

Article

Erreur de couverture au recensement : une évaluation démographique

par Réjean Lachapelle et Don Kerr

Juin 2000



Statistique
Canada

Statistics
Canada

Canada

Erreur de couverture au recensement : une évaluation démographique

Réjean Lachapelle et Don Kerr¹

Résumé

Le Recensement canadien de 1996 est corrigé en fonction de l'erreur de couverture estimée surtout par la contre-vérification des dossiers (CVD). Les auteurs montrent que de nombreux renseignements supplémentaires tirés de la contre-vérification des dossiers de 1996 ont une valeur immédiate pour l'estimation de la population. En plus de rendre possible une estimation de l'erreur de couverture, les résultats de la contre-vérification des dossiers permettent d'obtenir une autre estimation de la croissance démographique, avec décomposition éventuelle par composante. Cette fonction supplémentaire de la contre-vérification des dossiers est prometteuse pour l'évaluation de l'erreur estimative de couverture au recensement et pour l'élucidation des problèmes possibles d'estimation de composantes choisies du programme des estimations de population.

Mots clés : Erreur de couverture au recensement; estimations de population; contre-vérification des dossiers.

1. Introduction

Différentes formes de contre-vérification des dossiers (CVD) sont utilisées à Statistique Canada depuis les années 1960 pour l'estimation de l'erreur de couverture du Recensement canadien (Fellegi 1969; Brackstone et Gosselin 1973; Gosselin 1976; Burgess 1988; Carter 1990; Royce, Germain, Julien, Dick, Switzer et Allard 1994, Statistique Canada 1999). À l'aide de la CVD, Statistique Canada a préparé une longue série chronologique d'estimations de population, de 1971 jusqu'à présent, qui est intégralement corrigée pour le sous-dénombrement au recensement. Les auteurs montrent ci-dessous que la CVD comporte des renseignements supplémentaires qui, d'un point de vue démographique, peuvent être utilisés à des fins d'estimation de la population.

Le programme de données démographiques de Statistique Canada fait appel à des statistiques de l'état civil, au recensement le plus récent et à différentes sources administratives en vue de la préparation d'estimations de population très exactes et à jour. Des renseignements sur les naissances, les décès, l'immigration et l'émigration, entre autres composantes démographiques, permettent d'estimer la croissance démographique depuis le recensement antérieur. Lors de chaque recensement quinquennal, un cycle prend fin et l'exactitude des estimations est mise à l'essai (Romaniuc 1988). On peut comparer systématiquement ces estimations de la croissance à la croissance estimative obtenue en comparant des recensements subséquents (après correction pour l'erreur de couverture au recensement).

L'interprétation de la différence observée, que l'on appelle traditionnellement l'erreur en fin de période des

estimations de population intercensitaires, est loin d'être évidente. Une erreur importante en fin de période est une indication de problèmes d'estimation de la population, dont la nature précise n'est cependant pas du tout évidente pour ce qui est de savoir quelles composantes démographiques sont responsables de l'erreur. De plus, une évaluation franche de cette erreur en fin de période risque de déceler non seulement des problèmes d'estimation de la population, mais également des lacunes éventuelles des études de couverture au recensement elles-mêmes (au début ou à la fin de la période intercensitaire).

Les auteurs décrivent la possibilité d'une autre estimation de la croissance démographique, fondée explicitement sur les résultats de classification de la CVD. Des renseignements supplémentaires facilitent grandement l'interprétation et la décomposition de cette erreur en fin de période. Trois autres estimations de la croissance démographique pour la période intercensitaire sont présentées ci-dessous, y compris la croissance estimée dans le cadre du programme d'estimations de population, la croissance implicite obtenue par comparaison de recensements consécutifs et la croissance fondée explicitement sur les résultats de classification de la CVD. La section 3 montre comment cette estimation de la croissance fondée sur la CVD facilite la décomposition et l'interprétation de l'erreur en fin de période, et fournit des indications i) de biais dans des composantes choisies des estimations de population et ii) de problèmes éventuels dans les résultats de la CVD. La section 4 présente les résultats de cette décomposition et en décrit brièvement les répercussions tant pour la mesure de l'erreur de couverture au recensement que pour le programme des estimations de population.

1. Réjean Lachapelle, Division de la démographie, immeuble Principal, Parc Tunney, Statistique Canada, Ottawa (Ontario) K1A 0T6; Don Kerr, Department of Sociology, University of Western, London (Ontario) N6A 5C2.

2. Autres estimations de la croissance démographique

2.1 Estimations de la croissance fondées sur des documents administratifs : estimations postcensitaires

Le programme des estimations de population de Statistique Canada comporte l'inscription continue et l'estimation d'événements démographiques, d'après des statistiques de l'état civil et divers ensembles de données administratives. Ces événements sont additionnés ou soustraits de la population documentée lors du recensement antérieur (méthode des composantes). Dans l'estimation de la population d'une province au jour du Recensement de 1996 ($P_{\text{est}96}$) :

$$P_{\text{est}96} = P_{91} + B_{91-96} - D_{91-96} + I_{91-96} - E_{91-96} + \Delta \text{NPR}_{91-96} + \text{NM}_{91-96} \quad (1)$$

La population de base (P_{91}) de cette estimation se fonde sur le Recensement de 1991 après correction pour tout genre d'erreurs de couverture, y compris le sous-dénombrement net au recensement mesuré à l'aide de la CVD de 1991. On peut obtenir l'estimation postcensitaire en additionnant ou en soustrayant de cette valeur de base le nombre de naissances entre les recensements (B_{91-96}), le nombre de décès (D_{91-96}), les immigrants (I_{91-96}), les émigrants (E_{91-96}), la migration interprovinciale nette (NM_{91-96}), de même que le gain net ou la perte nette de résidents non permanents ($\Delta \text{NPR}_{91-96}$).

Les résidents non permanents (RNP) sont des personnes jouissant d'un statut juridique provisoire au Canada (titulaires d'une autorisation d'étude ou d'emploi, titulaires d'un permis ministériel, demandeurs du statut de réfugié et personnes à charge nées à l'extérieur du Canada). Contrairement à la migration interprovinciale, le gain net ou la perte nette de RNP n'est pas estimée à l'aide de données de type « flux » sur les mouvements d'entrée et de sortie de résidents non permanents, mais bien par comparaison dans le temps de données de type « stock » sur le nombre total de résidents non permanents demeurant au pays. On trouvera des renseignements supplémentaires sur la méthode, les sources et la qualité des données dans les publications trimestrielles et annuelles du programme des estimations de population (Statistique Canada 1999, 2000).

2.2 Estimation implicite de la croissance

On peut obtenir une estimation implicite de la croissance à l'aide des recensements de 1991 et de 1996, corrigés en fonction du sous-dénombrement net. Exception faite pour un petit nombre de réserves indiennes qui refusent, dont la population est estimée indépendamment, le sous-dénombrement brut a été estimé entièrement à l'aide de la CVD en 1996, tandis que le surdénombrement brut représente une

estimation combinée de trois études (la CVD, l'Étude sur les logements collectifs et l'Étude par appariement automatisé). En 1991, la CVD a été utilisée uniquement pour l'estimation du sous-dénombrement brut, tandis que le surdénombrement brut a été estimé à l'aide d'une étude plus petite, l'Étude sur les logements privés, combinée aux études par appariement automatisé et sur les logements collectifs de 1991. De plus, les personnes manquantes des réserves indiennes qui ont refusé ont été estimées dans le cadre de la CVD de 1991.

