



N° 91F0015MIF au catalogue — N° 001
ISSN: 1205-9978

Document de recherche

Documents démographiques

Une revue des méthodes d'estimation du sous-dénombrement net du recensement au Canada, aux États-Unis, Grande-Bretagne et en Australie

par Don Kerr

Division de la démographie
1710 Immeuble principal, Ottawa, K1A 0T6

Téléphone: 1 800 263-1136

Toutes les opinions émises par l'auteur de ce document ne reflètent pas nécessairement celles de Statistique Canada.



Statistique
Canada

Statistics
Canada

Canada

Une revue des méthodes d'estimation du sous-dénombrement net du recensement au Canada, aux États-Unis, en Grande-Bretagne et en Australie

par Don Kerr

91F0015MIF
ISSN : 1205-9978

Division de la démographie
1710 Immeuble principal, Ottawa, K1A 0T6
Statistique Canada

Comment obtenir d'autres renseignements:
Service national de renseignements: 1 800 263-1136
Renseignements par courriel : infostats@statcan.ca

Mars 1998

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'industrie, 2004

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Also available in english (Catalogue no. 91F0015MIE).

LA CONJONCTURE DÉMOGRAPHIQUE

DOCUMENTS DÉMOGRAPHIQUES

Les *Documents démographiques* constituent une collection de textes traitant de la population mais dans des genres divers (analytique, méthodologique, descriptif, etc.). Ces textes sont destinés à informer les personnes intéressées par les questions démographiques, des développements dans la connaissance soit des techniques d'analyse, soit de leurs résultats, soit des méthodes d'élaboration de données ou d'autres produits similaires.

Les lecteurs sont encouragés à faire part de leurs commentaires, suggestions et critiques aux auteurs ou au rédacteur en chef de la collection.

Les *Documents démographiques* sont distribués par les bureaux régionaux et les points de contact provinciaux. Ces textes peuvent être extraits du système Internet à www.statcan.ca.

Pour obtenir des documents sur papier, veuillez communiquer avec :

Division de la démographie
Section de la recherche et de l'analyse
Statistique Canada
Édifice Principal, salle 1708,
Parc Tunney, Ottawa,
Ontario, K1A 0T6
Canada

Téléphone : 1-613-951-2280

Télécopie : 1-613-951-2952

Signes conventionnels

- .. Données non disponibles.
- ... N'ayant pas lieu de figurer.
- Nul ou zéro.
- Nombres infimes.

Les données analysées étaient les plus récentes au moment de la rédaction.

Le papier utilisé dans la présente publication répond aux exigences minimales de l'«American National Standard for Information Sciences» - «Permanence of Paper for Printed Library Materials», ANSI Z39.48 - 1984.∞



Sommaire

Tous les pays qui organisent des recensements s'inquiètent post factum de sa qualité et de l'ampleur du sous-dénombrement net. Pour mesurer l'une et l'autre, différentes méthodes ont depuis longtemps été imaginées. Certaines mettent à profit des sources d'informations indépendantes du recensement lui-même, alors que d'autres contrôlent la cohérence des résultats. Ce sont des opérations lourdes et onéreuses.

Attendu que la population dans chaque pays est organisée différemment et que les structures administratives ne sont pas les mêmes, il n'y a pas de modèle universel qui puisse être appliqué pour obtenir ces mesures. La division de la Démographie de Statistique Canada a passé en revue les procédés utilisés dans quatre pays du monde industrialisé afin de comparer les méthodes et apprécier leurs avantages et leurs faiblesses : les États-Unis, le Royaume Uni, l'Australie et évidemment le Canada. Il ressort de la revue que l'analyse démographique peut être d'un grand secours pour déceler les incohérences quand vient le temps de comparer deux recensements successifs et que la technique des micro-données et le couplage des enregistrements sont indispensables pour évaluer l'importance du nombre de ceux qui ont été omis ou au contraire comptés deux fois le jour de l'énumération. Mais la conclusion la plus importante est la nécessité de conjuguer les efforts des démographes et des statisticiens pour évaluer des nombres dont la justesse demeurera toujours sujette à discussion.



LA CONJONCTURE DÉMOGRAPHIQUE

Équipe de production :

Directeur de la Division de la démographie
Rédacteur en chef
Coordonnateur technique
Composition et mise en page

Réjean Lachapelle
Jean Dumas
Carol D'Aoust
Danielle St-Germain

Ouvrages déjà parus :

Rapports sur l'état de la population du Canada (n° 91-209F au catalogue)

Rapport 1983, par J. Dumas - (Épuisé)

Rapport 1986, par J. Dumas - (139 pages, 15,00 \$) :

- *Le comportement fécond des femmes mariées, canadiennes de naissance*
- *La fécondité des célibataires*
- *Renforcement des positions majoritaires*, par R. Lachapelle

Rapport 1988, par J. Dumas - (166 pages, 21,00 \$) :

- *Les interruptions volontaires de grossesse dans la perspective démographique*
- *Conséquences à long terme du mariage et de la fécondité chez les adolescentes*, par C. Grinstaff

Rapport 1990, par J. Dumas - (116 pages, 26,00 \$) :

- *Évolution récente des populations canadiennes et américaines*

Rapport 1991, par J. Dumas avec la collaboration de A. Bélanger et C. Fortier - (187 pages, 26,00 \$) :

- *Brève revue des principaux mouvements migratoires dans le monde depuis la Seconde Guerre mondiale*

Rapport 1992, par J. Dumas - (162 pages, 26,00 \$) :

- *Structures démographiques en mutation, un bilan de deux siècles*, par Y. Lavoie

Rapport 1993, par J. Dumas - (227 pages, 26,00 \$) :

- *Les défis démographiques du Mexique (un aperçu)*

Rapport 1994, par J. Dumas et A. Bélanger avec la collaboration de G. Smith - (157 pages, 30,00 \$) :

- *La « génération sandwich » : Mythes et réalité*

Rapport 1995, par J. Dumas et A. Bélanger avec la collaboration de G. Smith - (200 pages, 30,00 \$) :

- *Ressemblances et dissemblances démographiques de l'Ontario et du Québec*

Rapport 1996, par J. Dumas et A. Bélanger avec la collaboration de G. Smith (192 pages, 30,00 \$) :

- *Les unions libres au Canada à la fin du XX^e siècle*



Hors série

La fécondité au Canada : Croissance et déclin, par A. Romaniuc - (n° 91-524F au catalogue, 1984, 158 pages, 16,00 \$)

Le revenu des immigrants au Canada, par R. Beaujot, K. G. Basavarajappa et R. B. P. Verma - (n° 91-527F au catalogue, 1988, 109 pages, 20,00 \$)

Les immigrants antillais, par A. H. Richmond - (n° 91-536F au catalogue, 1989, 85 pages, 25,00 \$)

Nouvelles tendances de la famille, par B. Ram - (n° 91-535F au catalogue, 1990, 98 pages, 25,00 \$)

Mariage et vie conjugale au Canada, par J. Dumas et Y. Péron - (n° 91-534F au catalogue, 1992, 181 pages, 38,00 \$)

Le vieillissement de la population et personnes âgées, par B. Desjardins - (n° 91-533F au catalogue, 1993, 130 pages, 40,00 \$)

La famille au long de la vie, par R. Beaujot, E. M. Gee, F. Rajulton et Z. R. Ravanera - (n° 91-543F au catalogue, 1995, 186 pages, 38,00 \$)

Document démographique

Document n° 1 : «Projection de la fécondité, Canada, provinces et territoires, 1993-2016», par R. B. P. Verma, S. Loh, S. Y. Dai et D. Ford - (n° 91F0015MPF au catalogue, 1996, 28 pages, 5,00 \$)

Document n° 2 : «La population en logements collectifs au Canada de 1971 à 1991» par G. Smith - (n° 91F0015MPF au catalogue, 1996, 52 pages, 5,00 \$)

Document n° 3 : «Nouvelles tables de mortalité par génération au Canada et au Québec, 1801-1991» par R. Bourbeau, J. Légaré et V. Emond - (n° 91F0015MPF au catalogue), 1997, 94 pages, 5,00 \$)

Document n° 4 : «Avantages de la variable de mobilité sur un an pour la ventilation de la migration interprovinciale par âge, sexe et état matrimonial» par M. Bédard et M. Michalowski - (n° 91F0015MPF au catalogue, 1997, 60 pages, 5,00 \$)

À VENIR

Rapport 1997, par J. Dumas et A. Bélanger (n° 91-209-XPF au catalogue) :

- *Impact de l'entourage des personnes âgées sur leurs conditions socio-économiques*

«Les relations numériques entre les générations dans les estimations de population» par C. Dionne et D. Kerr dans *Document démographique* - (n° 91F0015MPF au catalogue)



Remerciements

Ce document a été préparé sous la direction de R. Raby (chef de la section des estimations de la population) et de J. Dumas (rédacteur en chef de la Conjoncture démographique), tous deux de la Division de la Démographie de Statistique Canada.

L'auteur tient à remercier les nombreuses personnes qui ont révisé le manuscrit et fait part de leurs commentaires : G. Robinson (U.S. Bureau of the Census), A. Teague et J. Charlton (Office for National Statistics, U.K.), J. Paice (Australian Bureau of Statistics), S. Simpson (University of Manchester, U.K.), and P. Dick, J. Tourigny, B. Laroche, R. Lachapelle, G. Smith and D. Morissette (all of Statistics Canada). Il voudrait aussi témoigner sa gratitude à D. St-Germain et C. D'Aoust pour leur assistance technique.



Table des matières

	Page
Introduction	1
1. Canada	3
1.1 Techniques de collecte des données (recensement de droit)	3
1.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement au Canada	3
1.3 Méthodes démographiques appliquées aux données canadiennes	6
1.3.1 Application de la méthode démographique de base aux données canadiennes	7
1.3.2 Estimations postcensitaires canadiennes	8
1.3.3 Autres méthodes d'analyse démographique	9
2. États-Unis	10
2.1 Méthodes de collecte des données	10
2.2 Techniques utilisant les micro données pour l'évaluation du recensement aux États-Unis	11
2.3 Analyse démographique et évaluation de l'erreur de couverture du recensement aux États-Unis	13
2.3.1 Estimations démographiques et incertitude	16
3. Royaume-Uni	17
3.1 Techniques de collecte de données	17
3.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement en Grande-Bretagne	18
3.3 Méthodes démographiques pour l'évaluation de la couverture du recensement en Grande-Bretagne	20
4. Australie	22
4.1 Méthodes de collecte des données	22
4.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement en Australie	22
4.3 Méthodes démographiques pour l'évaluation de la couverture du recensement en Australie	24
5. Discussion	25
6. Conclusion	28
Références	30



PUBLICATIONS ÉLECTRONIQUES DISPONIBLES À
www.statcan.ca



Introduction

On s'attend à ce que le dénombrement d'une population par le recensement soit complet; cependant, on ne parvient jamais à cette perfection de sorte que les organismes statistiques publics utilisent divers moyens pour mesurer le sous-dénombrement. Plusieurs méthodes mises au point par les démographes et les statisticiens ont largement contribué à évaluer la «couverture» des données du recensement. Le but du présent rapport est de passer en revue certaines de ces méthodes utilisées au Canada, aux États-Unis, en Grande-Bretagne et en Australie.

Le but de l'analyse démographique, dans le cadre de l'estimation de la population, est de fournir des estimations indépendantes du recensement, en s'appuyant sur des données sur le nombre de naissances, de décès et de migrants. Les estimations démographiques ainsi obtenues peuvent alors être comparées aux chiffres du recensement, pour déterminer combien de personnes ont été omises le jour du dénombrement. La question est toutefois de savoir si les estimations sont suffisamment précises pour mesurer le sous-dénombrement du recensement, étant donné les incertitudes qui planent sur la qualité des données administratives et des statistiques de l'état civil sur lesquelles s'appuient habituellement ces estimations. Les résultats de l'analyse démographique peuvent également être comparés à ceux des méthodes utilisant les micro données mises au point par les statisticiens et qui consistent en l'appariement et le couplage des enregistrements pour chaque cas. Ces techniques ont pour but de comparer systématiquement un échantillon de la population qui aurait dû être recensée le jour du recensement au nombre effectivement dénombré. Lors du Recensement de la population de 1991 au Canada, l'utilisation d'une telle méthode utilisant

les micro données basée principalement sur la contre-vérification des dossiers (Statistique Canada, 1994) a servi à Statistique Canada pour l'ajustement officiel de l'erreur de couverture. La contre-vérification des dossiers est également la principale méthode utilisée pour estimer le sous-dénombrement du Recensement de 1996, tant à l'échelle nationale qu'au niveau provincial et territorial.

