antérieur		Base des naissances intercensi- taires		Base des immigrants intercensi- taires		Base des non- recensés	
	N	%	N	%	N	%	N	%
Recensées	28 551	88,7	1 587	89,3	870	64,9	757	87,3
Non recensées	1 320	4,1	35	2,0	115	8,6	131	4,4
Décédées	1 142	3,5	10	0,6	10	0,7	31	3,3
Émigrées ou à l'étranger	275	0,9	16	0,9	97	7,2	37	1,2
Non identifiables, non dépistées, non classifiables	912	2,8	128	7,2	249	18,6	105	3,8
Total	32 200	100,0	1 776	100,0	1 341	100,0	1 061	100,0

a. Il s'agit des chiffres non pondérés.

est des groupes de pondération (dans l'opération de post-stratification des bases de sélection), le nombre minimal de personnes choisies pour les groupes est fixé à 20. Pour le premier ajustement, le regroupement suit les mêmes règles que celles qui ont été décrites plus haut pour les groupes de réponses, mais les variables issues des opérations ne sont pas utilisées. L'état matrimonial est la première variable oubliée pour les regroupements; elle est suivie du sexe et de l'âge.

Dans les calculs des proportions des personnes non recensées, nous n'avons pas tenu compte du surdénombrement. Il n'existe pas en 1986 d'estimation du surdénombrement pour toute la population.

Le tableau des estimations du sous-dénombrement par âge et sexe (tableau 2) montre bien que le sous-dénombrement modifie sensiblement les profils de la population. Les résultats, cependant, sont sujets à caution du fait que le taux non pondéré de non-dépistés se situe à 3,8 %; il est donc du même ordre de grandeur que le taux de sous-dénombrement.

Comme en témoigne le tableau 3, la vérification indépendante (voir la section «Remarques») du nombre de personnes recensées révèle une différence qui ne peut pas être expliquée totalement par l'erreur échantillonnale. Cet écart peut avoir plusieurs causes : la non-exhaustivité des bases de

TABLEAU 2
Estimation du sous-dénombrement de la population^a par âge et sexe

Caractéristiques	Nombre de personnes non recensées ^b		Taux de sous-dénombrement de la population	
	Nombre estimé	Écart type	Taux estimé (%)	Écart type (%)
HOMMES	506 500	21 400	3,91	0,16
0-4 ans	21 000	6 500	2,22	0,67
5-14 ans	38 100	6 100	2,04	0,32
15-19 ans	42 800	8 100	4,18	0,75
20-24 ans	135 300	8 300	10,71	0,59
25-34 ans	138 200	10 400	5,81	0,41
35-44 ans	64 000	10 000	3,40	0,51
45-54 ans	26 000	6 900	2,00	0,52
55-64 ans	21 600	5 400	1,88	0,47
≥ 65 ans	19 600	6 100	1,70	0,52
FEMMES	377 500	22 300	2,87	0,16
0-4 ans	21 100	5 500	2,35	0,60
5-14 ans	39 200	6 100	2,21	0,33
15-19 ans	34 800	8 300	3,58	0,83
20-24 ans	88 400	9 300	7,33	0,71
25-34 ans	87 400	10 600	3,71	0,43
35-44 ans	25 200	5 900	1,37	0,32
45-54 ans	19 600	4 800	1,53	0,37
55-64 ans	28 100	6 500	2,28	0,51
≥ 65 ans	33 700	7 100	2,11	0,44

a Les Territoires sont exclus.

b Les estimations tiennent compte des ajustements prévus par le programme de mesure des erreurs de couverture du recensement canadien.

sélection; des erreurs de classification entre les catégories « recensé » et « non recensé »; l'exclusion de la CVD des illégaux qui se sont dénombrés eux-mêmes ou une augmentation significative des dédoublements de personnes au recensement de 1986 par rapport à celui de 1981. L'importance de ces sources potentielles d'erreur n'est pas connue.