Lors de l'évaluation initiale des études de couverture de 1996, la croissance implicite obtenue à l'aide des corrections ci-dessus a été considérée comme peu réaliste. Depuis, on a pu établir qu'une partie de l'estimation du sous-dénombrement net de 1991 était erronée, et qu'en réalité elle aurait été moins élevée si certaines améliorations méthodologiques avaient été introduites comme en 1996 (Tourigny, Clark et Provost 1998). On a pu montrer i) que le nombre de personnes d'abord classées comme manquantes en 1991 était trop élevé à cause d'une erreur de classification et ii) que l'estimation de 1991 du « surdénombrement » était trop bas. Par conséquent, les estimations de 1991 du sous-dénombrement et du surdénombrement ont été révisées de façon à refléter l'effet de ces modifications méthodologiques (${}^{\text{rev}}U_{91}$, ${}^{\text{rev}}O_{91}$). De plus, pour des raisons de cohérence relativement à 1996, des estimations modélisées distinctes des réserves indiennes qui ont refusé (indépendantes de la CVD) ont été ajoutées au Recensement en 1991.

Plus précisément, on obtient comme croissance implicite (Δ^I) :

$$\begin{aligned} \Delta^I &= P_{96} - P_{91} \\ &= \{P_{96}^c + U_{96} - O_{96} + \text{IR}_{96\text{M}} - \text{IR}_{\text{RRC}96}\} \\ &\quad - \{P_{91}^c + {}^{\text{rev}}U_{91} - {}^{\text{rev}}O_{91} + \text{IR}_{91\text{M}} - \text{IR}_{\text{RRC}91}\} \quad (2) \end{aligned}$$

où la population finale (P_{96} , P_{91}) est obtenue à l'aide de chiffres de recensement publiés antérieurement (P_{96}^c , P_{91}^c) et corrigés en fonction du sous-dénombrement (U_{96} , ${}^{\text{rev}}U_{91}$) et du surdénombrement brut (O_{96} , ${}^{\text{rev}}O_{91}$). Lors de l'ajout d'estimations modélisées indépendamment pour les réserves indiennes qui refusent ($\text{IR}_{96\text{M}}$, $\text{IR}_{91\text{M}}$), il est nécessaire d'éliminer la partie de l'estimation CVD du sous-dénombrement brut qui correspond à ces réserves ($\text{IR}_{\text{RRC}96}$, $\text{IR}_{\text{RRC}91}$). Les résultats présentés ici tiennent compte de ces changements.

2.3.1 Estimations de la croissance fondées sur la CVD

La contre-vérification des dossiers (CVD) est une procédure de couplage et d'appariement des dossiers qui sert à retrouver toutes les personnes d'un échantillon, à les interviewer et à obtenir une adresse pour le jour du recensement, tout en appariant les dossiers en fonction de documents particuliers du recensement. Cela exige la préparation d'un

échantillon pouvant représenter la même population cible que le recensement qu'il s'agit d'évaluer. Cette base de sondage, obtenue tout à fait indépendamment du recensement qu'il s'agit d'évaluer, est établie à l'aide du recensement antérieur, des naissances inscrites au cours de la période intercensitaire, des listes administratives d'immigrants intercensitaires et d'une liste à jour des résidents non permanents. Les personnes manquantes au cours du recensement antérieur sont représentées par un échantillon de cas classés comme « manquants » lors de la CVD antérieure, en l'absence d'une liste complète de ces personnes.

Compte tenu de cet échantillon, la CVD cerne toutes les personnes qui auraient pu faire partie de l'univers du Recensement de 1996. Sauf pour une très faible sous-population d'émigrants de retour (citoyens canadiens et immigrants reçus qui étaient à l'étranger lors du recensement antérieur), l'échantillon de la CVD est complet et tout à fait représentatif. La classification subséquente (manquant, dénombré, émigré, à l'étranger, décédé ou hors du champ d'observation) est appliquée à l'estimation des personnes « manquantes » du recensement en cours. Cette classification offre également la possibilité d'autres inférences, c'est-à-dire d'une autre estimation de la croissance démographique pour la période intercensitaire.

Au moment d'estimer la croissance démographique à l'aide de la CVD, il est utile de considérer les deux équations ci-dessous. Dans la première, la population cible du Recensement de 1991 (P_{91}^T) est exprimée en fonction de tous les résultats possibles de la classification de 1996. Dans la deuxième équation, on peut procéder dans le sens contraire, c'est-à-dire exprimer la population cible du recensement de 1996 (P_{96}^T) en fonction de toutes les situations possibles pour 1991 (ou encore, dans le cas des naissances et des immigrants, pour la période intercensitaire).

$$P_{91}^T = {}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91NP}C_{96PP} + {}^{91PP}D_{96} + {}^{91NP}D_{96} + {}^{91PP}E_{96FR} + {}^{91NP}E_{96FR} + {}^{91NP}E_{96EX} \quad (3)$$

$$P_{96}^T = {}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91NP}C_{96PP} + {}^{91-96}B_{96} + {}^{91EX}I_{96PP} + {}^{91EX}I_{96NP} + {}^{91FR}RE_{96PP} \quad (4)$$

où :

- ${}^{91}PP_{96}$ - les citoyens canadiens et les immigrants reçus au Canada en 1991, également ciblés par le Recensement de 1996
- ${}^{91}NP_{96}$ - les RNP au Canada en 1991, également ciblés par le Recensement de 1996 à titre de RNP
- ${}^{91NP}C_{96PP}$ - les RNP au Canada en 1991 qui sont devenus des immigrants reçus au cours de la période intercensitaire

- ${}^{91PP}D_{96}$ - les citoyens canadiens et les immigrants reçus au Canada en 1991 qui sont décédés au cours de la période intercensitaire
- ${}^{91NP}D_{96}$ - les RNP au Canada en 1991 qui sont décédés au cours de la période intercensitaire
- FR - les personnes autorisées à vivre en permanence au Canada (citoyens et immigrants reçus) qui ne font pas partie de la population cible du recensement en question
- ${}^{91PP}E_{96FR}$ - les citoyens canadiens et les immigrants reçus au Canada en 1991 qui ne font pas partie de la population cible du Recensement de 1996
- ${}^{91NP}E_{96FR}$ - les RNP au Canada en 1991 qui sont devenus des immigrants reçus ou des citoyens et qui ne font pas partie de la population cible du Recensement de 1996
- EX - les personnes qui n'ont jamais été citoyens ou immigrants reçus et qui ne font pas partie de la population cible du recensement en question
- ${}^{91NP}E_{96EX}$ - les RNP au Canada en 1991 qui ne sont pas devenus des immigrants reçus et qui ne font pas partie de la population cible du Recensement de 1996
- ${}^{91-96}B_{96}$ - les personnes nées au cours de la période 1991-1996 qui font partie de la population cible du Recensement de 1996
- ${}^{91EX}I_{96NP}$ - les personnes qui n'étaient pas au Canada en 1991, qui sont arrivées au cours de la période intercensitaire et qui sont des RNP de la population cible du Recensement de 1996
- ${}^{91EX}I_{96PP}$ - les immigrants qui ont été reçus au cours de la période intercensitaire et qui font partie de la population cible du Recensement de 1996
- ${}^{91FR}RE_{96PP}$ - les immigrants de retour, c'est-à-dire les citoyens canadiens et les immigrants reçus hors de l'univers du recensement en 1991 et faisant partie de l'univers du Recensement de 1996