Les procédures pour estimer l'intégralité de la couverture du recensement ont deux objectifs. Elles visent d'abord à estimer le nombre de personnes omises lors du recensement et à leur attribuer certaines caractéristiques du recensement (âge, sexe, état matrimonial, langue parlée à la maison, situation relative à l'emploi, lieu de résidence rural ou urbain, etc.). Elles servent ensuite à corriger la population sur laquelle sont basées les estimations démographiques officielles¹. En ce qui a trait au premier objectif, les techniques d'appariement et de couplage d'enregistrements des micro données sont les seules qui fournissent le type d'informations détaillées requises. Quant au deuxième objectif, l'analyse démographique peut compléter les résultats des techniques utilisant les micro données, en réduisant au minimum le biais dans les estimations démographiques selon l'âge et le sexe. Dans ce rapport, nous examinons l'utilité d'autres méthodes permettant d'atteindre le deuxième objectif, en améliorant la précision des estimations établies par âge et par sexe. Cet exercice n'est pas dénué d'importance, puisque ce sont ces chiffres ainsi corrigés qui servent à

1 Il faut pour cela des informations détaillées sur les personnes omises, selon l'âge, le sexe et le lieu de résidence.



l'établissement des estimations et des projections démographiques de

Statistique Canada. L'évaluation du sous-dénombrement relatif dans les provinces et territoires revêt également une importance capitale pour le calcul des paiements de transfert fiscal, puisque la part des paiements de péréquation fédéraux qui est allouée à une province varie selon que le sous-dénombrement de sa population est supérieur ou inférieur à la moyenne nationale.

Il est important d'insister sur l'ordre dans lequel s'effectuent les procédures pour établir les estimations démographiques de base selon l'âge et le sexe, entièrement corrigées en fonction du sous-dénombrement du recensement.

- 1) Au départ, tous les pays tentent d'effectuer un recensement de leur population qui soit le plus complet possible. Inévitablement, toutefois, des erreurs se glissent durant l'étape de la collecte, et on fait face à des données erronées ou incomplètes comme l'erreur dans la déclaration de l'âge ou l'absence de réponse à certaines questions.
- 2) Les données recueillies par le recensement sont ensuite vérifiées, afin de déceler les erreurs de déclaration évidentes. Il y a ensuite imputation des données manquantes à toutes les questions du recensement.
- 3) Les estimations de la couverture du recensement sont obtenues par diverses techniques, dont l'analyse démographique et l'appariement et le couplage d'enregistrements. Des comparaisons sont faites systématiquement entre la population de base du recensement et une norme ou un échantillon calculé de façon indépendante.

- 4) Enfin, les estimations du sous-dénombrement du recensement sont ajoutées à la population du recensement, à titre de correction finale.

En plus d'être complexe, ce processus comporte le risque qu'un biais soit introduit à chacune de ses étapes. Le présent rapport porte principalement sur les étapes 3 et 4. Évidemment, toute évaluation des facteurs responsables du sous-dénombrement selon l'âge et le sexe doit également s'intéresser au biais introduit aux étapes 1 et 2. Bien que des méthodes de collecte des données seront brièvement discutées avant d'examiner les méthodes démographiques et les techniques utilisant les micro données précises utilisées pour estimer le sous-dénombrement, l'examen systématique des techniques de vérification et d'imputation dépasse les propos du présent report.

Les méthodes démographiques peuvent, pour le moins, faciliter l'évaluation du succès des méthodes de couplage d'enregistrements utilisant les micro données, en fournissant des estimations indépendantes. Ces méthodes peuvent toutefois avoir beaucoup d'autres utilisations, par exemple remplacer les estimations erronées résultant des études postcensitaires, comme ce fut le cas en Grande-Bretagne, en 1991. Elles peuvent également accroître la confiance dans les estimations démographiques officielles, en corroborant des résultats obtenus par d'autres méthodes. Le rapport se terminera par la formulation de quelques recommandations sur les recherches qu'il faudrait mener au Canada à l'avenir sur les méthodes



démographiques servant à mesurer la couverture du recensement.

compléter le recensement même si le taux initial de non-réponse (environ 15 %) est relativement faible selon les normes internationales. Des contrôles de la qualité des données sont également prévus à toutes les étapes de la collecte.

1. Canada

1.1 Techniques de collecte des données (recensement de droit)

L'objectif étant d'obtenir un dénombrement complet de tous les logements, ménages et personnes du Canada, les recenseurs doivent d'abord compiler ou vérifier des listes exhaustives des logements dans chacun des 46 000 secteurs de dénombrement (SD). Dans chaque SD (quelque 220 ménages, en moyenne), les recenseurs identifient les logements résidentiels et les classent selon qu'il s'agit d'un logement privé ou collectif et selon que le logement est occupé ou inoccupé. Tous les occupants habituels de chaque logement occupé sont dénombrés, principalement par auto-déclaration. Pour recueillir cette information, deux méthodes sont utilisées au Canada : le questionnaire à retourner par la poste et l'enquête par interview.

Le recensement par les questionnaires à retourner par la poste est utilisé pour presque tous les logements; le recenseur livre un questionnaire du recensement dans chaque logement et demande aux occupants de le remplir le jour du recensement et de le retourner par la poste. La méthode par interview est réservée à une très faible proportion, soit les personnes qui vivent en régions éloignées, dans les réserves indiennes et dans certains quartiers des grandes villes. Des suivis sont prévus pour les deux méthodes de collecte, afin de

1.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement au Canada

Le Programme de mesure de l'erreur de couverture du Recensement de 1991 prévoit quatre études de base qui s'appuient sur le couplage d'enregistrements et l'appariement², soit :

- 1) la vérification des logements inoccupés;
- 2) la vérification portant sur les résidents temporaires;
- 3) l'étude du surdénombrement et, plus important encore;
- 4) la contre-vérification des dossiers.

Seules la contre-vérification des dossiers et l'étude du surdénombrement sont utilisés pour obtenir des chiffres officiels du sous-dénombrement net. Pour leur part, la vérification des logements inoccupés et la vérification portant sur les résidents

2 Pour être plus précis, il s'agit de quatre études qui s'appuient sur la démarche suivante: retrouver certaines personnes, les interviewer pour obtenir une adresse le jour du recensement et faire un appariement avec les données du recensement (Rapport technique du Recensement de 1991 sur les erreurs de couverture, Statistique Canada, 1993).



temporaires mènent à des ajustements directs de la base de données du recensement. Par conséquent, les estimations officielles du sous-dénombrement net, qui sont indiquées dans les publications de Statistique Canada, mesurent les erreurs de couverture, excluant celles documentées dans deux des quatre études³.

La contre-vérification des dossiers (CVD) est une méthode de couplage d'enregistrements complète, qui sert à estimer le sous-dénombrement brut officiel au Canada. Elle consiste à construire un échantillon destiné à représenter la même population-cible que celle du recensement, c'est-à-dire une base de sondage complète obtenue de manière à être totalement indépendante du recensement évalué et incluant diverses sources de données, dont la population dénombrée lors du recensement précédent, les enregistrements de naissances, les listes administratives des immigrants et des résidents non permanents, ainsi que les personnes omises selon la précédente CVD. En 1991, après prélèvement d'un échantillon de quelque 56 000 personnes qui auraient dû être recensées, une recherche a été faite dans la base de données du recensement pour identifier les

personnes non recensées. Ces personnes ont été retracées et interviewées et les résultats ont été utilisés pour déterminer le nombre et les caractéristiques des personnes non recensées. En complément à la CVD, l'étude du surdénombrement constitue le principal moyen de déceler les dénombremens erronés⁴; elle consiste en un échantillonnage de personnes résidant dans des logements privés et collectifs, dans le but de déterminer s'il y avait lieu de recenser les personnes inscrites sur les formulaires du recensement, d'obtenir de l'information sur d'autres adresses de vérification et de procéder à l'appariement automatisé avec la base de données du recensement pour identifier les ménages et les personnes comptés deux fois. Les résultats de l'étude du surdénombrement fournissent des estimations du surdénombrement brut. Ensemble, les résultats de la CVD et de l'étude du surdénombrement fournissent les estimations officielles du sous-dénombrement net.

La CVD est considérée comme la source d'information la plus fiable sur le sous-dénombrement brut au Canada, et ce tant à l'échelle du pays que des provinces. La CVD a été utilisée pour estimer le sous-dénombrement brut de 1971 à 1991, puis en 1993 pour modifier la base de toutes les estimations démographiques établies par Statistique Canada pour cette période (Statistique

3 En bref, la vérification des logements inoccupés consiste en un échantillonnage stratifié à plusieurs degrés des SD, où tous les logements sélectionnés sont vérifiés pour déterminer si les recenseurs les ont correctement répartis entre les logements occupés et inoccupés. Cette étude donne lieu à un ajustement immédiat de la base de données du recensement, pour les ménages et les personnes omises à la suite d'erreurs de la part des recenseurs. L'étude sur les résidents temporaires consiste en un échantillonnage stratifié de toutes les personnes qui ont déclaré avoir été temporairement absentes de leur lieu de résidence habituel le jour du recensement sur un questionnaire conçu à cette fin et distribué dans les hôtels, les aéroports, etc. S'il s'avère que ces personnes n'ont pas été recensées à leur lieu habituel de résidence, la base de données du recensement est alors corrigée pour les inclure. L'étude sur les résidents temporaires a été annulée en 1996.

4 Double compte, fausse déclaration, ménages «fabriqués», personnes décédées avant le jour du recensement, etc.



Canada, 1994)⁵. La répartition par âge du sous-dénombrement net en 1991 indique un taux relativement faible parmi les enfants, mais un taux nettement plus élevé chez les personnes de 20 à 24 ans et de 25 à 29 ans des deux sexes, ainsi que de très faibles taux chez les Canadiens vers la fin de la quarantaine et dans la cinquantaine. Comme nous l'indiquerons ultérieurement, ce profil est très similaire à ceux observés dans deux autres pays à l'étude, soit l'Australie et la Grande-Bretagne; par contre, il diffère sensiblement du profil américain. Il convient par ailleurs de souligner que le profil mis en évidence par la CVD, selon lequel le sous-dénombrement est plus élevé chez les hommes que chez les femmes, en particulier chez les jeunes adultes, demeure inchangé depuis 1971.

Malgré l'utilisation de procédures rigoureuses, la CVD comporte quelques limites et, fait peu surprenant, la principale source d'erreurs est l'erreur d'échantillonnage de la méthode de CVD, qui augmente avec le niveau de désagrégation. En outre, bien que l'erreur-type soit toujours indiquée dans les publications de Statistique Canada, l'erreur non due à l'échantillonnage n'est habituellement pas quantifiée. À titre d'exemple, un pourcentage significatif (4,9 %) de l'échantillon non pondéré de la CVD n'a pu en fait être «retrouvé» par la technique d'appariement (en d'autres mots, la personne choisie avait déménagé de la dernière adresse connue et ni elle, ni aucun membre du ménage, n'ont pu être retrouvés à partir de

l'information disponible). Il a donc été impossible de classer ces personnes comme étant soit «recensées avec succès», soit «omisées». Ces enregis-trements «non retrouvés» ont été imputés par l'une des caractéristiques suivantes : «recensée», «omise», «décédée» ou «émigrée/à l'étranger/hors du champ d'observation», en utilisant une distribution à peu près égale à celle observée chez les personnes qui avaient déménagé de la dernière adresse connue mais qui avaient pu être retrouvées. Cette imputation comporte un biais de grandeur inconnue, car il se peut que les personnes qui ont été retrouvées diffèrent de celles qui ne l'ont pas été, en termes de migration⁶.

Les études de couverture de Statistique Canada sont reconnues pour produire des estimations de haute qualité du sous-dénombrement net, aux échelons national et provincial. Certes, l'erreur d'échantillonnage associée à ces études de couverture augmente avec le niveau de désagrégation, de sorte qu'il n'est pas surprenant que la CVD ne puisse pas produire d'estimations précises pour des catégories détaillées selon l'âge et le sexe. Cependant, au moment d'analyser les corrections en fonction du sous-dénombrement net à de tels niveaux de désagrégation, il faut également tenir compte des techniques de collecte des données pour le recensement, ainsi que de la qualité des réponses indiquées pour certaines variables comme la date de naissance et le sexe. En 1991, par exemple, il a fallu imputer les données pour la date de naissance ou le sexe pour plus de 727 000 personnes. Ce nombre représente les personnes de quelque 200 000

5 Comme l'étude du surdénombrement a été utilisée pour la première fois pour le Recensement de 1991, nous ne disposons d'aucune mesure réelle du surdénombrement avant cette période. Cette nouvelle série d'estimations démographiques (1971-1991), corrigées en fonction du sous-dénombrement net, est donc basée sur une estimation du surdénombrement qui date d'avant 1991 et qui repose sur l'hypothèse voulant que la distribution soit restée inchangée. Pour plus de détails sur les estimations pour cette série, voir Michalowski, 1993.

6 Burgess (1988) présente un examen de certaines difficultés associées aux versions antérieures de la technique de CVD. La plupart des problèmes soulevés par Burgess en rapport avec le Recensement de 1986 ont été corrigés, incluant les problèmes liés à la petite taille de l'échantillon et aux lacunes importantes dans la base de sondage de la CVD.



ménages pour lesquels aucune information n'a été recueillie (la taille et les caractéristiques du ménage ont toutes deux dû être imputées), en plus des cas de non-réponse aux questions sur l'âge et le sexe pour plus de 150 000 personnes pour lesquelles des données partielles ou invraisemblables ont été recueillies. Outre ces données imputées pour plus de 727 000 personnes, les résultats (profils) de l'étude sur les résidents temporaires (auprès de 92 584 personnes) et de la vérification des logements inoccupés (auprès de 134 668 personnes) sont ajoutés à la base de données du recensement par addition aléatoire, une technique que l'on peut comparer à l'imputation. Le total auquel on parvient est de plus de 950 000 personnes. Ne sont pas quantifiées ici les erreurs des répondants (répondants substitués), les erreurs liées à la saisie des données et les erreurs systématiques de traitement, qui ne peuvent être décelées par les systèmes ou l'analyse de certification. Il est donc difficile de déterminer si les problèmes observés au niveau désagrégé sont imputables au biais non dû à l'échantillonnage et à la variance d'échantillonnage des études de couverture ou au biais non dû à l'échantillonnage des chiffres du recensement.