LE SURDÉNOMBREMENT

Le surdénombrement est défini comme suit : est considérée comme surdénombrée a) toute personne incluse dans la couverture du recensement qui est dénombrée deux fois ou plus; ou b) toute personne dénombrée au moins une fois sans faire

TABLEAU 3
Comparaison de la CVD de 1986 avec d'autres sources

Population	CVD de 1986		Autres sources	
	Estima- tions (x 1000)	Écart type (x 1000)	Estima- tions (x 1000)	Sources
Dénombrée	24 737	49	25 064	Recensement ^a
Décédée	906	25	883	Statistique de l'état civil ^b
Émigrée	288	19	235	Estimation par une méthode démographique
Non recensée dans les logements inoccupés	81	11	94	Vérification des logements inoccupés

a. Les chiffres du recensement excluent l'ajustement que l'on effectue pour tenir compte des réserves indiennes partiellement dénombrées et les ajustements prévus par le programme de mesure de l'erreur de couverture du recensement de 1986.

b. Source : Statistique Canada, 1988.

partie de la couverture. Nous nous concentrerons sur le groupe des « citoyens canadiens et immigrants reçus ayant un lieu habituel de résidence au Canada le jour du recensement », puisqu'une forme de vérification du surdénombrement existe déjà pour les personnes appartenant aux deux autres catégories énoncées dans la section 1.

La méthodologie expliquée dans cet article sert à estimer seulement la première composante du surdénombrement. Par conséquent, lorsque nous parlerons de surdénombrement dans ce qui va suivre, il faudra comprendre qu'il s'agit de celui qui est dû au fait qu'une personne incluse dans la couverture de recensement a été dénombrée deux fois ou plus. Il existe trois causes possibles de surdénombrement. Une personne peut être dénombrée dans deux logements différents si : 1) elle a déménagé durant la période de référence du recensement; 2) elle a été dénombrée à un lieu temporaire de résidence le jour du recensement; et 3) elle a été dénombrée à tort par une personne de son entourage (membre ou non de sa famille), qui a mal interprété les instructions concernant les personnes à inclure (même si la personne concernée ne séjournait pas à cet endroit). La méthodologie de l'étude doit permettre d'obtenir des

informations auprès des personnes choisies en ce qui concerne les déménagements, les lieux de séjour le jour du recensement ou les adresses (de personnes apparentées ou non) où il serait plausible de trouver du surdénombrement.

Nous avons décidé de recueillir les informations nécessaires pour déterminer le sous-dénombrement par le truchement d'une enquête post-énumérative. Toutefois, cette enquête sert à plusieurs évaluations. Par conséquent, la stratification utilisée a de multiples objectifs qui ne sont pas reliés au surdénombrement. Pour cette raison, nous ne présentons pas l'allocation ni la sélection de l'échantillon. Disons seulement que la base de sondage est l'ensemble de tous les enregistrements de la base des logements occupés autres que les institutions, les pénitenciers, les hôpitaux, etc. Dès que nous choisissons un logement, toutes les personnes y résidant le jour du recensement sont incluses dans l'échantillon.

Questionnaire et collecte

L'enquête est menée par voie d'entrevues personnalisées. L'interviewer obtient d'abord les noms de toutes les personnes qui habitent le logement sélectionné au moment de l'interview et doit ensuite remplir un questionnaire comportant, entre autres, quatre questions reliées à la couverture du ménage. Ces questions permettent avant tout de connaître la composition du ménage le jour du recensement; elles servent aussi à obtenir des informations sur des adresses où il y aurait pu y avoir du surdénombrement. La semaine de référence de l'étude est la huitième semaine après le recensement.

Traitement

Le traitement comporte trois phases : l'appariement, le dépistage d'adresses et la consultation des fichiers administratifs. Puisque la population visée par l'étude correspond à l'ensemble des personnes dénombrées dans les logements privés le jour du recensement, il faut lier les informations obtenues lors des entrevues aux enregistrements de la base de données du recensement des logements sélectionnés, ces enregistrements composant l'échantillon réel. Les personnes qui figurent dans le questionnaire de l'étude mais non dans celui du recensement sont éliminées pour les fins de l'étude. Celles qui se trouvent dans le questionnaire du recensement

mais non dans le questionnaire de l'enquête sont considérées comme des non-répondants. Deux types de non-réponses sont identifiés. Le premier type (type I) couvre les personnes des logements sélectionnés où il a été impossible d'établir un contact parce que les ménages ont été absents pendant toute la durée de l'enquête. Le deuxième type (type II) couvre, premièrement, les cas où aucune information n'est donnée pour expliquer l'absence d'une personne dans le questionnaire de l'enquête mais où une personne du même ménage se trouve dans les deux questionnaires; et, deuxièmement, les ménages qui ont déménagé entre le jour du recensement et la semaine de référence de l'enquête.

Dès que l'appariement est complété, nous effectuons le dépistage des adresses inscrites dans les questionnaires de l'étude en réponse aux questions sur la couverture. Après quoi nous décidons pour chaque personne choisie si elle est dénombrée ou non. Les non-réponses de type I sont traitées à l'étape de la pondération.