On peut obtenir une estimation de la croissance (Δ^{RRC}) en soustrayant la première équation de la deuxième :

$$\begin{aligned} \Delta^{RRC} = & {}^{91-96}B_{96} + {}^{91EX}I_{96PP} + {}^{91EX}I_{96NP} \\ & - {}^{91PP}D_{96} - {}^{91NP}D_{96} - {}^{91PP}E_{96FR} - {}^{91NP}E_{96FR} \\ & - {}^{91NP}E_{96EX} + {}^{91FR}RE_{96PP}. \end{aligned} \quad (5)$$

À l'aide des résultats de classification et des bases de sondage introduites antérieurement, il est possible d'estimer, directement à partir de la CVD de 1996 elle-même, tous les termes (exception faite pour le dernier : les

émigrants de retour). On peut obtenir une approximation de la population cible du recensement de 1991 à l'aide de l'échantillon tiré des bases de sondage du recensement et des personnes manquantes – avec identification des résultats de classification pertinents. On peut obtenir une approximation de la population cible du recensement de 1996 à l'aide de toutes les personnes classées parmi les unités dénombrées ou manquantes en 1996. On peut obtenir le terme final (les émigrants de retour) indépendamment de la CVD à l'aide de la variable de mobilité quinquennale du Recensement de 1996, en identifiant toutes les personnes qui étaient à l'extérieur du pays il y a cinq ans (à l'exclusion des immigrants récents et des RNP). On peut exprimer cette même estimation de la croissance démographique, fondée sur la CVD, au niveau provincial en incorporant une estimation de la migration interprovinciale. Puisque la CVD se fonde sur des fichiers de soins de santé pour les deux territoires nordiques du Canada (le Yukon et les TN-O) avec des listes administratives d'adresses courantes pour le recensement en question, cette estimation de la croissance n'est pas possible pour les populations relativement peu nombreuses du Grand Nord canadien.

Il subsiste dans la CVD un problème mineur qui risque d'introduire un faible biais dans les résultats de classification. Malheureusement, il n'est pas possible d'identifier tous les RNP dans l'échantillon de la CVD, d'où le risque d'un chevauchement inconnu des bases de sondage (entre les bases de sondage du recensement, des RNP et des immigrants). Puisque les RNP, dans le recensement, ne peuvent être identifiés qu'à l'aide du questionnaire détaillé du recensement, qui est distribué à 20% environ des ménages, il est possible que quelques RNP demeurant au Canada en 1991 et sélectionnés dans la base de sondage du recensement ont également été choisis soit dans la base des immigrants, soit dans celle des RNP sans avoir été identifiés comme tels. Bien que l'on tente de corriger la CVD pour un tel chevauchement, en identifiant toute personne de ce genre dans les bases des immigrants et des RNP, il existe un biais inconnu dans la mesure où cette démarche est infructueuse. Cette difficulté d'élucidation du chevauchement entraîne la possibilité d'un nombre trop élevé d'immigrants et/ou de RNP dans l'échantillon, ou d'un nombre trop faible, si l'on élimine trop de personnes des bases de sondage mentionnées. Ce dernier résultat risque de provoquer une déflation subséquente de l'estimation de la croissance, démographique et du sous-dénombrement brut (parmi d'autres résultats de classification), tandis que la première possibilité entraîne le résultat contraire.

2.3.2 Estimation de la croissance fondée sur la CVD : une décomposition plus détaillée

Même si les estimations de la croissance démographique tant postcensitaire que fondée sur la CVD devraient être très comparables, les termes particuliers au sein de chacune ne sont pas censés être directement équivalents. Ainsi, les

naissances dans les estimations postcensitaires représentent toutes les naissances intercensitaires relevant d'une population – peu importe si les personnes nées se déplacent ou meurent – tandis que les naissances dans une équation discrète représentent toutes les personnes nées et demeurant au Canada à la fin de la période intercensitaire. Dans un tel contexte, il est possible d'élargir l'équation fondée sur la CVD de façon à obtenir des termes qui sont plus comparables à ceux figurant dans les estimations postcensitaires. Dès lors, l'estimation de la croissance démographique fondée sur la CVD peut servir à évaluer les composantes de la croissance démographique figurant dans la méthode des composantes.

Pour élargir cette équation, il convient de commencer par les naissances, exprimées encore une fois en fonction des résultats de classification possibles de la CVD. Comme il a été mentionné antérieurement, la naissance comme terme de l'équation (5) représente une partie seulement des naissances survenues au cours de la période intercensitaire. De façon plus générale, on peut exprimer toutes les naissances comme suit :

$$B^{91-96} = {}^{91-96}B_{96} + {}^{91-96B}D_{96} + {}^{91-96B}E_{96FR} \quad (6)$$

où :

B^{91-96} = toutes les naissances intercensitaires

${}^{91-96}B_{96}$ = toutes les naissances intercensitaires classées à terme parmi les unités dénombrées ou manquantes en 1996

${}^{91-96B}D_{96}$ = décès de naissances intercensitaires

${}^{91-96B}E_{96FR}$ = personnes hors de la population cible en 1996 mais nées au Canada au cours de la période intercensitaire

De même, on peut exprimer tous les immigrants comme suit :

$$I^{91-96} = {}^{91EX}I_{96PP} + {}^{91NP}C_{96PP} + {}^{91-96I}D_{96} + {}^{91-96I}E_{96FR} \quad (7)$$

où :

${}^{91EX}I_{96PP}$ = immigrants intercensitaires classés à terme parmi les unités dénombrées ou manquantes en 1996

${}^{91NP}C_{96PP}$ = tous les RNP en 1991 qui deviennent immigrants reçus et qui sont classés à terme parmi les unités dénombrées ou manquantes en 1996

${}^{91-96I}D_{96}$ = décès d'immigrants reçus au cours de la période intercensitaire

${}^{91-96I}E_{96FR}$ = émigrants parmi les immigrants intercensitaires (peu importe s'ils demeureraient ou non au Canada à titre de RNP en 1991)