Même si les erreurs observées entre des recensements successifs peuvent être très différentes, voire contradictoires, pour une même cohorte, certaines corrections donnent lieu à des incohérences dans les chiffres du recensement, lorsque l'on suit des cohortes au fil des ans. Alors que les chiffres de population sont généralement acceptables pour le Canada et les provinces, ils le sont moins pour certains groupes d'âge particuliers. Ainsi, les chiffres du recensement pour des cohortes suivies au fil des ans, corrigés en fonction des résultats des études de couverture, ne sont pas nécessairement cohérents avec les données des démographes sur la mortalité et la migration dans ces cohortes. Or si les données sur la mortalité et la

migration qui servent à établir les estimations démographiques sont assez fiables et valables, les chiffres corrigés du recensement pour certaines cohortes deviennent alors hautement improbables. Par exemple, la différence entre les estimations corrigées du recensement pour les hommes de 20 à 24 ans en 1986 et les estimations corrigées du recensement pour la même cohorte en 1991 semble presque inexplicable, si l'on se fie aux statistiques de l'état civil et aux données sur l'immigration. En effet, la différence observée entre les chiffres corrigés du recensement correspondait dans l'ensemble à moins de 20 % de la différence estimée par méthodes démographiques, pour cette cohorte.

Les méthodes démographiques semblent donc indiquer que d'autres améliorations sont souhaitables, en particulier pour assurer une cohérence intercensitaire entre les chiffres du recensement au Canada. C'est là un objectif important des modèles démographiques utilisés pour corriger les données du recensement.

1.3 Méthodes démographiques appliquées aux données canadiennes

Dans le contexte canadien, les méthodes démographiques semblent offrir deux avantages. Elles permettent d'abord d'évaluer les études de couverture, puis elles fournissent des estimations plus précises de la population du Canada, selon des catégories détaillées par âge et par sexe. Il convient de souligner que l'évaluation des résultats de la CVD par l'analyse démographique a déjà mené à la révision des estimations du sous-dénombrement net selon la CVD, pour certaines catégories selon l'âge et le sexe. En 1991, par exemple, les résultats de la CVD pour les enfants de 0 à 4 ans indiquaient un taux de sous-dénombrement net estimé beaucoup plus élevé chez les enfants de sexe féminin (3,75



%) que masculin (2,26 %), et ce sans aucune raison apparente si ce n'est l'erreur d'échantillonnage. L'analyse démographique, menée par la suite sur la base d'informations précises concernant le rapport de masculinité à la naissance et les données de mortalité obtenues des statistiques de l'état civil, a clairement montré que l'ajout des résultats de la CVD aux résultats du recensement donnait des chiffres peu vraisemblables. Le nombre d'enfants de sexe féminin a donc été corrigé de manière à correspondre au rapport de masculinité chez les enfants de 0 à 4 ans, établi par l'analyse démographique. De même, l'analyse démographique a semblé indiquer certains problèmes avec les chiffres corrigés du recensement pour les groupes plus âgés. Elle a donc établi les estimations du sous-dénombrement net par groupe quinquennal, pour les Canadiens âgés de plus de 55 ans, en établissant le sous-dénombrement total dans ces groupes d'âge (selon la CVD) en proportion avec la taille de la cohorte. Comme la CVD a été conçue uniquement pour estimer le nombre de personnes omises par le recensement, la taille de l'échantillon utilisé n'est pas suffisamment grande pour fournir des estimations exactes pour les groupes plus âgés. Des techniques d'analyse brutes ont donc été utilisées pour établir les estimations pour les groupes plus âgés, avec répartition proportionnelle selon la distribution par âge de la population des 55 ans et plus.

1.3.1 Application de la méthode démographique de base aux données canadiennes

Comme nous l'avons indiqué précédemment, l'application des méthodes démographiques pour l'évaluation de la couverture du recensement est tributaire de la qualité des données sur la fécondité, la mortalité et la migration. L'équation de comptabilité démographique de base part d'une cohorte à la naissance, à laquelle sont ajoutées et soustraites

les diverses composantes du changement démographique (décès, immigration et émigration) pour chaque année, jusqu'à l'année du recensement à l'étude. En ce qui a trait aux données canadiennes, nous disposons d'une assez longue série chronologique, puisque des données sur les composantes du changement démographique sont compilées depuis 1921. À partir de cette série chronologique, Bender (1992) a estimé la population du Canada pour plusieurs des dernières années du recensement, par sexe et groupe d'âge quinquennal. Ces estimations ont ensuite été comparées aux données



non corrigées du recensement, pour obtenir des estimations indépendantes du sous-dénombrement net à l'échelle nationale, selon l'âge et le sexe.

Pour la résumer brièvement, la méthode de comptabilité démographique de base s'appuie sur l'équation d'équilibre suivante :

$$C_t = B_y - D_{y,t} + I_{y,t} - E_{y,t},$$

où C_t représente la taille de la cohorte durant l'année t , B_y représente la taille de la cohorte à son origine (nombre de naissances durant l'année y) et $D_{y,t}$, $I_{y,t}$ et $E_{y,t}$ désignent le nombre cumulatif d'événements (décès, immigrants et émigrants) qui influent sur la taille de la cohorte entre les années y et t . Comme les données existent sur les composantes du changement démographique depuis 1921, la méthode d'analyse démographique pure peut fournir des estimations pour le groupe des 0 à 4 ans en 1926, les groupes des 0 à 4 ans et des 5 à 9 ans en 1931, les groupes des 0 à 4 ans, 5 à 9 ans et 10 à 14 ans en 1936, et ainsi de suite depuis le groupe des 0 à 4 ans jusqu'au groupe des 65 à 69 ans inclusivement, en 1991.

Cette technique soulève toutefois certains problèmes, du fait que les données de Statistique Canada sur l'enregistrement des naissances selon le sexe et sur l'enregistrement des décès selon l'âge et le sexe, ainsi que les estimations de l'immigration, de l'émigration, des résidents non permanents et des Canadiens de retour selon l'âge et le sexe, comportent des erreurs qui, par surcroît, augmentent à mesure que la série chronologique s'allonge. Heureusement, les deux composantes fondamentales, à savoir les naissances et les décès, sont également celles qui sont mesurées avec le

plus de précision⁷. En revanche, la qualité des estimations sur la migration internationale selon l'âge et le sexe diminue à mesure que l'on remonte dans le temps, et plus particulièrement celles sur l'émigration.

Les divers problèmes liés à la qualité des données ont pour effet de rendre peu fiables les estimations du sous-dénombrement net obtenues par la méthode de comptabilité, pour les groupes plus âgés. Ainsi, de très hauts taux de surdénombrement ont été estimés pour les Canadiens nés avant 1951, ce qui n'a rien de surprenant puisque à la Division de la démographie on a moins confiance dans la série chronologique avant 1971. Il est possible que des estimations acceptables, indépendantes des données du recensement, ne puissent être établies que pour les groupes d'âge plus jeunes car, pour les personnes plus âgées, l'effet cumulatif de l'erreur dans les composantes risque de produire des estimations peu fiables.

1.3.2 Estimations postcensitaires canadiennes

Pour l'évaluation du Recensement de 1986, Romaniuc (1988) a utilisé la méthode des composantes pour établir les estimations postcensitaires,

7 Aucune étude n'a jamais évalué l'intégralité de l'enregistrement des naissances et des décès au Canada, si ce n'est une étude limitée sur l'enregistrement des naissances durant les années 30, qui a été menée par Enid Charles (1940). Des études préparatoires portant sur quelques secteurs de recensement semblent indiquer un taux d'omission de 3 %. Les analystes des données sur les naissances au Canada estiment que le système d'enregistrement était déjà très fiable au moment où l'introduction du programme d'allocations familiales universelles à la fin de années 40 est venue accroître l'incitation à enregistrer les naissances. Cependant, l'allocation familiale universelle a récemment été remplacée par un crédit d'impôt pour enfants et on ignore si ceci aura un effet sur l'incitation à enregistrer les naissances.



les erreurs de fermeture et les estimations correspondantes du sous-dénombrement, dans les populations nationale et provinciales. Cette méthode diffère de la formule de comptabilité démographique de base, en ce qu'elle consiste à ajouter les naissances et les immigrants, et à soustraire les décès et les émigrants, de la population de base du recensement réalisé cinq ans auparavant, plutôt que de partir d'une cohorte à la naissance. Dans son évaluation du Recensement de 1991, Bender (1992) a, elle aussi utilisé cette méthode pour estimer les taux de sous-dénombrement net selon l'âge et le sexe en 1981, 1986 et 1991. Tous deux ont fait ressortir les lacunes et les mérites de cette méthode, pour l'estimation du sous-dénombrement net. Lorsqu'on la compare à la formule de comptabilité, cette méthode comporte manifestement moins de risques d'erreur au niveau des composantes de la croissance démographique, car les composantes se limitent à la dernière période intercensitaire. Par contre, la population de base du recensement joue un rôle majeur dans la détermination des estimations subséquentes, car on présume que cette population reflète l'effet net de toutes les composantes antérieures du changement démographique, pour chaque cohorte.

La comparaison des chiffres corrigés par CVD du recensement de 1991 aux estimations postcensitaires ne peut fournir une estimation du sous-dénombrement net selon l'âge et le sexe, que si l'on admet au départ que la population de l'année de base 1986 a été parfaitement mesurée. Les événements intercensitaires selon l'âge et le sexe ont été mesurés ou estimés sans erreur. Ces deux hypothèses sont irréalistes dans le contexte canadien, en particulier la première. En effet, rien n'indique que l'erreur dans les chiffres corrigés par CVD en 1986 soit moindre que celle dans les chiffres corrigés par CVD en 1991, puisque des

méthodes similaires ont été utilisées. En résumé, bien que cette méthode donne une indication raisonnable du degré d'incohérence intercensitaire dans les chiffres corrigés du recensement, elle fournit peu d'indications sur les moyens de corriger les écarts observés. Il est toutefois possible d'assouplir l'hypothèse sur l'exactitude des estimations corrigées par CVD en 1986, en tenant compte des taux relatifs de sous-dénombrement net plutôt que des chiffres absolus. Il est instructif, par exemple, de savoir que la couverture en 1991 diffère de celle de 1986, même si l'on ne connaît pas les taux précis de couverture de l'un ou l'autre recensement.

1.3.3 Autres méthodes d'analyse démographique

Outre celles mentionnées précédemment, peu de recherches ont porté sur l'utilisation des méthodes démographiques pour l'évaluation des données du recensement. Les estimations postcensitaires, en complément immédiat du programme d'estimations démographiques de Statistique Canada, sont depuis longtemps utilisées pour évaluer les résultats provisoires du recensement et elles continueront de l'être. Certaines recherches préparatoires intéressantes ont également été menées en rapport avec les recensements précédents, pour examiner la pertinence des données administratives⁸ pour le processus d'évaluation. Bien que ces recherches s'annoncent prometteuses à première vue, en particulier pour les groupes plus âgés, davantage de recherches devront être effectuées pour déterminer où se situent les problèmes éventuels dans l'utilisation des données administratives. La collecte et la vérification de la plupart des données administratives sont de la

8 Mentionnons, par exemple, les données sur l'assurance-maladie, les fichiers d'impôt, les allocations familiales pour les enfants de 0 à 14 ans et les prestations de la sécurité de la vieillesse pour les personnes de 65 ans et plus (Fortier et Raby, 1989; Michalowksi, 1992).



responsabilité des provinces; cependant, les priorités de chacune diffèrent, et il en va de même des ressources allouées au contrôle de la qualité des données. C'est ce qui explique que la validité des fichiers de données varie considérablement, d'une province à une autre⁹.

Il y a plus de 25 ans, Lapierre-Adamcyk (1970) s'est engagée dans un domaine de recherche intéressant, qui avait pour but d'appliquer aux données canadiennes les méthodes démographiques utilisées aux États-Unis par Ansley Coale (1955). Ce dernier est généralement reconnu comme le pionnier de l'application de l'analyse démographique pour l'évaluation de la couverture du recensement aux États-Unis (Himes et Clogg, 1992)¹⁰. Utilisant l'itération de Coale basée sur l'hypothèse des «erreurs similaires», Lapierre-Adamcyk a obtenu d'autres estimations du sous-dénombrement net, ventilées selon de grands groupes d'âge pour les Canadiens de naissance; ces estimations ont

ensuite été comparées à d'autres estimations obtenues antérieurement par CVD. La qualité douteuse des statistiques sur les naissances et les décès utilisées pour les estimations fut l'un des principaux problèmes de cette chercheuse, et ce même problème vaut encore aujourd'hui pour les estimations démographiques. Lapierre-Adamcyk avait alors insisté sur la «nécessité d'étudier l'intégralité relative de l'enregistrement des naissances». Si Statistique Canada avait suivi cette recommandation, l'agence serait en mesure de fournir des estimations plus précises, par des techniques directes, pour ces groupes d'âge qui ont été les plus difficiles à dénombrer lors du Recensement de 1996, à savoir les jeunes adultes.