Le deuxième type de non-réponse ne peut pas être traité simplement par voie d'ajustement des facteurs de pondération. Il peut s'agir d'un simple oubli, ou peut-être ces personnes ne résidaient-elles que temporairement dans le logement sélectionné : si tel est le cas, elles ont une plus grande chance d'être surdénombrées. Pour traiter ces cas, nous tirons un échantillon de personnes et nous tentons un appariement avec des dossiers administratifs afin d'obtenir des adresses plausibles. Les résultats tirés de cet échantillon sont inférés à l'ensemble des non-répondants du deuxième type.

Pondération

Puisque toutes les personnes qui sont incluses dans le questionnaire du recensement dans les logements sélectionnés font partie de l'échantillon, nous attribuons à toutes ces personnes le facteur de pondération du logement choisi. Encore une fois, nous faisons deux ajustements des facteurs de pondération. Le premier ajustement sert à traiter les non-réponses de type I. Premièrement, nous transférons (comme pour le sous-dénombrement) les facteurs de pondération des non-réponses de type I aux autres personnes choisies. Après cet exercice, nous transférons les facteurs de pondération des non-réponses de type II qui n'ont pas été échantillonnées de nouveau à celles qui ont fait partie de cette sous-enquête.

Le deuxième ajustement a pour but de calibrer l'échantillon pour qu'il puisse donner, par sommation des facteurs de pondération, le nombre connu de personnes recensées. Cet ajustement se fait au niveau des provinces.

Estimation

Lorsqu'un cas de surdénombrement est décelé, deux interventions sont prévues. Premièrement, si les informations recueillies dans l'enquête permettent de déterminer à coup sûr lequel des deux logements identifiés est la résidence habituelle de la personne choisie, et que cette résidence n'est pas le logement sélectionné pour l'enquête, alors l'indicateur de surdénombrement est fixé à 1 pour cet enregistrement. Si par contre les informations ne permettent pas d'identifier la résidence habituelle, alors l'indicateur est fixé à 0,5 pour le logement sélectionné. La justification de cet algorithme réside dans le fait qu'il faut examiner deux logements, dont celui qui a été sélectionné, pour identifier du surdénombrement. L'autre logement, même s'il n'est pas sélectionné, est représenté tout de même dans l'échantillon du fait de la pondération. Cela signifie que si l'on somme les poids de tous les enregistrements où il y a un surdénombrement, on estime le double du nombre de personnes surdénombrées. Peut-être un exemple illustrera-t-il la justification de l'algorithme. Supposons que notre population contient trois personnes, A, B et C, et que la personne A est surdénombrée dans un recensement. Nous observons en réalité quatre personnes, A, B, C et D (D étant la même personne que A). Supposons que, par manque d'information, nous ne puissions pas déterminer la résidence habituelle des personnes. Prenons un échantillon d'une personne dans la liste du recensement et supposons que nous puissions déterminer à coup sûr si cette personne est surdénombrée. Puisqu'il y a quatre personnes dans la liste, la personne choisie a un facteur de pondération de quatre. Si B ou C est choisie, nous obtenons une estimation du surdénombrement de zéro. Si par contre A ou D est choisie, nous obtenons une estimation de quatre (soit l'indicateur de surdénombrement 1 multiplié par le facteur de pondération 4). En faisant la moyenne des estimations des quatre échantillons possibles (ce qui est, théoriquement, ce que mesure l'enquête), nous obtenons la valeur deux. Donc si nous ne divisons pas les estimations obtenues par deux, nous générons un biais très important.

Remarques

Le traitement de la non-réponse de type II de l'étude post-énumérative peut invalider l'enquête si ce type de non-réponse est fréquent. Il faut alors repenser les questions. Cette méthodologie n'est bonne que dans la mesure où la composition du ménage est déterminée avec une grande précision. Il faut voir que le surdénombrement est une caractéristique très rare, et puisque cette non-réponse englobe de façon intrinsèque une partie du surdénombrement, il faut absolument la minimiser pour que la méthodologie réussisse.

Dans le cas d'une personne énumérée à deux endroits différents, la méthode d'estimation utilisée suppose que de l'information concernant l'autre lieu de résidence sera fournie à chacun des deux endroits où la personne a demeuré. Si tel n'est pas le cas, le nombre total de personnes surdénombrées pourrait être plus élevé que celui qui est estimé.