Si l'on combine les équations 5, 6 et 7, on peut réexprimer la croissance démographique comme suit :

$$\begin{aligned}
 P_{96}^T - P_{91}^T &= B^{91-96} - {}^{91PP}D_{96} - {}^{91NP}D_{96} - {}^{91-96B}D_{96} \\
 &\quad - {}^{91-96I}D_{96} + I^{91-96} + {}^{91FR}RE_{96PP} - {}^{91NP}C_{96PP} \\
 &\quad - {}^{91PP}E_{96FR} - {}^{91NP}E_{96FR} - {}^{91-96B}E_{96FR} \\
 &\quad - {}^{91-96I}E_{96FR} - {}^{91NP}E_{96EX} - {}^{91EX}I_{96NP}. \quad (8)
 \end{aligned}$$

Le terme final de l'équation (8) correspond à :

$$\begin{aligned}
 {}^{91EX}I_{96NP} &= NP_{96} - NP_{91} + {}^{91NP}D_{96} + {}^{91NP}E_{96EX} \\
 &\quad + {}^{91NP}C_{96PP} + {}^{91NP}E_{96FR}. \quad (9)
 \end{aligned}$$

Par conséquent :

$$\begin{aligned}
 P_{96}^T - P_{91}^T &= \\
 &B^{91-96} - {}^{91PP}D_{96} - {}^{91NP}D_{96} - {}^{91-96B}D_{96} - {}^{91-96I}D_{96} \\
 &+ I^{91-96} + {}^{91FR}RE_{96PP} - {}^{91NP}C_{96PP} \\
 &- {}^{91PP}E_{96FR} - {}^{91NP}E_{96FR} - {}^{91-96B}E_{96FR} \\
 &- {}^{91-96I}E_{96FR} - {}^{91NP}E_{96EX} + NP_{96} \\
 &- (NP_{91} - {}^{91NP}D_{96} - {}^{91NP}E_{96EX} - {}^{91NP}C_{96PP} - {}^{91NP}E_{96FR}) \quad (10)
 \end{aligned}$$

ou :

$$\begin{aligned}
 P_{96}^T - P_{91}^T &= \\
 &(B^{91-96}) - ({}^{91PP}D_{96} - {}^{91-96B}D_{96} - {}^{91-96I}D_{96}) + (I^{91-96}) \\
 &- ({}^{91PP}E_{96FR} + {}^{91-96B}E_{96FR} + {}^{91-96I}E_{96FR} - {}^{91FR}RE_{96PP}) \\
 &+ (NP_{96} - NP_{91}). \quad (11)
 \end{aligned}$$

Cette version élargie de l'équation (5) donne lieu à une ventilation de la croissance démographique au niveau national, et permet des comparaisons plus utiles avec des composantes estimées à l'aide de documents administratifs. Tous les termes sauf ${}^{91FR}RE_{96PP}$ et NP_{91} peuvent être obtenus directement de la CVD de 1996. La lacune susmentionnée pour la base de sondage de la CVD nécessite une estimation indépendante des émigrants de retour, tandis que la nature de la base de sondage pour les RNP explique l'absence du dernier terme. La CVD ne se fonde pas sur une liste de tous les RNP arrivant au Canada au cours de la période intercensitaire (comme c'était le cas pour les immigrants), mais bien sur la liste administrative la plus à jour des RNP en vue de l'établissement de sa base de sondage (sans information sur le nombre de RNP demeurant au Canada en 1991).

Les estimations postcensitaires permettent de documenter la croissance démographique grâce à l'inscription « continue » et à l'estimation d'événements démographiques au fil des ans. La CVD estime la croissance à l'aide d'information sur l'état des unités identifiées lors de deux dates « discrètes » au moins (au début et à la fin de la période intercensitaire). Malgré cette distinction conceptuelle mineure entre l'estimation « continue » et l'estimation « discrète », chaque terme de l'équation 11 (chaque couple de parenthèses) correspond plus ou moins à une composante distincte documentée à l'aide de dossiers administratifs. Le premier terme sert à identifier toutes les naissances intercensitaires (la somme pondérée de la base des naissances); le deuxième terme englobe les décès (résultats de classification pour la base des naissances, la base des personnes manquantes, la base du recensement et la base des immigrants); le troisième terme comprend tous les immigrants (la somme pondérée de la base des immigrants); le quatrième terme comprend les émigrants (résultats de classification pour la base des naissances, la base des immigrants, la base des personnes manquantes et les bases du recensement, de même que la composante des émigrants de retour); le cinquième terme correspond au gain net ou à la perte nette de RNP. Puisque le nombre de RNP demeurant au Canada en 1991 n'est pas disponible dans la CVD de 1996, pour le moment, ce dernier terme est obtenu à l'aide du dénombrement de 1991, après correction pour le sous-dénombrement. Encore une fois, il est possible d'exprimer cette équation au niveau provincial.

L'équation (11) permet une évaluation détaillée du programme des estimations postcensitaires. Ainsi, s'il subsiste des différences entre les estimations fondées sur la CVD et les estimations postcensitaires, il est possible de déterminer dans quelle mesure les différences de croissance estimative sont attribuables à des différences de migration (typiquement assez difficiles à estimer dans le programme des estimations postcensitaires) et dans quelle mesure elles sont attribuables à des différences d'accroissement naturel. Bref, le tableau 1 englobe toutes les estimations de la croissance mentionnées ci-dessus, y compris la croissance implicite, la croissance fondée sur des documents administratifs et les deux autres estimations de la croissance fondées sur la CVD (équations simplifiée et élargie). Il existe de légères différences entre l'équation simplifiée et l'équation élargie – mais pas du tout de la même importance que pour les autres estimations (implicite, postcensitaire). En guise d'explication des différences entre les deux estimations fondées sur la CVD, disons que l'équation simplifiée ne suppose pas la même classification détaillée que l'équation élargie; elle n'est pas aussi biaisée par le problème du chevauchement des bases de sondage mentionné ci-dessus, et elle ne s'appuie pas sur le dénombrement des RNP du Recensement de 1991. Les différences observées pour les estimations qui restent font l'objet de la décomposition en cours.

Tableau 1
Autres estimations de la croissance,
1991-1996, Canada et les provinces et territoires

	Croissance implicite	Estimations de la population documents administratifs	CVD simplifiée	CVD élargie
T.-N.	-17 997	-9 263	-17 897	-17 751
Î.-P.-É.	5 404	5 483	2 568	1 583
N.-É.	15 781	24 271	17 075	16 860
N.-B.	7 714	13 097	12 017	11 276
QUÉBEC	206 307	300 849	261 357	252 014
ONTARIO	659 349	766 568	668 443	655 572
MAN.	23 682	24 981	7 377	6 288
SASK.	15 953	11 098	11 524	9 312
ALB.	186 594	186 986	151 944	159 907
C.-B.	505 025	466 611	465 864	472 342
YUKON	3 085	2 329	N/A	N/A
T.N.-O.	6 837	5 864	N/A	N/A
Provinces (sauf les terr.)	1 607 771	1 790 681	1 580 273	1 567 404
Canada	1 617 693	1 798 874	N/A	N/A