Conscients des faiblesses des données utilisées pour établir les estimations démographiques, C. Dionne a proposé d'autres techniques basées sur l'approche des composantes par cohorte, pour tenter de donner un caractère multidimensionnel aux estimations subséquentement obtenues (Dionne, 1995; Dionne et Kerr, 1995a). Selon Dionne, «multidimensionnelle» s'entend d'une technique qui consiste à tenir compte simultanément de diverses relations entre des cohortes ventilées par sexe et de niveaux variables de sous-dénombrement, pour estimer la taille relative des groupes d'âge durant l'année de recensement. Le but de tels modèles est d'obtenir des estimations du sous-dénombrement net défendables, malgré les problèmes de qualité des données. Le modèle intergénérationnel, qui offre plusieurs applications empiriques, est la plus complète de ces méthodes multidimensionnelles (Dionne et Kerr, 1995b). Les recherches se poursuivent en vue de déterminer l'utilité de telles méthodes pour l'évaluation et peut-être aussi la correction des distributions par âge et par sexe des données du Recensement.

9 Outre l'utilisation des dossiers administratifs pour l'évaluation des données du recensement, le projet de comparaison des dossiers administratifs (CDA) de Statistique Canada a cherché à déterminer si les données administratives pouvaient être utilisées en remplacement des données du recensement (Statistique Canada, 1993). Pour le dénombrement de la population canadienne en 1991, la CDA a été faite d'abord à partir des dossiers d'impôt de Revenu Canada, avec imputation pour les non-déclarants. À partir des estimations indépendantes ainsi obtenues, des comparaisons systématiques ont été faites avec les chiffres de population du recensement de 1991. La CDA ayant révélé des écarts significatifs entre les données, à tous les niveaux géographiques, on en est arrivé à la conclusion que les dossiers administratifs complètent les données du recensement de la population, mais ne peuvent les remplacer (Statistique Canada, 1994). La difficulté vient de la non-universalité de la plupart des ensembles de données administratives au Canada et, donc, du besoin d'imputer les données pour les populations omises.

10 La recherche de Ansley Coale sera examinée à la section sur les techniques américaines.



Aux fins d'évaluation des résultats du Recensement, la cohérence entre les diverses techniques pourrait être un critère de correction des données. D'autres recherches devront toutefois être effectuées pour déterminer la solidité des techniques en regard des problèmes de qualité des données et pour expliquer les écarts observés.

2. États-Unis

2.1 Méthodes de collecte des données

Avant de procéder au recensement qui se fait par le biais d'un questionnaire envoyé et retourné par la poste, le *U.S. Census Bureau* déploie beaucoup d'efforts afin de s'assurer que ses listes de recensement, de même que celles établies par les services postaux et plusieurs entreprises commerciales de listes de distribution, sont les plus complètes possible. Le point de départ du recensement aux États-Unis diffère de celui du recensement du Canada, où des recenseurs sont envoyés directement sur le terrain pour vérifier et établir les listes avant le jour du recensement.

Depuis quelques recensements, les États-Unis enregistrent des taux initiaux de non-réponse qui sont beaucoup plus élevés qu'au Canada et ils ont donc investi des ressources considérables afin que le recensement soit le plus complet possible. S'étant fixés pour objectif d'obtenir un dénombrement universel de droit de tous les logements, ménages et personnes, les États-Unis ont affecté plus de 300 000 recenseurs lors du Recensement de 1990 dans ce pays pour faire le suivi des non-réponses. Compte tenu des coûts générés par la non-réponse, les plans pour le recensement de l'an 2000 dans ce pays prévoient une plus grande utilisation des méthodes statistiques pour la collecte des données auprès des ménages les plus difficiles à dénombrer. En termes

plus précis, les États-Unis s'attendent à un taux initial de non-réponse d'environ 60 %. Lorsque 90 % de ces cas de non-réponse initiale auront été dénombrés, le bureau américain prévoit utiliser des techniques d'échantillonnage pour estimer les caractéristiques des 10 % restants; cela signifie qu'il y aura imputation des données pour environ 10 à 15 millions d'enregistrements.

En alliant ainsi des techniques d'échantillonnage à des contrôles améliorés de la qualité, par le biais d'une enquête postcensitaire et d'une analyse démographique, le *U.S. Census Bureau* espère obtenir un «recensement avec un nombre unique» où tous les chiffres pour chaque unité auront été corrigés en fonction de la couverture. En ne suivant qu'un échantillon du segment de la population qui est le plus difficile à recenser, le bureau américain espère libérer des ressources qui pourront être affectées à l'application de méthodes plus rigoureuses de contrôle de la qualité et ainsi obtenir un recensement de meilleure qualité, à plus faible coût¹¹. On ignore quelle erreur risque d'être introduite par les méthodes d'échantillonnage et d'imputation.

2.2 Techniques utilisant les micro données pour l'évaluation du recensement aux États-Unis

Bien que les États-Unis aient fait l'essai de diverses techniques utilisant les micro données, l'enquête postcensitaire demeure le principal outil utilisé par le bureau du recensement de ce pays pour mesurer la couverture. Cette enquête n'est pas pour autant fixe. Elle évolue à chaque recensement. Ainsi, l'enquête postcensitaire de 1990 prévoit une révision des enquêtes postcensitaires précédentes,

11 Pour plus de détails, voir *U.S. Bureau of the Census*, 1996.



dont la première à l'échelle du pays remonte à 1950¹². Après avoir étudié la possibilité d'utiliser une technique de couplage des enregistrements comme la CVD, l'organisme américain en est arrivé à la conclusion qu'une telle méthode ne convenait pas, dans un pays où les recensements sont conduits à intervalles de dix ans. Le couplage des enregistrements s'appuie en effet largement sur la recherche des enregistrements échantillonnés lors du recensement précédent, ce qui serait difficile à réaliser après dix ans. Pour le recensement de 1990, les États-Unis ont procédé à une enquête postcensitaire auprès d'un échantillon de quelque 166 000 unités de logement, soit 400 000 personnes¹³. De ce nombre, 22 000 unités de logement étaient inoccupées; 144 000 logements ont été identifiés comme étant occupés et 142 000 ménages ont en fait été interviewés (Hogan et coll., 1993, p.1049).

L'enquête postcensitaire de 1990 a été menée auprès de deux échantillons de base, soit l'échantillon P (échantillon de la population américaine obtenu par échantillonnage aréolaire) et l'échantillon E (échantillon du recensement prélevé de la

base de données du recensement). L'échantillon P représente la population-cible du recensement aux États-Unis et il est construit par échantillonnage aréolaire de grappes géographiques (en 1990, un échantillon de 5 300 grappes a été utilisé), alors que l'échantillon E est formé d'enregistrements du recensement choisis parmi les mêmes grappes. La proportion de l'échantillon P qui est en fait dénombrée, laquelle proportion donne une estimation du sous-dénombrement brut, est déterminée comme suit : après établissement d'une nouvelle liste des logements, il y a réinterview des ménages dans certains îlots, puis interview de suivi et appariement avec les enregistrements du recensement. Des techniques d'appariement automatique sont appliquées à l'échantillon E pour obtenir une estimation du surdénombrement, et un suivi distinct est effectué pour déceler les cas problèmes par le biais d'interviews sur place.

En 1990, les estimations du sous-dénombrement et du surdénombrement obtenues des échantillons P et E ont été combinées pour obtenir des estimations du sous-dénombrement net basées sur deux systèmes. Tout comme au Canada, cette technique utilisant les micro données est sujette à la fois à des erreurs d'échantillonnage et à d'autres non dues à l'échantillonnage. Tout a été fait en vue de réduire au minimum ces erreurs, en mettant au point des techniques d'imputation exhaustives pour les données manquantes, en ajustant les procédures et en utilisant d'autres stratégies de stratification a posteriori. Le bureau américain a par ailleurs fait de nombreuses études sur les effets des diverses sources d'erreurs, notamment une analyse de sensibilité de l'effet de la technique d'imputation pour les données manquantes. Ces erreurs ont été combinées en une estimation de l'erreur globale, à l'aide d'un modèle de l'erreur totale (Mulry et Spencer, 1993), dont le but est de permettre l'évaluation subséquente de

12 Pour un examen complet des méthodes utilisées pour l'enquête postcensitaire aux États-Unis, voir *Journal of the American Statistical Association*, volume 88, n°423, 1993.

13 Un échantillon beaucoup plus petit (environ 56 000 personnes) a été choisi pour la CVD au Canada, en 1991. Au moment de comparer la taille des deux échantillons, il est important toutefois de tenir compte des différences dans l'efficacité statistique des deux plans d'échantillonnage. L'effet du plan de sondage de l'enquête postcensitaire américaine (c.-à-d. le ratio entre la variance d'échantillonnage d'une estimation nationale basée sur le plan et la variance d'échantillonnage de la même estimation basée cette fois-ci sur un échantillon aléatoire simple de même taille) est supérieur à celui de la CVD, dont la taille aurait dû être de 15 % supérieure si on avait voulu utiliser le même plan (Dick, 1996).



l'effet de toutes les sources d'erreur dans l'enquête postcensitaire, tant à l'échelle du pays que des États.

Le bureau américain a produit trois séries d'estimations différentes pour le recensement de 1990, qui sont essentiellement le résultat des discussions menées par ses propres statisticiens et démographes au sujet des questions de méthodologie¹⁴. Les premières estimations du sous-dénombrement net, établies par l'enquête postcensitaire, ont été publiées en juillet 1991; ces estimations ont été révisées en janvier 1992, puis à nouveau en juillet 1992. Les estimations révisées de janvier 1992 visaient à régler les problèmes techniques ou de calcul décelés dans les estimations de juillet 1991, ainsi que les taux élevés de variance d'échantillonnage dans ces estimations et les problèmes concernant les techniques de lissage utilisées en grande partie pour réduire la variabilité d'échantillonnage. Les estimations révisées de juillet 1992 différaient sensiblement. Ainsi, les 1 392 estimations initiales directes par stratification a posteriori ont été regroupées et leur nombre a été réduit à 357, pour diminuer la variabilité d'échantillonnage et éliminer le recours au lissage, et l'ensemble de données de base de l'enquête postcensitaire a été modifié pour réduire les biais (voir Hogan et coll., 1993, p.1052-53). La révision de juillet 1992 a fourni les estimations officielles du sous-dénombrement net qui sert actuellement pour la pondération d'un grand nombre d'enquêtes menées par le *U.S. Census Bureau*, dont la *Current Population Survey* (CPS). L'analyse démographique a révélé des incohérences dans les estimations obtenues par l'enquête postcensitaire, au niveau des profils âge-sexe et des taux relatifs de

sous-dénombrement net par région. En outre, le profil âge-sexe du sous-dénombrement net différait de celui observé au Canada. Les estimations de juillet 1992 faisaient ainsi état d'un sous-dénombrement net parmi les enfants de race blanche (0 à 17 ans) qui était presque identique à celui observé chez les jeunes adultes (18 à 29 ans). Or, au Canada, le sous-dénombrement net des jeunes adultes était bien des fois supérieur à celui des enfants¹⁵.

Comme on fera un plus grand usage des méthodes statistiques pour le processus de collecte des données lors du recensement de l'an 2000 aux États-Unis, il s'impose d'intensifier les recherches pour mettre au point des méthodes plus rigoureuses de contrôle de la qualité. Comme le prévoit *The Plan for Census 2000* (*U.S. Census Bureau*, 1996), un «deuxième» recensement de 750 000 adresses sera mené auprès d'un nombre suffisant d'unités de logement dans chaque État pour assurer l'intégrité du processus de répartition et l'intégralité du recensement de l'an 2000. Comme des techniques d'échantillonnage seront utilisées pour le processus de collecte des données, il faudra également prévoir les méthodes les meilleures pour évaluer la couverture. Le *U.S. Census Bureau* participe actuellement à des recherches visant à évaluer les répercussions de ce changement sur les techniques de recherche et d'appariement utilisées pour estimer le sous-dénombrement. Les résultats provisoires du recensement d'essai de 1995 semblent indiquer la nécessité de poursuivre les recherches sur les estimations et les révisions par double système, de manière à obtenir des estimations plus stables et sans biais. Des techniques susceptibles de remplacer les estimations par double système ont

14 L'enquête postcensitaire de 1980 aux États-Unis a produit douze séries différentes d'estimations par double système (Fay et coll., 1988).

15 La répartition selon l'âge et le sexe n'est pas conforme non plus avec celles observées au Royaume-Uni ou en Australie.



été testées, mais sans succès, durant l'élaboration d'un programme de mesure intégré pour le recensement de l'an 2000 (Robinson, 1996).

2.3 Analyse démographique et évaluation de l'erreur de couverture du recensement aux États-Unis

Outre l'enquête postcensitaire, des méthodes démographiques sont utilisées depuis plusieurs décennies pour évaluer la couverture du recensement aux États-Unis. Comme indiqué précédemment, la logique qui sous-tend l'utilisation des méthodes démographiques est simple; le bureau américain a toutefois rencontré plusieurs problèmes pratiques qui ont empêché la production d'estimations qui soient, à la fois, indépendantes de l'enquête postcensitaire et des méthodes du recensement et aussi d'une qualité acceptable.