Pour plus de détails sur la méthodologie et les résultats de l'enquête, voir Statistique Canada, 1990.

CONCLUSION

La contre-vérification des dossiers est jusqu'à maintenant le meilleur outil pour estimer le sous-dénombrement du recensement canadien. Ses estimations constituent la base de la vérification de la qualité des comptes du recensement. Cependant, cette méthodologie a des limites conceptuelles, théoriques et pratiques. Sans doute peut-on encore en améliorer certains aspects, mais même une CVD « améliorée » comportera toujours des difficultés. Néanmoins, les résultats de la CVD sont utiles car ils permettent d'étudier les caractéristiques des personnes non recensées en vue de connaître les causes d'erreurs et de déterminer par le fait même les faiblesses du dénombrement. Par ailleurs, grâce aux personnes choisies qui se révèlent avoir été recensées (provenant de la base de sélection du recensement), il est possible d'évaluer le contenu de certaines questions posées aux deux recensements et de réaliser des analyses longitudinales.

D'un autre côté, Statistique Canada fait des recherches en vue de parvenir à une estimation fiable de l'erreur de couverture nette du recensement de 1991. Il faudra donc prévoir une étude pour établir une estimation du surdénombrement pour toute la

population. L'expérience acquise en 1986 sera utile. Mais mesurer les erreurs de couverture n'est pas chose facile, car on se sert essentiellement pour cela d'études mesurant le biais d'un total ou d'une proportion. Les caractéristiques sur lesquelles portent les études visant à mesurer la couverture (soit les caractéristiques «non recensé» et «surdénombré») sont des événements très rares; c'est pourquoi toutes ces études peuvent être invalidées par un taux de non-réponse non négligeable : car la non-réponse (qui survient par exemple lorsqu'une personne est «non dépistée») est ou peut être très reliée à ces caractéristiques.

RÉFÉRENCES BIBLIOGRAPHIQUES

- COCHRAN, W.M., 1977. *Sampling Techniques*. New York, John Wiley & Sons, 3e édition.
- COWAN, C.D. et R.E. FAY, 1984. «Estimate of Undercount in the 1980 Census». In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 566-571.
- FELLEGI, I.P., 1981. «Should the Census Count be Adjusted for Allocation Purposes?—Equity Considerations». *Current Topics in Survey Sampling*, New York, Academic Press, 47-76.
- STATISTIQUE CANADA, 1988. *Estimations annuelles postcensitaires de la population suivant l'état matrimonial, l'âge, le sexe et composantes de l'accroissement, Canada, provinces et territoires au 1er juin 1988*. Volume 6, 6e édition, catalogue 91-210.
- STATISTIQUE CANADA, 1990. *Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986*, catalogue 99-135F.
- WOLTER, K.M., 1985. *Introduction to Variance Estimation*. New York, Springer-Verlag.

RÉSUMÉ - SUMMARY - RESUMEN**BOUDREAU Jean-René — MESURE DES ERREURS DE COUVERTURE
DU RECENSEMENT CANADIEN**

Les erreurs de couverture sont dans une certaine mesure inévitables dans un recensement de la population, et peuvent avoir de sérieux effets sur les effectifs et les distributions, ainsi que sur les estimations intercensitaires et les projections démographiques. Il est par conséquent important de mesurer la nature et l'ampleur des erreurs de couverture. Plusieurs méthodes ont été conçues à cette fin. Cet article décrit les méthodes utilisées par Statistique Canada pour mesurer les erreurs de couverture du recensement canadien.

**BOUDREAU Jean-René — COVERAGE ERROR MEASUREMENT OF
THE CANADIAN CENSUS**

Coverage errors are to some extent unavoidable in a Census of Population, and can have serious effects on counts, distributions, intercensal estimates and demographic projections. Consequently, it is important to determine the nature and the magnitude of coverage errors. Methods have been designed to achieve this goal. This paper describes the methods used by Statistics Canada to measure coverage errors in the Canadian Census.

**BOUDREAU Jean-René — MEDIDA DE LOS ERRORES DE
COBERTURA DEL CENSO CANADIENSE.**

En cierta medida, los errores de cobertura de un censo de población son inevitables y pueden tener efectos graves sobre los efectivos y su distribución, así como sobre los cálculos intercensales y las proyecciones demográficas. Por lo tanto, es importante medir la naturaleza y amplitud de los errores de cobertura. A este fin, se han concebido varios métodos. El artículo describe los métodos utilizados por Estadística Canadá para medir los errores de cobertura del censo canadiense.