3. Décomposition de l'erreur en fin de période

La croissance implicite pour la période 1991-1996 est obtenue uniquement après correction des recensements pour l'erreur de couverture, y compris les chiffres révisés de 1991 sur le sous-dénombrement et le surdénombrement bruts et les raffinements pour les réserves indiennes qui refusent. D'autre part, l'estimation de la croissance fondée sur la CVD (version simplifiée) est obtenue à l'aide d'approximations des populations cibles de 1991 et de 1996, c'est-à-dire des bases de sondage des recensements et des personnes manquantes de la CVD de 1996 et de toutes les personnes classées parmi les unités manquantes ou dénombrées dans cette étude. C'est pourquoi il existe entre les deux estimations des différences mineures qu'il y a lieu de cerner plus clairement dans une décomposition intégrale de l'erreur en fin de période. À cet égard, il est utile d'exprimer la croissance implicite obtenue à l'aide de chiffres définitifs de la population en fonction de ces approximations (bases de sondage et résultats de classification). De même, puisque l'erreur en fin de période représente la différence entre la croissance implicite et la croissance associée aux estimations postcensitaires, l'erreur en fin de période peut également être exprimée en fonction de ces approximations.

Par souci de simplification, soit δ , tous les termes de croissance négative possibles de l'équation (5) et η tous les termes de croissance positive possibles :

$$\delta =$$

$$({}^{91PP}D_{96} + {}^{91NP}D_{96} + {}^{91PP}E_{96FR} + {}^{91NP}E_{96FR} + {}^{91NP}E_{96EX}) \quad (12)$$

$$\eta = ({}^{91-96}B_{96} + {}^{91EX}I_{96PP} + {}^{91EX}I_{96NP} + {}^{91FR}RE_{96PP}) \quad (13)$$

Les populations dénombrées dans les deux recensements se laissent représenter comme suit :

$${}^{91}P_{96} = ({}^{91}PP_{96} + {}^{91}NP_{96} + {}^{91NP}C_{96PP}). \quad (14)$$

Puisque les chiffres définitifs de la population (P_{91} , P_{96}) utilisés dans l'estimation de la croissance implicite comportent des estimations modélisées distinctes pour les réserves indiennes qui refusent, il est utile de réexprimer l'estimation de la croissance fondée sur la CVD après avoir délimité ces réserves bien précisément. En cernant les personnes demeurant dans des réserves qui ont refusé en 1996 et qui faisaient partie de la population cible en 1991 comme des ${}^{91}IR_{96}$, en exprimant la croissance de ces réserves sous l'effet de la migration ou des naissances suivant l'estimation de la CVD sous la forme η_{IR} , et en redéfinissant ${}^{91}P_{96}$ de façon à exclure toutes les personnes associées à ces deux termes, il est possible de retourner aux équations (3)-(5) sous la forme :

$$P_{91}^T = {}^{91}P_{96} + {}^{91}IR_{96} + \delta \quad (15)$$

$$P_{96}^T = {}^{91}P_{96} + {}^{91}IR_{96} + \eta_{IR} + \eta \quad (16)$$

$$\Delta^{RRC} = P_{96} - P_{91} = \eta + \eta_{IR} - \delta. \quad (17)$$

Lorsqu'on exprime la croissance implicite en fonction des bases de sondage et des résultats de classification de la CVD, il est utile de s'appuyer sur l'estimation de la croissance de la CVD (équation 17) pour définir la croissance totale en commençant par P_{91} au lieu de P_{91}^T . Si l'on reconnaît que l'estimation définitive de la population (P_{91}) équivaut aux bases de sondage du recensement et des personnes manquantes (P_{91}^T) moins le surdénombrement (${}^{REV}O_{91}$) plus les raffinements pour les réserves indiennes qui refusent ($IR_{91M} - IR_{RRC91}$), on a :

$$P_{96}^T - P_{91} = \eta + n_{IR} - \delta + {}^{REV}O_{91} + (IR_{RRC91} - IR_{91M}). \quad (18)$$

Par contre, cette population cible (P_{96}^T) peut également s'exprimer comme suit :

$$P_{96}^T = EN_{96} + U_{96} + {}^{91FR}RE_{96PP} \quad (19)$$

où EN_{96} est une estimation du nombre de personnes recensées en 1996. En rappelant selon l'équation 2 que :

$$P_{96} = P_{96}^c + U_{96} - O_{96} + (IR_{96M} - IR_{RRC96}) \quad (20)$$

on peut exprimer la croissance implicite (Δ^I) en fonction des estimations de la croissance fondées sur la CVD, comme suit :

$$\begin{aligned} \Delta^I = P_{96} - P_{91} &= (P_{96} - P_{96}^T) + (P_{96}^T - P_{91}) = \{(\eta - \delta)\} \\ &+ \{\eta_{IR} - (IR_{91M} - IR_{RRC91}) + (IR_{96M} - IR_{RRC96})\} \\ &+ \{(P_{96}^c - EN_{96} - {}^{91FR}RE_{96PP} + {}^{REV}O_{91} - O_{96})\}. \end{aligned} \quad (21)$$

La croissance implicite (Δ^I) peut se définir comme la somme i) d'une estimation de la croissance fondée sur la CVD (à l'exclusion des réserves indiennes qui refusent), ii) d'un second terme qui dépend de la décision d'estimer les réserves indiennes qui refusent à l'aide d'un modèle indépendant et iii) d'un troisième terme comportant une comparaison entre l'estimation des personnes recensées d'après la CVD et le nombre de personnes effectivement recensées en 1996.

Ce dernier terme (la différence pour ce qui est des personnes recensées) donne lieu à une interprétation intéressante, et il est considéré comme un élément essentiel de l'évaluation de la CVD (Tourigny, Bureau et Clark 1998; Royce 1993). Des différences appréciables pour ce terme laissent supposer soit des erreurs d'échantillonnage et/ou un biais, soit une erreur de classification et/ou des problèmes de tirage de l'échantillon. Pour que cette comparaison soit utile, on élimine le surdénombrement de 1996 et l'estimation des émigrants de retour, puisque ni l'un ni l'autre ne peut faire partie de l'estimation des personnes recensées. De même, puisque la CVD tire une partie de son échantillon du recensement antérieur, elle reporte inévitablement un certain surdénombrement inhérent à ses poids, qu'il faut ensuite éliminer de son estimation des personnes recensées. Ces corrections sont comprises dans le troisième terme (le troisième couple de crochets) de l'équation 21.

Il y a inflation de l'estimation des personnes recensées sous l'effet des poids associés au surdénombrement de la base de sondage du Recensement de 1991, mais une partie seulement est associée directement à cette estimation, le reste étant réparti parmi les autres résultats de classification. Par conséquent, tous les résultats de classification figurant dans les équations mentionnées ci-dessus sont aussi légèrement surestimés. Pour la présente décomposition, cette distinction mineure est laissée de côté. C'est là une autre raison, bien que mineure, pour laquelle l'estimation de la croissance fondée sur la CVD est différente de l'estimation implicite, cette dernière n'étant pas biaisée par un tel surdénombrement.