La question est de savoir s'il est possible d'obtenir des estimations démographiques suffisamment précises pour donner une bonne indication de la nature et de l'étendue du sous-dénombrement du recensement. Pour avoir confiance dans la qualité des estimations démographiques du sous-dénombrement net, les démographes doivent d'abord être assurés de la qualité des statistiques de l'état civil et des données de migration qui sont utilisées pour produire ces estimations. Par le passé, les analyses démographiques de la couverture du recensement aux États-Unis ont largement consisté à évaluer et à corriger les séries chronologiques, en s'appuyant sur la cohérence interne prouvée des données

démographiques, d'un recensement à un autre. Pour l'application de l'équation démographique de base aux données américaines, plusieurs ajustements ont été apportés aux séries chronologiques. Les recherches sur l'intégralité de l'enregistrement des naissances ont notamment mené à une correction systématique de la série chronologique des naissances.

Ansley Coale (1955) a été le premier démographe à appliquer l'analyse démographique aux données américaines, pour l'évaluation du recensement de 1950. Son intérêt pour la comptabilité démographique a amené certains auteurs à dire de lui qu'il avait jeté les bases de l'application de l'analyse démographique aux données du recensement aux États-Unis (Himes et Clogg, 1992). Lors d'une recherche antérieure, l'application directe de la méthode par composantes de base s'était soldée par des résultats insatisfaisants, imputables à l'absence de preuves quant à l'intégralité des données sur l'enregistrement des naissances avant 1940, ainsi que de preuves concluantes sur la qualité des autres séries chronologiques. La qualité des estimations démographiques directes pour la plupart des cohortes était ainsi compromise par le biais associé, entre autres, au sous-enregistrement des naissances. En conséquence, plusieurs techniques ont été mises au point pour estimer la population sur la base de données imparfaites. La recherche de Coale est représentative de celles qui ont été menées durant les années 50 et 60, bien que celle-ci ait depuis été révisée et poussée plus loin par plusieurs démographes affiliés au *U.S. Census Bureau* pour les recensements de 1960, 1970 et 1980 (Siegel et Zelnik, 1966; *U.S. Bureau of the Census*, 1977; Fay, Passel et Robinson, 1988).

Lorsqu'il a appliqué pour la première fois les techniques démographiques à l'analyse de la



couverture du recensement, Coale a eu accès aux résultats de deux tests du système d'enregistrement des naissances, qui avaient été menés conjointement avec les recensements de 1940 et de 1950¹⁶. Ces tests indiquaient un sous-enregistrement de l'ordre de 7 % à 8 % de toutes les naissances en 1940 et de 2 % à 3 %, en 1950. Or un tel taux de sous-enregistrement ne pouvait qu'influer sur toutes estimations démographiques subséquentes. Fort de cette information, Coale a pu estimer directement la taille des cohortes plus jeunes en 1950; la qualité des données sur les naissances associées aux cohortes plus âgées à leur origine restait cependant inconnue. Vu l'absence de données sur la précision des données sur les naissances avant 1940, Coale a élaboré des stratégies pour pallier ce problème¹⁷. Coale et Zelnik (1963) ont notamment appliqué la technique de rétropolation aux données sur la population native du pays. Ils ont ainsi fait une rétropolation des cohortes d'un grand nombre de recensements, à partir des tables de survie existantes, pour obtenir d'autres estimations du nombre de naissances associées à chaque cohorte; ils ont ensuite fait la moyenne de ces estimations, les ont corrigées en fonction du sous-enregistrement et les ont

prolongées jusqu'au moment présent. Des recherches similaires ont été menées par Coale et Rives (1973), cette fois-ci dans le but d'obtenir de nouvelles estimations sur la population noire, ainsi que par Whelpton (1950) pour établir des estimations des naissances selon la race. C'est sur ces estimations que se sont basés les analystes du *U.S. Census Bureau* pour l'évaluation de certaines cohortes plus âgées (55 à 64 ans) lors du recensement de 1990.

Bien que les estimations établies notamment par Coale, Whelpton et Zelnik ont joué un rôle fondamental dans les premières applications de l'analyse démographique pour l'évaluation de la couverture du recensement, une proportion croissante de la population américaine a, à chaque recensement subséquent, été estimée directement à partir des coefficients de survie des naissances corrigés en fonction du sous-enregistrement¹⁸. En outre, en raison du cumul des erreurs qui se produit lors de l'utilisation de techniques démographiques pour faire des estimations pour les groupes plus âgés, le bureau américain a utilisé, pendant plusieurs recensements, les données administratives¹⁹ pour les personnes de 65 ans et plus. En 1990, les estimations démographiques correspondent essentiellement à la somme des naissances, des décès ainsi que de l'immigration et de l'émigration légales, complétée par les données sur le personnel des forces armées à l'étranger, les estimations des immigrants non inscrits et les données administratives précitées. Grâce aux estimations calculées à l'échelle nationale selon l'âge, le sexe et la race, l'application directe de la méthode par composantes

16 Ceci diffère de la situation de Lapierre-Adamcyck, qui n'avait pratiquement aucune information sur le degré de sous-enregistrement des naissances.

17 Citons, par exemple, l'hypothèse de travail des erreurs similaires d'un recensement à l'autre, par Coale. Sur la base de données imparfaites, les estimations pour les cohortes plus âgées ont été obtenues en présumant d'un profil constant du taux de sous-dénombrement net, d'un recensement à un autre, selon l'âge et le sexe. Les taux de sous-dénombrement estimés pour les cohortes plus jeunes en 1950 ont été appliqués aux cohortes du même âge des recensements précédents. Les chiffres révisés ont ensuite été appliqués à 1950, pour obtenir des estimations pour les cohortes plus âgées. Ce processus a été répété jusqu'aux cohortes les plus âgées; malheureusement, on observe une accumulation des erreurs dans les cohortes plus âgées.

18 En 1990, seules les estimations de Whelpton sur les naissances ont été utilisées pour les estimations démographiques des personnes âgées de 55 à 64 ans.

19 Données sur l'assurance-maladie, corrigées en fonction de la sous-inscription.



de base permet d'obtenir des estimations indépendantes du recensement²⁰.

Les résultats de l'analyse démographique diffèrent sensiblement de ceux de l'enquête postcensitaire, tout comme ils diffèrent des résultats obtenus au Canada. L'analyse démographique indique, par exemple, un surdénombrement estimé à +1,66 % dans le groupe des 15 à 19 ans et de +0,02 % chez les 20 à 24 ans lors du recensement de 1990 (Robinson et coll., 1993). Or ce surdénombrement estimé chez les 20 à 24 ans tranche nettement avec le profil depuis longtemps établi au Canada (par la CVD) et qui fait plutôt état d'un sous-dénombrement net de 7 % dans le groupe des 20 à 24 ans en 1991. Par ailleurs, l'analyse démographique et l'enquête postcensitaire ont toutes deux produit des taux de sous-dénombrement net plus élevés chez les enfants (0 à 17 ans) que chez les jeunes adultes (18 à 29 ans). Le taux de sous-dénombrement estimé chez les enfants, selon l'enquête postcensitaire, était ainsi de 3,2 %, comparativement à 3,0 % chez les jeunes adultes et les taux estimés par l'analyse démographique s'établissaient respectivement à 2,3 % et 1,1 %. Au Canada, par contre, les estimations du sous-dénombrement net chez les enfants, selon la CVD, ont toujours été nettement inférieures à celles chez les jeunes adultes (2 % à 3 % chez les enfants de 0 à 9 ans, comparativement au taux maximal de 7 % mentionné précédemment chez les 20 à 24 ans). Malgré l'absence d'explications satisfaisantes, ces écarts pourraient notamment être dus à des différences dans les opérations du recensement ainsi que dans les techniques et les données

utilisées pour mesurer le sous-dénombrement net. Il se peut également que les personnes omises lors du recensement aux États-Unis soient contrebalancées par le nombre relativement élevé de personnes comptées plus d'une fois -- que le profil du surdénombrement selon l'âge et le sexe diffère ou non sensiblement d'un pays à l'autre.

Le bureau du recensement a également examiné divers moyens de combiner les résultats de l'analyse démographique aux estimations du sous-dénombrement obtenues de l'enquête postcensitaire. Bell (1993), par exemple, a tenté de modifier les estimations de l'enquête postcensitaire de 1990 en utilisant les estimations démographiques des rapports de masculinité nationaux selon l'âge. Il a ainsi testé différentes stratégies pour incorporer l'information sur les rapports de masculinité, notamment des méthodes prévoyant un assouplissement d'une hypothèse fondamentale de l'enquête postcensitaire concernant l'indépendance des échantillons P et E. En fait, si la sélection des échantillons P et E satisfait le critère d'indépendance, il est possible d'obtenir des estimations basées sur le double système, qui soient exemptes de plusieurs sources de biais, incluant les personnes omises lors du recensement et de l'enquête postcensitaire subséquente (biais de corrélation entre les opérations du recensement et celles de l'enquête postcensitaire qui suit). On peut utiliser un point de départ plus réaliste, pour lequel cette hypothèse est quelque peu assouplie, en introduisant des rapports de masculinité très précis selon l'âge. À titre d'exemple, on peut présumer que le critère d'indépendance ne vaut que pour les femmes, dont la couverture est probablement meilleure, alors que les estimations pour les hommes sont établies à partir des rapports de masculinité obtenus par analyse démographique. Bell a poussé plus loin des recherches menées précédemment par Wolter (1990), qui a été le premier à utiliser les rapports de

20 En raison des lacunes des données sur la migration interne aux États-Unis, les recherches se poursuivent sur l'utilisation potentielle des estimations démographiques et des données administratives pour l'évaluation de la couverture du recensement au niveau des États et de vastes comtés (voir Robinson, 1995).



masculinité pour faire une révision systématique des estimations par double système. Conscient que les rapports de masculinité constituent un outil robuste pour l'évaluation et la correction des données du recensement, le bureau américain poursuit ses recherches dans ce domaine en préparation du recensement de l'an 2000.

2.3.1 Estimations démographiques et incertitude

Comme la fiabilité des estimations démographiques est d'autant réduite par l'incertitude associée aux composantes sous-jacentes, les analystes du bureau américain ont décidé de s'attaquer à ce problème en mettant au point ce que Das Gupta (1991) a désigné un modèle d'incertitude de l'analyse démographique. Le but de ce modèle est d'établir des estimations par intervalle de la population et du sous-dénombrement net²¹, à partir d'une théorie pseudostatistique basée sur des estimations par intervalle analogues, mais qualitativement différentes, des intervalles de confiance fréquentistes classiques. Ces intervalles sont considérés utiles pour évaluer la précision des estimations démographiques, sans pour autant être calculés directement de ces estimations au plan de la probabilité statistique.

Pour documenter ce modèle d'incertitude, le bureau américain a préparé une série de rapports qui font une évaluation de l'incertitude associée à la mesure de chaque composante de leurs estimations démographiques (Robinson, 1991a, 1991b, 1991c, 1991d; Woodrow, 1991a, 1991b; Robinson et Lapham, 1991; Robinson, Woodrow et Ahmed, 1991; Robinson, Ward et Spencer, 1991).

21 La probabilité est très forte (95 % ou 99 %) que la population réelle des États-Unis et le sous-dénombrement net se situent à l'intérieur des intervalles.

Le but était d'établir des estimations par intervalle pour chaque composante, à partir du consensus établi au sein des spécialistes, de l'erreur de mesure et des méthodes d'estimation du bureau américain. L'étude menée par Robinson (1991) illustre bien ce courant de recherches. Ce dernier a utilisé l'analyse de sensibilité pour examiner l'effet de sources hypothétiques d'erreur dans les données sur l'enregistrement des naissances. Comme toutes les naissances sont corrigées en fonction des enregistrements incomplets, il était logique pour Robinson d'examiner diverses sources d'erreur dans les tests sur l'enregistrement des naissances, incluant l'erreur d'échantillonnage et celle non due à l'échantillonnage²². Il s'est posé la question suivante: «Quel taux d'erreur d'échantillonnage et d'erreur non due à l'échantillonnage peut-on considérer comme raisonnable, ou réaliste, et le cas échéant quelle est l'incidence d'une telle erreur sur les estimations globales des naissances?». L'effet combiné de ces erreurs détermine les limites inférieure et supérieure du modèle global²³. La réplication de simulations et de jugements similaires concernant d'autres composantes permet d'obtenir une mesure globale de l'incertitude des estimations finales du sous-dénombrement net.

Certes, l'intervalle de confiance associé à l'enregistrement des naissances était relativement étroit, tout comme l'étaient ceux associés aux décès

22 Biais d'appariement, biais de corrélation entre les ensembles de données étudiés, biais d'extrapolation ou d'interpolation pour les années pour lesquelles les tests sur les enregistrements des naissances n'étaient pas disponibles, etc.

23 Après de multiples simulations, Robinson est arrivé à des intervalles pour les naissances variant de 1,3 % en 1965-70 à 6,6 % en 1935-40, parmi la population blanche. Des intervalles beaucoup plus larges, atteignant 16,7 % durant la période de 1935 à 1940, ont été observés parmi la population de race noire.



intercensitaires et à leur répartition selon l'âge, le sexe et la race. Par contre, les données sur la migration étaient sujettes à des erreurs d'estimation relativement grandes, compte tenu de la difficulté de calculer les estimations indirectes de l'immigration, de l'émigration et, plus particulièrement, des résidents non inscrits. Après avoir défini les intervalles d'incertitude associés à l'analyse démographique, il est apparu que certaines estimations du sous-dénombrement, établies par l'enquête postcensitaire, se situaient en dehors de ces intervalles. Ces écarts ayant été définis, il s'agit maintenant de poursuivre les recherches, afin d'expliquer les différences obtenues par différentes méthodes. Das Gupta (1990) a également élargi cette recherche en assignant des intervalles de confiance aux rapports de masculinité estimés selon l'âge.