D'après ce qui précède, l'erreur en fin de période équivaut à :

$$\begin{aligned} \Delta_{91-96}^D - \Delta_{91-96}^I = & \\ & [\Delta_{91-96}^D - \{(\eta - \delta)\} \\ & - \{\eta_{IR} - (IR_{91M} - IR_{RRC91}) + (IR_{96M} - IR_{RRC96})\}] \\ & - [\{(P_{96}^c - EN_{96} - {}^{91FR}RE_{96PP} + {}^{rev}O_{91} - O_{96})\}]. \quad (22) \end{aligned}$$

Dans la décomposition de l'erreur en fin de période, le premier terme entre crochets [] souligne la différence entre l'estimation postcensitaire de la croissance et l'estimation de la croissance combinée de la CVD (englobant les réserves qui refusent, avec raffinements pour les estimations modélisées). Le deuxième terme (la différence pour ce qui est des personnes recensées) met en évidence les difficultés éventuelles des études de couverture. Théoriquement, en l'absence d'erreur d'échantillonnage et d'erreur non due à l'échantillonnage dans la CVD, ce dernier terme devrait être négligeable.

4.1 Résultats de la décomposition : erreur en fin de période

Le tableau 2 présente l'erreur en fin de période lorsque la population de 1991 et celle de 1996 ont été estimées. Si l'on ajoute le sous-dénombrement net aux chiffres publiés du Recensement de 1996, de même que les estimations indépendantes des réserves indiennes qui refusent, la population du Canada au jour du Recensement de 1996, corrigée pour l'erreur de couverture, est estimée à 29 619 539. Ce chiffre est appréciablement inférieur à l'estimation pour le jour du recensement préparée dans le cadre du programme des estimations postcensitaires : 29 800 720. La différence entre les deux chiffres, qui correspond à la différence mentionnée ci-dessus entre la croissance implicite et la croissance fondée sur des documents administratifs, est plus élevée que prévu d'après l'expérience passée, atteignant 181 181 (c'est-à-dire 0,61% de la population au jour du Recensement de 1991).

Parmi les provinces et les territoires, l'erreur en fin de période est particulièrement marquée à Terre-Neuve (1,56%), dans le Nord canadien (-2,38% au Yukon et -1,44% dans les T.N.-O.) et, chose assez étonnante, dans les trois provinces les plus importantes (1,30% au Québec, 0,97% en Ontario et -0,99% en Colombie-Britannique). Dans les régions, on observe des erreurs en fin de période plus grandes que la moyenne nationale dans l'Est et le Centre du Canada (sauf en Î.-P.-É.), tandis que les provinces de l'Ouest comportent des erreurs en fin de période inférieures à la moyenne nationale. Ce sont précisément ces erreurs que la décomposition courante cherche à évaluer et à expliquer.

Le tableau 3 présente les résultats de cette décomposition, selon i) la différence entre l'estimation de la croissance fondée sur des documents administratifs et l'estimation fondée sur la CVD (version simplifiée) et ii) la différence pour les personnes recensées. On y trouve également l'erreur-type associée aux estimations de la CVD.

4.2 Comparaison des estimations de la croissance

Dans toutes les provinces (sauf la Saskatchewan), la croissance estimée en fonction de documents administratifs est plus élevée que ne l'indique l'estimation fondée sur la CVD. À l'échelle nationale (sauf les territoires), cette divergence en matière de croissance (210 408) semble beaucoup plus importante pour l'explication de l'erreur en fin de période que la divergence en matière de personnes recensées (-27 498). Pour de nombreuses provinces, la différence de croissance tombe bel et bien dans les limites prévues en fonction de l'erreur d'échantillonnage, mais pour d'autres provinces il faut trouver une meilleure explication. Ainsi, la différence de croissance en Ontario est importante (98 125), représentant près de la moitié de la différence observée à l'échelle nationale. De même, Terre-Neuve, le Québec, l'Alberta et le Manitoba, pris ensemble, expliquent une bonne partie de cette différence.

À titre d'indication des facteurs responsables de ces différences, le tableau 4 présente des comparaisons d'après

l'équation 11 (équation détaillée). D'autres estimations sont fournies pour i) les naissances, ii) les décès, iii) l'immigration, iv) l'émigration, v) la migration interprovinciale et vi) le changement net du nombre de résidents non permanents. Les problèmes majeurs de l'explication de l'erreur en fin de période sont évidents dans le tableau 4, en particulier pour l'émigration.

Puisque le Canada ne possède pas de système complet d'inscription à la frontière, l'émigration est clairement l'élément le plus faible du programme des estimations de population. En l'absence de renseignements immédiats sur le nombre de personnes quittant le Canada, la CVD, avec ses procédures exhaustives de dépistage, de couplage des

dossiers et d'interviews directes, est considérée comme supérieure à toute autre source de données actuellement accessible. Bien que la CVD comporte certains problèmes comme le chevauchement des bases de sondage mentionné antérieurement, l'évaluation en cours révèle une erreur manifeste dans les estimations postcensitaires, c'est-à-dire une sous-estimation des personnes quittant le Canada. Dans l'ensemble, la différence observée à l'échelle nationale (206 228) explique en grande partie l'erreur en fin de période documentée en 1996. De même, pour l'Ontario, les difficultés d'estimation de l'émigration semblent être fondamentales (la différence atteignant 119 947).

Tableau 2

Résultats des études de couverture relativement à l'estimation de la population (1996 – jour du recensement)

	{1}	{2}	{3}	{4 = 1 + 2 + 3}	{5}	{6 = 5 - 4}	{7 = 6/4 * 100}
	Dénombrement de 1996 avec ajouts aléatoires	Sous-dénombrement net de 1996	Réserves indiennes	Recensement de 1996 corrigé pour la CVD	Estimation postcensitaire de 1996 (i)	Erreur en fin de période	Erreur en fin de période (%)
T.-N.	551 792	9 424	0	561 216	569 950	8 734	1,56
Î.-P.-É.	134 557	1 149	175	135 881	135 960	79	0,06
N.-É.	909 282	20 821	0	930 103	938 593	8 490	0,91
N.-B.	738 133	14 225	518	752 876	758 259	5 383	0,71
QUÉBEC	7 138 795	116 750	12 427	7 267 972	7 362 514	94 542	1,30
ONTARIO	10 753 573	301 368	20 849	11 075 790	11 183 050	107 260	0,97
MAN.	1 113 898	18 881	315	1 133 094	1 134 393	1 299	0,11
SASK.	990 237	28 051	586	1 018 874	1 014 019	-4 855	-0,48
ALB.	2 696 826	66 327	11 287	2 774 440	2 774 832	392	0,01
C.-B.	3 724 500	142 443	3 136	3 870 079	3 831 665	-38 414	-0,99
YUKON	30 766	1 022	0	31 788	31 032	-756	-2,38
T.N.-O.	64 402	3 024	0	67 426	66 453	-973	-1,44
Canada	28 846 761	723 485	49 293	29 619 539	29 800 720	181 181	0,61

(i) Estimations postcensitaires du 14 mai obtenues avec des composantes définitives pour les estimations intercensitaires.

Estimations définitives (24 septembre 1998) du sous-dénombrement net, 1991 et 1996.

Tableau 3

Décomposition de l'erreur en fin de période

Province/Territoire	Erreur en fin de période	Différence entre		Différence pour les pers. rec.	Erreur-type des estimations
		Estimation de la croissance Dém. et CVD	Erreur-type des estimations		
T.-N.	8 734	8 634	4 889	100	5 176
Î.-P.-É.	79	2 915	2 425	-2 836	2 462
N.-É.	8 490	7 196	9 011	1 294	9 455
N.-B.	5 383	1 080	7 793	4 303	7 918
QUÉBEC	94 542	39 492	25 493	55 050	29 310
ONTARIO	107 260	98 125	41 212	9 135	51 300
MAN.	1 299	17 604	10 108	-16 305	10 370
SASK.	-4 855	-426	9 187	-4 429	10 200
ALB.	392	35 042	19 067	-34 650	21 618
C.-B.	-38 414	747	20 518	-39 161	22 996
YUKON	-756	N/A	N/A	-108	270
T.N.-O.	-973	N/A	N/A	-284	464
Canada sauf les terr.	18 2910	210 408	43 951	-27 498	58 724
Canada	18 1181	N/A	N/A	-27 890	58 762

Tableau 4
Composantes estimées (1991-1996) compilées par la Division de la démographie et mesure discrète (détaillée) de la CVD

	T.-N.	Î.-P.-É.	N.-É.	N.-B.	QUE	ONT	MAN	SASK	ALB	C.-B.	CANADA (sauf les terr.)
Naissances											
Démographie	31 748	8 803	55 994	44 444	453 556	730 520	81 485	70 382	199 484	229 511	1 905 927
CVD	31 779	8 782	55 984	44 444	454 332	729 744	81 485	70 382	199 484	229 511	1 905 927
Différence	-31	22	10	0	-776	776	0	0	0	0	0
Décès											
Démographie	-19 286	-5 692	-37 677	-28 567	-252 628	-376 760	-45 858	-40 652	-75 798	-126 935	-1 009 853
CVD	-18 530	-6 913	-43 820	-29 354	-273 617	-400 047	-56 108	-40 143	-74 640	-138 433	-1 081 605
Différence	-756	1 221	6 143	787	20 989	23 287	10 250	-509	-1 158	11 498	71 752
Immigration											
Démographie	3 411	771	14 489	3 359	189 905	618 869	22 004	11 282	84 130	213 506	1 161 726
CVD	3 538	820	14 058	3 614	189 905	618 870	22 129	11 157	84 130	216 892	1 165 113
Différence	-127	-49	431	-255	0	-1	-125	125	0	-3 386	-3 387
Émigration											
Démographie	-671	-206	-2 297	-2 429	-15 490	-48 609	-5 684	-2 493	-19 718	-17 834	-115 431
CVD	-2,227	-455	-7 334	-3 889	-55 766	-168 556	-10 871	-7 133	-33 689	-31 739	-321 659
Différence	1,556	249	5 037	1 460	40 276	119 947	5 187	4 640	13 971	13 905	206 228
Migration interprovinciale											
Démographie	-23 074	1 643	-5 288	-3 255	-51 176	-40 850	-25 336	-26 644	7 155	167 809	984
CVD	-32 767	-886	-1 479	-2 933	-49 395	-37 505	-29 765	-25 095	-10 321	191 222	1 076
Différence	9 693	2 529	-3 809	-322	-1 781	-3 345	4 429	-1 549	17 476	-23 413	-92
Résidents non permanents											
Démographie	-1 406	164	-950	-455	-23 353	-116 602	-1 630	-777	-8 267	554	-152 722
CVD	455	236	-549	-606	-13 445	-86 934	-582	144	-5 057	4 890	-101 448
Différence	-1 861	-72	-401	151	-9 908	-29 668	-1 048	-921	-3 210	-4 336	-51 274
Total											
Démographie	-9 263	5 483	24 271	13 097	300 849	766 568	24 981	11 098	186 986	466 611	1 790 681
CVD	-17 751	1 583	16 860	11 276	252 014	655 572	6 288	9 312	159 907	472 343	1 567 404
Différence	8 488	3 900	7 411	1 821	48 835	110 996	18 693	1 786	27 079	-5 731	223 277

Sans être décisive, la décomposition courante indique également qu'il existe d'autres composantes problématiques à part l'émigration pour ce qui est de l'explication de l'erreur en fin de période pour certaines provinces. Ainsi, les résultats indiquent que les estimations de la migration interprovinciale comportent des erreurs pour la Colombie-Britannique et Terre-Neuve compte tenu des différences observées pour ces composantes et les erreurs en fin de période correspondantes. Dans l'ensemble, l'acceptation de la CVD relativement à ces migrations plus difficiles à estimer permettrait d'expliquer non seulement une bonne partie de cette différence de croissance, mais également une bonne partie de l'erreur en fin de période de 1996. Pour ce qui est de l'erreur en fin de période qui reste, il convient de se pencher sur la différence observée pour les personnes recensées. Par le fait même, on s'éloigne des difficultés éventuelles des estimations postcensitaires.

4.3 Comparaison des estimations de personnes recensées

Bien que la différence pour les personnes recensées que l'on observe à l'échelle nationale soit beaucoup plus faible que la différence documentée pour ce qui est de la croissance, pour la moitié environ des provinces, cette différence est de taille comparable sinon supérieure. Quant à l'interprétation de cette constatation, il est reconnu que la

CVD n'a jamais été conçue en fonction de la population « recensée ». La priorité étant accordée à la documentation du nombre de « personnes manquantes » dans le recensement, le plan d'échantillonnage de la CVD comporte une surreprésentation des « groupes difficiles à recenser » (les jeunes adultes célibataires par exemple), de même qu'une sous-représentation des personnes « faciles à recenser ». Dans l'ensemble, la comparaison pour les personnes recensées confirme l'exactitude de la CVD, les différences d'une province ou d'un territoire à l'autre n'étant pas significatives. Néanmoins, les différences constatées pour quelques provinces sont préoccupantes, et se rapprochent beaucoup du seuil statistiquement significatif à 95% au Québec (différence positive), de même que du seuil statistiquement significatif en Colombie-Britannique, en Alberta et au Manitoba (différences négatives).

Quant à l'évaluation des résultats des études de couverture de 1991, deux autres hypothèses ont été formulées pour expliquer les différences observées pour les personnes recensées (Royce 1993). À l'une des extrémités, on pourrait insister pour que toute la différence (pour une province donnée) soit expliquée en fonction de la représentativité de l'échantillon de la CVD, laissant supposer une erreur d'échantillonnage ou des lacunes quelconques dans la base de sondage. À l'autre extrémité, on pourrait insister pour que toute la différence soit attribuée à l'absence de

documentation du véritable rapport entre les personnes recensées et les autres résultats de classification, ce qui laisserait sous-entendre une certaine erreur de classification ou un biais dû à l'absence de correction pour le dépistage. Une correction pour la première des deux hypothèses exerce une influence relativement mineure sur l'estimation des personnes manquantes, tous les résultats de classification subissant une inflation ou une déflation correspondant à la différence proportionnelle pour les personnes recensées. Une correction pour la deuxième hypothèse pourrait avoir un effet assez marqué puisque l'absence d'estimation du véritable rapport laisse sous-entendre que toute la différence soit attribuée à d'autres catégories.