Malgré la recommandation du *U.S. Census Bureau*, qui s'appuyait ainsi sur de vastes études sur la couverture du recensement aux États-Unis, le *Department of Commerce* a décidé de ne pas corriger les données du recensement de 1990. En conséquence, le directeur du bureau du recensement a décidé, en 1992, de ne pas corriger les estimations démographiques postcensitaires en fonction du sous-dénombrement. Ceci laissait toutefois au *Bureau of Labour Statistics* la possibilité de corriger les estimations démographiques aux fins des enquêtes de suivi, en particulier pour la *Current Population Survey* (CPS) qu'il a menée en janvier 1994. C'est ainsi que, pour la première fois aux États-Unis, une série officielle d'estimations démographiques corrigées en fonction du sous-dénombrement net est devenue disponible, à l'échelle du pays et des États, aux fins de la CPS.

3. Royaume-Uni

3.1 Techniques de collecte des données

Avec pour objectif le dénombrement universel de la population, le recensement décennal britannique s'appuyait auparavant sur les deux méthodes fondamentales de recensement, soit celles basées sur la population de droit (résidence habituelle) et la population de fait (présente le soir du recensement). Lors du recensement de 1991, les recenseurs ont recueilli l'information selon le lieu habituel de résidence. Jusqu'ici, la Grande-Bretagne a utilisé la méthode de la livraison et ramassage par le recenseur, pour toutes ses opérations concernant la population. Cette technique a été abandonnée au Canada, et les États-Unis et la Grande-Bretagne sont actuellement à revoir leurs méthodes. Pour le recensement d'essai de 1997, le Royaume-Uni prévoit faire l'essai de nouvelles méthodes de dénombrement, notamment l'établissement de listes d'adresses pré-imprimées pour le personnel sur le terrain, ainsi que la livraison des questionnaires du recensement et leur retour par le courrier. En outre, dans une petite région éloignée, une méthode basée entièrement sur l'envoi par le courrier sera mise à l'essai.

En 1991, les recenseurs devaient localiser toutes les propriétés pertinentes et en faire la liste, en plus de recueillir directement les questionnaires du recensement. Plusieurs vérifications de l'intégrité des données ont été menées dans le cadre des opérations sur le terrain, y compris des suivis par téléphone et en personne et la vérification des listes. Pour la première fois en Grande-Bretagne, on a également tenté, en 1991, d'inclure les ménages «à leur domicile», dont les membres étaient absents le soir du recensement. Pour ce faire, il a été permis pour la première fois que les formulaires du recensement soient remplis en retard, après le retour du ménage à son adresse, ou par imputation en copiant l'enregistrement d'un ménage apparié qui avait rempli le formulaire de bon gré. Comme dans les autres pays, les estimations du sous-



dénombrement net obtenues directement des études postcensitaires n'incluent pas les personnes omises, qui ont par la suite été ajoutées à la base de données du recensement par imputation. Le taux d'imputation au Royaume-Uni demeure élevé, à plus de trois fois l'estimation initiale du sous-dénombrement établie par l'enquête postcensitaire. Pour plus de détails, consulter le *British 1991 Census User's Guide* (Dale et Marsh, 1993).

3.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement en Grande-Bretagne

Les Britanniques effectuent une étude postcensitaire de chaque recensement depuis 1961, bien que leurs premiers efforts dans ce domaine aient fait l'objet de critiques. Au nombre des difficultés techniques qui ont surgi durant les premières études postcensitaires, mentionnons la petite taille de l'échantillon, les difficultés à recueillir des données exactes, les retards dans la mise en œuvre, ainsi que le manque apparent d'indépendance totale entre les techniques de collecte des données du recensement et des études postcensitaires. À cause de ces difficultés techniques, il a fallu consacrer beaucoup d'efforts à la vérification de ces estimations par analyse démographique, ainsi qu'à leur comparaison avec les données administratives. La position officielle de l'*Office of Population Censuses and Surveys* (OPCS, aujourd'hui *Office for National Statistics*, ONS) est la suivante : les estimations du sous-dénombrement obtenues par analyses utilisant les

micro données avant 1991 sont exactes et cohérentes avec les analyses de données agrégées.

Les estimations du sous-dénombrement, établies à partir de ces premières études, étaient prudentes. Bien que les enquêtes postcensitaires plus récentes soient plus rigoureuses, en ce qui a trait au plan d'enquête et aux méthodes de collecte de données, le taux de sous-dénombrement déclaré demeure relativement faible. En 1991, par exemple, l'estimation obtenue directement de l'enquête postcensitaire établissait ce taux à seulement un demi pour cent environ, ce qui est étonnement bas. Il convient toutefois de souligner que ce chiffre n'inclut pas les ménages dont les données étaient manquantes ou ont été imputées et qui ont été ajoutées directement à la base de données du recensement; si ces imputations avaient été incluses, le sous-dénombrement net aurait été bien des fois supérieur.

Un des facteurs proposés pour expliquer partiellement ces estimations prudentes du sous-dénombrement est le fait que l'étude postcensitaire britannique est dépendante, et non indépendante, du recensement (Diamond, 1994). Cette étude part en effet du recensement pour en vérifier les méthodes. Il devient donc très difficile de faire des estimations raisonnables sur les personnes omises à la fois du recensement et de l'étude postcensitaire (Diamond, 1994, p.135). Cette méthode fait contraste avec l'enquête postcensitaire par saisie et re-saisie qui est menée aux États-Unis et pour laquelle il y a sélection d'un échantillon initial de la population entièrement indépendant du recensement. Aux États-Unis, les estimations par double système ont été conçues expressément dans le but de résoudre ce problème, de manière à pouvoir identifier la



population omise à la fois lors du recensement et de l'enquête postcensitaire²⁴.

L'enquête postcensitaire britannique de 1991 a été basée sur un échantillon à plusieurs degrés de quelque 1 200 districts de recensement, devant faire l'objet d'un suivi systématique. Les registres des recenseurs ont été utilisés comme base de sondage, pour sélectionner les ménages devant faire l'objet d'une vérification. Le caractère dépendant de cette étude se voit au fait que celle-ci a consisté à vérifier les registres des recenseurs et l'information recueillie sur les questionnaires du recensement²⁵. La question importante est donc de savoir s'il résulte un biais dans l'étude postcensitaire, du fait que l'on utilise comme point de référence le registre et les questionnaires du recensement. Si, lors de l'enquête postcensitaire, l'intervieweur obtient des renseignements en accord avec l'enregistrement initial ou le formulaire du recensement, rien n'amène à douter de l'exactitude des données. À titre d'exemple, si les membres

extrêmement mobiles d'un ménage sont omis au départ, la probabilité de les retrouver lors de l'enquête postcensitaire peut également être faible.

Or si les personnes sont omises à la fois du recensement et de l'enquête postcensitaire, les estimations du sous-dénombrement qui en découlent comporteront un biais par défaut. Les estimations finales du sous-dénombrement net établies par l'organisme en 1991 sont beaucoup plus élevées que les estimations initiales; les méthodes d'analyse démographique ont démontré que le taux de sous-dénombrement estimé par l'étude postcensitaire était trop faible, puisque jusqu'à un million de personnes additionnelles avaient été omises lors du recensement. Comme rien n'indique que des problèmes similaires existaient en 1981, les difficultés plus grandes qu'ont éprouvées les recenseurs lors du recensement et de l'étude postcensitaire de 1991 ont donné lieu à bien des conjectures, mettant notamment en cause l'effet des impôts britanniques, une plus grande mobilité des cohortes plus jeunes, ainsi qu'un plus grand éloignement. Là encore, les données à l'appui de ce changement ont été obtenues par analyse démographique (Wiggins, 1993).

24 Même si l'échantillon pour l'enquête postcensitaire aux États-Unis a été choisi indépendamment du recensement, il faut reconnaître que les probabilités réelles de saisie n'étaient pas toujours entièrement indépendantes. À titre d'exemple, l'analyse démographique a révélé un biais de corrélation en ce qui a trait aux hommes de race noire, dont un nombre significatif ont pu être omis à la fois des enregistrements du recensement et lors des interviews de l'enquête postcensitaire.

25 L'étude postcensitaire britannique consiste à établir une liste des secteurs de dénombrement échantillonnés pour les comparer au registre initial des recenseurs, à vérifier les bâtiments qualifiés de non résidentiels, à vérifier la classification des ménages entre les ménages «absents au recensement» et les logements «inoccupés lors du recensement», à interviewer de nouveau les ménages lorsque le registre indique que le formulaire du recensement a été retourné pour comparer systématiquement les réponses à celles du questionnaire initial, ainsi qu'à vérifier les logements initialement classés parmi ceux à occupants multiples.

3.3 Méthodes démographiques pour l'évaluation de la couverture du recensement en Grande-Bretagne

Alors qu'aux États-Unis les débats sur la façon d'ajuster correctement les chiffres du recensement ont donné lieu à une abondance de données parfois contradictoires (ainsi qu'à de nombreux litiges), certaines techniques démographiques ont été utilisées avec succès en Grande-Bretagne pour évaluer et corriger les chiffres officiels de population. Il s'agit de l'équation d'équilibre classique, combinée à des informations de base sur les rapports de masculinité prévus. En



Grande-Bretagne, les derniers recensements ont été systématiquement évalués à l'aide d'une variété de mesures, incluant une comparaison systématique des chiffres du recensement avec les estimations postcensitaires correspondantes, obtenues à partir des données du recensement précédent entièrement corrigé sur la base des événements intercensitaires. Dans l'ensemble, les chiffres officiels des recensements de 1971 et de 1981, corrigés à l'aide des études postcensitaires, ont été jugés d'une grande exactitude, en grande partie du fait qu'une telle comparaison a établi un haut degré de cohérence intercensitaire.

L'OPCS a toutefois noté un écart significatif, lorsque les données du recensement de 1991 ont été comparées aux estimations postcensitaires basées sur le recensement de 1981. Les estimations dépassaient ainsi de 1 047 000 personnes les chiffres du recensement de 1991 corrigés en fonction du sous-dénombrement d'après l'enquête postcensitaire (*Office of Population Censuses and Surveys*, 1995, p. 113). Cet écart, qui semble indiquer un sous-dénombrement d'environ 2,2 %, était nettement supérieur à l'estimation de l'enquête postcensitaire qui établissait à moins de 300 000 (0,4 %) le nombre de personnes omises. Pour expliquer l'erreur de clôture observée en 1991, il a fallu réexaminer plusieurs sources d'erreur potentielle, sans écarter la possibilité de problèmes au niveau du recensement et de l'enquête postcensitaire de 1981, de problèmes liés à la mesure et à l'estimation des composantes utilisées pour établir les estimations postcensitaires de 1981 à 1991 et de problèmes au niveau du recensement et de l'enquête postcensitaire de 1991.

On ne peut se fier aux estimations postcensitaires pour évaluer la couverture du recensement, que s'il est possible d'identifier les problèmes dans la collecte ou l'estimation des données. Pour

expliquer l'écart d'environ un million de personnes en 1991, l'OPCS (1993) a examiné différentes hypothèses, notamment qu'il y ait pu y avoir surdénombrement significatif lors du recensement de 1981, qu'un nombre comparable de personnes aient été omises lors du recensement et de l'enquête postcensitaire de 1991 et que l'effet net des événements intercensitaires (décès, naissances et migration) ait été nettement exagéré. Enfin, même si l'on n'a pas écarté la possibilité que les enquêtes postcensitaires de 1981 et 1991 aient toutes deux sous-évalué le niveau réel de sous-dénombrement, cette possibilité n'a jamais été officiellement acceptée.

L'OPCS a rejeté la possibilité d'une erreur significative dans les études de couverture du recensement de 1981. Par ailleurs, l'hypothèse de travail concernant les statistiques de l'état civil implique que l'enregistrement soit complet, même s'il n'est fait mention d'aucune étude sur l'enregistrement des naissances ou des décès. En ce qui a trait à la composante la plus douteuse des estimations démographiques – la migration internationale – la position officielle de l'OPCS est que le biais prévu, s'il en est un, va plutôt dans la direction opposée à celle suggérée par l'erreur de clôture. Si l'erreur de clôture observée était attribuable entièrement à des erreurs de déclaration au niveau de la migration internationale nette, alors l'arrivée nette de quelque 400 000 personnes durant la période 1981 à 1991 serait nécessairement remplacée par un exode net de plus de 600 000. Selon le consensus qui se dégage parmi les analystes de Grande-Bretagne, il est plus probable que l'immigration soit sous-évaluée que surévaluée.

Les rapports de masculinité déclarés selon l'âge, pour les chiffres corrigés et non corrigés du recensement, sont venus étayer l'hypothèse de l'existence de problèmes avec les chiffres de 1991.