Si la deuxième hypothèse s'applique, une correction pourrait réduire l'erreur en fin de période pour neuf des douze provinces et territoires, c'est-à-dire pour toutes les provinces dans lesquelles l'erreur en fin de période survient dans le même sens que la différence pour les personnes recensées. Par contre, si la différence est attribuable à des problèmes de représentativité de l'échantillon, une correction subséquente aurait un effet négligeable ou de légère inflation de l'erreur en fin de période dans la plupart des provinces. De plus, l'évaluation est rendue plus complexe par le fait qu'il est difficile d'établir les chiffres comparables du recensement. Une erreur risque d'être introduite par diverses sources, y compris l'estimation des émigrants de retour fondée sur le recensement (${}^{91FR}RE_{96PP}$), une correction exagérée ou insuffisante du chevauchement des bases de sondage, l'erreur d'échantillonnage et l'erreur non due à l'échantillonnage pour l'estimation du sous-dénombrement de 1991 et de 1996, l'erreur d'échantillonnage et l'erreur non due à l'échantillonnage pour l'estimation du surdénombrement, de même que l'erreur éventuelle de classification des personnes recensées selon la province. Dans un tel contexte, il semble justifié de poursuivre la recherche sur la véritable nature des erreurs de l'estimation des personnes recensées fondée sur la CVD.

5. Conclusion

Les auteurs ont montré que le programme de mesure de la couverture du recensement canadien offre des renseignements supplémentaires qui ont une valeur appréciable pour l'estimation de la population. À part la possibilité d'estimer le sous-dénombrement au recensement, on peut élargir les résultats de classification tirés de ces études de façon à obtenir une autre estimation de la croissance démographique, possiblement décomposée par composante. À l'aide de la plus importante des études de couverture (la contre-vérification des dossiers de 1996), les auteurs ont présenté une nouvelle méthode qui permet une estimation indépendante de la croissance démographique pour la période intercensitaire. La contre-vérification des dossiers permet non seulement d'établir des estimations jugées très exactes pour l'erreur de couverture au recensement, en évitant une partie du biais de corrélation qui a nui aux études postcensitaires

dans d'autres pays, mais également d'obtenir des renseignements très utiles sur l'ampleur de certaines migrations importantes pour l'estimation de la population.

L'élément clé de la contre-vérification des dossiers est qu'elle se fonde sur un échantillon représentatif de toutes les personnes qui, théoriquement, auraient pu se trouver au Canada le jour du recensement, avec des lacunes mineures seulement vu la qualité élevée des statistiques de l'état civil et des données sur l'immigration au Canada. À l'aide de procédures exhaustives de dépistage et d'interviews, on obtient alors des renseignements précieux sur le nombre et les caractéristiques des personnes dûment recensées, manquantes et dénombrées plus d'une fois, de même que des renseignements utiles sur le nombre de personnes quittant le pays (temporairement ou de façon permanente), le nombre de personnes décédées, demeurant dans une autre province, et ainsi de suite. Compte tenu de l'échantillon relativement grand et du savoir-faire et de l'effort appréciables consacrés à la minimisation de tous les types d'erreur, les résultats de classification obtenus peuvent favoriser le programme des estimations de population. Cela est particulièrement vrai pour certaines migrations difficiles à estimer.

La minimisation de toute erreur dans le programme de mesure de la couverture du recensement demeure un objectif prioritaire de la planification du Recensement de 2001. Puisque ces études ont été conçues principalement en fonction d'une estimation de la population « manquante » plutôt que d'autres résultats de classification (personnes émigrées, décédées, etc.), la nouvelle stratégie démographique présentée par les auteurs soulève la question de savoir si l'on devrait remanier en quelque sorte le plan actuel, vu que son usage courant est élargi. Il est intéressant à cet égard de constater que ces études de couverture semblent fournir une autre estimation de la croissance qui fait concurrence à l'estimation courante accessible dans le cadre du programme des estimations de population, et qui est probablement meilleure pour ce qui est de certaines composantes. Il semble justifié de poursuivre la recherche sur les façons de mieux exploiter cette constatation afin d'améliorer la qualité du programme des estimations de population.

Remerciements

Nous tenons à remercier R.G. Carter et P. Dick, tous deux de Statistique Canada, de même que G. Robinson, du Bureau of the Census des États-Unis, des remarques formulées au sujet d'une version antérieure du présent document. Les remarques et suggestions du rédacteur adjoint et de deux examinateurs sont également appréciées.

Bibliographie

- Brackstone, G.J., et Gosselin, J.-F. (1973). *Census Evaluation Program, 1971 RRC: Methodology Report*. Statistique Canada. Ottawa, Ontario.

- Burgess, R.D. (1988). Évaluation des estimations du sous-dénombrement obtenues par la contre-vérification des dossiers du recensement du Canada. *Techniques d'enquête*, 14, 147-167.
- Carter, R.G. (1990). The Measurement of net coverage error in Canadian censuses. *Recueil: Symposium 90, Measure et amélioration de la qualité des données*, Statistique Canada.
- Fellegi, I.P. (1969). A theory of record linkage. *Journal of the American Statistical Association*, 64(328), 1183-1210.
- Gosselin, J.-F. (1976). The methodology of the 1971 Reverse Record Check. *Techniques d'enquête*, 2, 180-194.
- Romaniuc, A. (1988). Une approche démographique à l'évaluation du recensement de 1986 et des estimations de population pour le Canada. *Techniques d'enquête*, 14, 169-185.
- Royce, D. (1993). Evaluation of the May 1993 Revised Results of the 1991 Census Coverage Studies. Division des méthodes d'enquêtes sociales, document de travail, Statistique Canada, Ottawa, Ontario.
- Royce, D., Germain, M.-F., Julien, C., Dick, P., Switzer, K. et Allard, B. (1994). *Couverture: Rapports techniques du recensement de 1991*. No. 92-341F au catalogue, Ottawa: Statistique Canada.
- Statistique Canada (1999). *Couverture: Rapports techniques du recensement de 1996*. No. 92-370-XPB au catalogue.
- Statistique Canada (2000). *Statistique démographiques annuelles*. No. 21-213-XPB au catalogue.
- Tourigny, J., Clark C. et Provost, M. (1998). Evaluation of the March 1998 Preliminary Results of the 1996 Census Coverage Studies. Division des méthodes d'enquêtes sociales, document de travail, Statistique Canada, Ottawa, Ontario.
- Tourigny, J., Bureau, M. et Clark, C. (1998). Revised Direct Estimates of 1991 Census Coverage Studies. Sept 24th release. Division des méthodes d'enquêtes sociales, document de travail, Statistique Canada, Ottawa, Ontario.