En effet, les rapports de masculinité déclarés chez les jeunes adultes semblent peu vraisemblables, compte tenu de ce que l'on sait du rapport de masculinité à la naissance, ainsi que de la mortalité et de la migration chez les jeunes hommes et les jeunes femmes. Alors que les rapports de masculinité selon les chiffres du recensement de 1981 et les estimations postcensitaires de 1991 étaient en accord avec les prévisions, des rapports de masculinité beaucoup plus faibles que prévus ont été observés chez les personnes âgées de 20 à 25 ans lors du recensement de 1991. Cette observation concorde avec la conclusion que l'étude postcensitaire n'a pas réussi à déceler toute l'erreur de couverture chez les jeunes hommes. Le consensus au sein du personnel de l'OPCS étant que l'exactitude des chiffres basés sur le recensement de 1991 laissait à désirer, l'OPCS a pris une décision sans précédent de ne pas ajuster les chiffres de population sur le recensement de 1991.

Dans le groupe des 0 à 44 ans, l'écart entre les estimations postcensitaires et les chiffres du recensement a été attribué entièrement à des problèmes liés du recensement et à l'étude postcensitaire de 1991 (OPCS, 1993b, p.25). Pour justifier le fait qu'aucune portion de l'erreur ne soit attribuée au recensement de 1981 ou aux composantes intercensitaires, on a fait valoir que la taille de ces erreurs était impossible à quantifier et que toute valeur serait arbitraire. Malgré l'absence de données empiriques indiquant quelle proportion de l'erreur observée aurait pu être due à des problèmes en 1981 ou à des erreurs dans les données sur la migration – plutôt qu'aux chiffres du recensement de 1991 – les estimations postcensitaires pour ces groupes d'âge ont néanmoins été acceptées. Là encore, cet ajustement a été justifié par l'évidence d'une cohérence intercensitaire entre les données du recensement et les rapports de masculinité observés. Pour les groupes plus âgés

(45 à 79 ans), ce sont les chiffres du recensement corrigés par l'étude postcensitaire qui ont été acceptés. Enfin, chez les personnes de 80 ans et plus, l'analyse démographique a révélé un déficit dans les données du recensement corrigées par les résultats de l'enquête (Heady, Smith et Avery, 1993, p.39); les chiffres de l'analyse démographique ont donc été retenus.

Le profil du sous-dénombrement net selon l'âge et le sexe en Grande-Bretagne, établi par les estimations postcensitaires, se rapproche davantage des résultats de la CVD au Canada que des résultats obtenus aux États-Unis. Ainsi, chez les jeunes adultes, le taux de sous-dénombrement net culmine chez les jeunes femmes dans le début de la vingtaine et chez les jeunes hommes – dont les taux sont plus élevés – vers la fin de la vingtaine. Autant chez les hommes que chez les femmes, les taux de sous-dénombrement étaient comparables aux taux observés au Canada, pour ces groupes d'âge. À un âge plus avancé (à l'exception de la population des 80 ans et plus), un taux négligeable de sous-dénombrement a été observé chez les deux sexes, tandis que le sous-dénombrement net chez les enfants se situe entre le taux des groupes plus âgés et celui des jeunes adultes.

Enfin, comme aux États-Unis et au Canada, les méthodes démographiques utilisées pour estimer, à l'échelle nationale, le sous-dénombrement net selon l'âge et le sexe ne peuvent pas, pour l'instant, être appliquées au niveau infranational, en raison principalement de l'imprécision des données sur la migration interne²⁶. De même, on ne peut pas présumer que le taux et le profil âge-sexe du sous-dénombrement observé à l'échelle nationale s'appliquent également

26 Comme nous le verrons ci-après, la même généralisation vaut pour l'Australie.



à l'échelon infranational²⁷. Il faut donc utiliser d'autres méthodes pour répartir le sous-dénombrement estimé entre les régions. Pour ce faire, on a utilisé une distribution proportionnelle du nombre dans chaque groupe selon l'âge et le sexe entre toutes les régions, afin d'obtenir des estimations provisoires puis, à partir de l'information sur les rapports de masculinité selon l'âge pour l'ensemble du pays et certaines régions en 1971, 1981 et 1991, les chiffres distribués au prorata ont été corrigés de façon systématique pour éviter d'obtenir des rapports de masculinité invraisemblables (OPCS, 1993).

4. Australie

4.1 Méthodes de collecte des données

Tout comme au Canada, le recensement en Australie a lieu tous les cinq ans. Bien qu'il s'agisse d'un recensement basé sur la population de fait et exécuté selon la formule de livraison et ramassage des questionnaires, certaines données sont disponibles sur une base de droit. Pour le dénombrement de la population selon l'emplacement des gens le soir du recensement, le bureau australien confie aux recenseurs la tâche de dresser la liste des logements, d'établir des contacts directs avec les occupants de tous les logements occupés, de remettre aux résidents les formulaires du recensement à remplir soi-même, puis de recueillir tous les formulaires. Pour assurer un recensement le plus complet possible, plusieurs vérifications sont incluses dans les opérations sur le terrain, notamment un suivi par téléphone et en personne fait par les recenseurs, ainsi que la vérification des

listes par les superviseurs des recenseurs (Castles, 1989).

4.2 Techniques utilisant les micro données pour l'évaluation de la couverture du recensement en Australie

La correction des chiffres du recensement en Australie, comme en Grande-Bretagne, a suscité peu de controverse, les Australiens ayant recours à l'enquête postcensitaire, combinée à l'analyse démographique, pour toutes les estimations démographiques depuis 1971. Le plan d'enquête est lui aussi demeuré relativement constant au fil des ans; il consiste essentiellement en un échantillonnage aréolaire à plusieurs niveaux de la population australienne, de manière à obtenir un échantillon des logements privés qui soit le plus indépendant possible des opérations du recensement. L'enquête postcensitaire australienne est conçue de manière à déterminer à la fois le sous-dénombrement et le surdénombrement dans l'échantillon aréolaire choisi.

Si l'on examine maintenant l'enquête postcensitaire australienne en regard de la distinction que fait Diamond (1994) entre enquête dépendante et indépendante, l'enquête australienne se rapproche davantage de la dernière, puisqu'on s'efforce de réduire au maximum l'interaction entre le recensement et l'enquête postcensitaire. Ainsi, l'enquête postcensitaire australienne, comme son homologue aux États-Unis, n'est ni effectuée par les recenseurs sur le terrain, ni n'utilise les listes et les questionnaires du recensement. Dans la mesure où l'étude postcensitaire et le recensement sont statistiquement indépendants, «les facteurs qui contribuent à l'omission ou au surdénombrement d'une personne lors du recensement peuvent ou non se retrouver également dans l'enquête postcensitaire».

27 Pour une description complète des méthodes utilisées pour obtenir les estimations démographiques corrigées en fonction du sous-dénombrement à l'échelon infranational, voir OPCS, 1993b.



re, ce qui donnera lieu à une estimation biaisée du sous-dénombrement».

En 1991, un échantillon de quelque 40 000 logements privés a été prélevé. Les données auprès des ménages choisis ont été recueillies par le biais d'interviews sur place, réalisés trois semaines après le jour du recensement. Les personnes qui résidaient dans le pays le jour du recensement ont été questionnées sur leur lieu de résidence habituel, ainsi que sur leur adresse le jour du recensement et de toute autre adresse susceptible d'être indiquée sur un formulaire du recensement. L'appariement systématique de l'échantillon de l'enquête postcensitaire et du recensement a permis d'estimer le nombre de personnes qui auraient dû être dénombrées lors du recensement, ainsi que le nombre de personnes qui l'ont réellement été. L'enquête postcensitaire australienne ayant été conçue de manière à estimer à la fois le nombre de personnes omises et le nombre de personnes surdénombrées, elle permet d'établir des estimations nationales après stratification a posteriori, selon l'âge et le sexe. L'erreur d'échantillonnage a manifestement été plus élevée aux niveaux inférieurs de stratification géographique, et plus particulièrement pour les catégories détaillées selon l'âge et le sexe. Dans les cas où les données sur l'adresse étaient vagues et où il a été impossible de faire un appariement avec les formulaires pertinents du recensement, les données ont été imputées (Madden, 1995).

Alors que les premières études postcensitaires menées en Australie ne faisaient état que de très faibles taux de sous-dénombrement, les taux en 1976 étaient devenus comparables à ceux enregistrés au Canada ou aux États-Unis. En 1991, le sous-dénombrement net total en Australie a été estimé à environ 1,8%, ce qui représente une baisse négligeable par rapport à 1986 (1,9 %). En ce qui a

trait à la répartition du sous-dénombrement selon l'âge et le sexe, les chiffres de 1991 pour l'Australie indiquent un profil très similaire à celui du Canada, qui se caractérise par un taux maximal d'environ 4,4 % chez les personnes de 20 à 24 ans -- hommes et femmes; notons toutefois que le taux réel est loin d'être aussi élevé qu'au Canada (*Australian Bureau of Statistics*, 1995). Ensuite, une comparaison systématique est faite entre les résultats de l'enquête postcensitaire et les sources de données administratives, comme les dossiers d'allocations familiales et les dossiers d'assurance-maladie, et les corrections nécessaires sont apportées aux cellules âge-sexe appropriées de l'enquête postcensitaire, si les données administratives ou l'analyse démographique le justifient.

Comme l'erreur d'échantillonnage associée aux estimations du sous-dénombrement net total, selon l'âge et le sexe, au niveau infranational (États et territoires) a été considérée d'un niveau élevé inacceptable, le bureau australien a eu recours à la technique d'ajustement proportionnel itératif (Purcell et Kish, 1989; Purcell, 1979; Bagnall, 1983) pour estimer le sous-dénombrement net dans les États et territoires, selon l'âge et le sexe. Cette technique consiste à utiliser les estimations démographiques nationales, ventilées selon l'âge et le sexe et corrigées en fonction des résultats de l'enquête postcensitaire, les estimations corrigées par État et territoire pour chaque sexe, ainsi que les chiffres non corrigés du recensement pour chaque État et territoire, par âge et sexe. Les chiffres non corrigés du recensement sont ajustés aux deux autres sources de données, par le biais de l'algorithme d'ajustement proportionnel itératif²⁸.

28 Cette procédure a été décrite en détails pour la première fois dans un document technique sur les estimations démographiques, publié en 1983 par le Bureau australien (Bagnall, 1983).



Une des contraintes à respecter était de ne pas avoir de surdénombrement dans aucun groupe d'âge. En 1991, la même procédure a été utilisée pour estimer la couverture du recensement au niveau des États et des territoires.

4.3 Méthodes démographiques pour l'évaluation de la couverture du recensement en Australie

Les estimations démographiques ont pu être compilées indépendamment du recensement de 1991, pour les personnes de 0 à 64 ans, en utilisant la banque nationale de données démographiques, qui est mise à jour par l'*Australian Bureau of Statistics* et qui réunit des données sur les naissances, les décès et la migration internationale depuis 1926. Les estimations pour les groupes plus âgés ont été établies à partir d'une longue série chronologique sur les naissances, complétée de données détaillées sur la mortalité ainsi que sur les arrivées de l'étranger et les départs vers l'étranger, selon l'âge et le sexe. L'hypothèse formulée est que la série chronologique historique sur les naissances et les décès en Australie est assez complète et qu'elle ne contient pas d'erreurs par omission (Hall, 1969). L'information sur la migration internationale étant elle aussi considérée fiable, les estimations démographiques selon l'âge et le sexe ont été utilisées à plusieurs reprises pour évaluer les estimations du sous-dénombrement net par l'enquête postcensitaire.

Cependant, la qualité de ces estimations démographiques diminue chez les groupes plus âgés, car la série chronologique qui doit alors être utilisée est plus longue²⁹. L'Australie est l'un des rares pays dont la proportion de la population née à

l'étranger est supérieure à celle du Canada³⁰. Aussi n'est-il pas surprenant que l'efficacité de l'analyse démographique de base régresse chez les groupes plus âgés. Des comparaisons détaillées avec l'enquête postcensitaire et les données administratives (par exemple, allocations familiales, assurance-maladie) ont mené à la conclusion que les résultats de l'enquête postcensitaire sont probablement plus valables, en ce qui concerne le sous-dénombrement net total, pour la plupart des catégories âge-sexe des groupes plus âgés.

Tout comme au Canada, les récentes estimations démographiques officielles ont été corrigées en mettant à profit l'analyse démographique. À titre d'exemple, le rapport de masculinité chez les enfants, qui avait été établi à partir des estimations postcensitaires de 1986, était considéré inacceptable, compte tenu de ce qu'on savait du rapport de masculinité à la naissance et des différences entre les sexes au niveau de la mobilité et de la migration chez les plus jeunes. Le nombre estimé de sujets masculins omis parmi les 0 à 4 ans a donc été augmenté quelque peu, pour se rapprocher de l'estimation démographique. Pour les mêmes raisons, les chiffres pour les autres groupes d'âge ont été modifiés légèrement, le nombre de jeunes hommes et de jeunes femmes étant révisé sur la base des rapports de masculinité établis par l'analyse démographique. En 1991, une évaluation systématique du sous-dénombrement du recensement a été faite à partir de l'enquête postcensitaire, des estimations démographiques décrites précédemment et des rapports de masculinité basés sur les données de l'assurance-maladie. Les Australiens continuent d'avoir une grande confiance dans leurs estimations postcensitaires de la couverture, sous réserve de

29 Cette observation vaut particulièrement pour les femmes, en 1991.

30 En juin 1995, la proportion des Australiens nés à l'étranger atteignait 23 %.



certaines rajustements pour certains groupes d'âge particuliers. Dans la dernière série d'estimations de 1991, un poids a été attribué à chaque source de données, afin qu'aucune ne l'emporte sur une autre lorsque la correction risquerait de donner des résultats irréalistes (McLennan, 1995, p.16). Dans son résumé de la fiabilité de l'enquête postcensitaire, Madden (1995, p.4) indique que des rajustements ont été apportés aux estimations de cette enquête pour les hommes de 60 à 64 ans, ainsi que pour les femmes de 35 à 39 ans, de 40 à 44 ans, de 65 à 69 ans et de 75 ans et plus³¹.

5. Discussion

La répartition actuelle du sous-dénombrement net selon l'âge et le sexe, définie par la méthode de contre-vérification des dossiers (CVD), est similaire aux résultats enregistrés en Grande-Bretagne et en Australie, et elle se caractérise par des taux de sous-dénombrement net particulièrement élevés chez les jeunes adultes, notamment chez les jeunes hommes. Les données américaines tranchent toutefois sur les autres dans le cadre de comparaisons internationales; on remarque en effet que le taux de sous-dénombrement chez les jeunes adultes est relativement faible -- qu'il soit établi par l'analyse démographique ou l'enquête postcensitaire. Cet écart pourrait être dû à un certain nombre de facteurs, notamment à des différences dans les opérations du recensement, par exemple dans la manière dont l'information est recueillie auprès des ménages et dont les listes sont établies, les

techniques d'imputation utilisées pour les données manquantes, le niveau relatif et la nature de la non-réponse et les stratégies prises pour réduire cette non-réponse, le biais de corrélation lorsque les mêmes personnes sont omises de l'enquête postcensitaire et du recensement, etc. Dans l'ensemble, les comparaisons internationales de la couverture du recensement sont extrêmement complexes, à cause notamment de toutes ces sources possibles de différence mais aussi d'autres facteurs, par exemple de la taille et des conditions de logement des familles et des ménages, des relations qu'entretiennent les organismes gouvernementaux des statistiques avec le public, du degré d'apathie ou d'aliénation du public, du niveau général d'alphabétisation de la population, etc.

La CVD étant considérée comme un nouveau dénombrement indépendant du recensement de la population du Canada³², il est peu probable que les problèmes observés en Grande-Bretagne en 1991 se présentent ici. Ces problèmes, dus à l'interdépendance du recensement et des dénombremens qui ont suivi, ont donné lieu à l'omission des mêmes personnes lors des deux opérations. Dans la mesure où les bases de sondage de la CVD sont complètes et établies indépendamment des opérations du recensement, il ne devrait y avoir qu'un nombre négligeable de personnes omises à la fois par le recensement et la CVD. Bien que les avantages de cette méthode soient évidents, peu d'études ont porté sur les limites conceptuelles, théoriques et pratiques de la CVD, par exemple sur les effets de la difficulté à retrouver et à apparier les enregistrements ou encore sur les effets des techniques d'imputation subséquentes. Dans un tel

31 Le Canada et l'Australie ont tous deux corrigé leurs estimations du sous-dénombrement net, à la lumière de l'analyse démographique. Cette pratique diffère de celle utilisée aux États-Unis où pareils ajustements n'ont pas été faits malgré certaines incohérences manifestes entre les estimations de l'enquête postcensitaire et de l'analyse démographique, pour les groupes détaillés ventilés selon l'âge, le sexe et la race.

32 Il faudrait plutôt parler d'un «nouveau dénombrement indépendant» du «recensement actuel», puisque de toute évidence une grande partie de l'échantillon de la CVD est choisi parmi des personnes dénombrées lors du recensement précédent.



contexte, les recherches américaines sur l'élaboration d'un modèle de l'erreur totale aux fins des enquêtes postcensitaires s'avèrent particulièrement instructives (Mulry et Spencer, 1993). Le bureau américain a alloué des ressources considérables pour l'évaluation systématique de l'erreur d'échantillonnage et de l'erreur non due à l'échantillonnage dans l'enquête postcensitaire, au niveau de l'ensemble du pays et des différents États³³.

Cependant, même si la CVD est considérée comme la mesure la plus fiable et la plus valable du sous-dénombrement net au Canada, cette méthode a pour autant des limites. La principale source d'erreur de la CVD se situe au niveau de l'erreur d'échantillonnage, qui augmente avec le degré de désagrégation. Un des principaux avantages de l'analyse démographique tient donc à sa capacité de fournir des estimations démographiques plus détaillées, ventilées par catégorie d'âge et de sexe. Ces améliorations sont le fruit des recherches qui ont été menées dans tous les pays dont les programmes ont été revus -- depuis les rajustements faits au Canada et en Australie, par exemple pour corriger les rapports de masculinité invraisemblables chez les enfants de 0 à 4 ans, à la décision de l'OPCS en Grande-Bretagne de ne pas modifier la population de base du recensement de 1991 en fonction des résultats de l'évaluation du recense-

ment et de l'enquête postcensitaire. Aux États-Unis, les résultats de l'analyse démographique devaient fournir un repère pour juger les résultats de l'enquête postcensitaire en 1990. Récemment, on a utilisé l'analyse démographique pour évaluer le succès du recensement d'essai de 1995 aux États-Unis, en prévision du recensement de 2000 (Robinson et coll., 1996).

La situation du Canada et de l'Australie diffère de celle des États-Unis et de la Grande-Bretagne sous un aspect important, celui de la migration internationale des dernières décennies. Les lacunes dans la mesure de la migration internationale rendent plus difficile l'application de la méthode d'analyse démographique de base. Comme indiqué précédemment, ces lacunes ont pour effet de fausser, au Canada et en Australie, les estimations du sous-dénombrement net établies par la méthode démographique de base pour les groupes plus âgés. Bien que les estimations de la migration internationale pour une période particulière puissent être acceptables, leur précision diminue à cause des erreurs cumulatives dans les estimations de la taille des groupes plus âgés. Par ailleurs, même si les niveaux d'immigration ne sont pas aussi élevés aux États-Unis, il existe néanmoins dans ce pays certains segments de la population qui sont très difficiles à dénombrer, en particulier la population résidente non inscrite. Il s'agit d'une population qui n'est pas sans importance et sur laquelle on possède peu d'information fiable, notamment en ce qui a trait à sa répartition par âge et par sexe. Pour sa part, la Grande-Bretagne, dont la proportion de personnes nées à l'étranger est similaire à celle des États-Unis, n'a pas ce problème de segments de la population non inscrits, ce qui élimine un obstacle majeur dans l'établissement des estimations démographiques.

33 Les problèmes entourant l'enquête postcensitaire aux États-Unis ont fortement ravivé l'intérêt pour l'étude des biais résultant de l'erreur de repérage et de l'erreur non due à l'échantillonnage. Les problèmes dus notamment aux personnes qui déménagent, aux résidents temporaires, aux logements inoccupés, etc. sont tous significatifs. Comme les personnes forment l'unité d'échantillonnage de la CVD au Canada, celles-ci sont faciles à retrouver pour résoudre les problèmes. La CVD n'est pas pour autant dépourvue de biais, lesquels sont manifestement concentrés dans les personnes impossibles à retrouver. Des problèmes similaires existent avec l'enquête postcensitaire américaine.



Il s'impose de poursuivre les recherches sur l'utilisation des techniques démographiques pour l'évaluation du recensement au Canada. Quant à la pertinence des données administratives à cette fin, il faut dans ce domaine également approfondir les recherches sur la qualité de ces données et la façon de les utiliser pour l'évaluation. Comme on l'a constaté aux États-Unis, il devient particulièrement difficile d'obtenir des estimations démographiques fiables pour les groupes plus âgés, en raison de l'effet de l'erreur cumulative dans les composantes de la croissance. C'est pourquoi les États-Unis ont utilisé les dossiers corrigés de l'assurance-maladie pour évaluer la couverture des groupes âgés de 65 ans et plus. Il est donc recommandé de poursuivre les recherches au Canada sur des applications similaires des données administratives, en particulier pour les groupes plus âgés. Il reste toutefois à savoir si les ajustements et les imputations nécessaires pour travailler avec ces données peuvent être justifiés dans le contexte canadien.

Certaines recommandations formulées par Lapierre-Adamcyk (1970), il y a plus de 25 ans, valent encore aujourd'hui, notamment celle engageant à poursuivre les recherches sur la qualité des données utilisées pour l'établissement des estimations démographiques, et plus particulièrement sur l'exactitude des statistiques sur les naissances et les décès et des données administratives sur la migration internationale. Ces travaux sont jugés depuis longtemps prioritaires aux États-Unis, comme en font foi les recherches importantes qui y ont été menées. Le modèle d'incertitude de Das Gupta (1991), par exemple, témoigne des nombreux efforts qui ont été faits en vue d'évaluer la précision des composantes démographiques au fil des ans. Il reste maintenant à savoir dans quelle mesure de telles recherches pourraient s'appliquer au contexte canadien.

C'est dans cette optique que Dionne (1995) a proposé l'utilisation de modèles multidimensionnels pour estimer la population et évaluer la couverture du recensement. L'objectif est d'obtenir des estimations robustes du sous-dénombrement net, indépendamment de la CVD, en établissant une structure des relations intergénérationnelles, essentiellement à partir des statistiques de l'état civil. Le succès de cette méthode dépendra de la sensibilité des modèles aux problèmes liés à la qualité des données. Lorsqu'on travaille avec des données imparfaites, la question est de savoir s'il est possible d'obtenir des estimations raisonnables du sous-dénombrement net, de déterminer dans quelle mesure ces estimations diffèrent lorsqu'on utilise des méthodes différentes (incluant des méthodes démographiques différentes) et de pouvoir expliquer les écarts observés. Les recherches sur ces questions se poursuivront.

L'examen présente ici des méthodes utilisées dans les autres pays est loin d'être complet, car bon nombre de pays d'Europe de l'ouest procèdent eux aussi à des enquêtes postcensitaires pour évaluer leur recensement. Dans un examen des méthodes utilisées en Europe de l'Ouest, Begeot et coll. (1993) parlent ainsi d'enquêtes postcensitaires menées en Belgique, en Grèce, en Espagne, en France, en République d'Irlande, en Italie, au Luxembourg, au Portugal, en Autriche, en Norvège, en Suède et en Finlande. Outre l'Europe, la Nouvelle-Zélande a l'intention elle aussi de corriger ses chiffres du recensement en fonction du sous-dénombrement net en 1996. Dans ces pays, la méthode la plus souvent utilisée à cette fin est l'enquête postcensitaire, à laquelle s'ajoutent, dans bon nombre de pays d'Europe de l'Ouest, des comparaisons systématiques avec les registres de population. De tous les pays mentionnés, seule la République d'Irlande a recours exclusivement aux



méthodes démographiques, utilisant les estimations postcensitaires pour évaluer les résultats du recensement.

6. Conclusion

Plusieurs méthodes, mises au point par des démographes et des statisticiens, ont été utilisées pour évaluer la couverture des données du recensement. Le but du présent rapport était de faire une revue des méthodes utilisées au Canada, aux États-Unis, en Australie et en Grande-Bretagne. Nous y avons décrit brièvement les techniques de collecte des données dans chaque pays, ainsi que les techniques de utilisant les micro données (couplage d'enregistrements et recherche), puis nous avons examiné les méthodes d'analyse démographique qui constituent le volet principal de notre étude. L'importance accordée aux méthodes démographi-ques, par rapport aux autres techniques, varie d'un pays à l'autre, depuis une dépendance presque totale – comme pour l'estimation de la population de 1991 en Grande-Bretagne – à leur utilisation pour évaluer et corriger les résultats des méthodes utilisant les micro données, comme cela se fait en Australie.

En conclusion, l'auteur suggère d'entreprendre des recherches dans les domaines suivants, afin d'améliorer les estimations faites par Statistique Canada de la couverture du recensement.

- 1) Il faudrait faire une évaluation plus exhaustive des séries chronologiques disponibles sur les naissances, l'immigration, l'émigration, les Canadiens de retour, les résidents non permanents et les décès, incluant une étude de l'intégralité de l'enregistrement des naissan-ces et des décès.
- 2) Les résultats de la CVD et des méthodes démographiques devraient être évalués et comparés de façon plus systématique, en regard notamment des incohérences qui ressortent de l'utilisation seulement des méthodes démographiques.
- 3) Il faudrait poursuivre les recherches sur d'autres méthodes démographiques suffisamment robustes pour faire face aux problèmes de qualité des données (p. ex. le modèle inter-générationnel).
- 4) L'utilité d'autres sources de données administratives devrait être évaluée de façon systématique, en particulier pour l'estimation de la couverture des groupes plus âgés (dossiers d'assurance-maladie, d'impôt et de sécurité sociale).
- 5) L'utilisation des rapports de masculinité devrait être envisagée plus sérieusement, ceux-ci offrant un outil robuste pour l'évaluation des données du recensement.
- 6) Il faut pousser plus loin les recherches sur les problèmes éventuels associés à la contre-vérification des dossiers, en particulier en ce qui a trait à l'erreur non due à l'échantillonnage et à l'utilisation possible de l'analyse de sensibilité pour en évaluer les effets.
- 7) Enfin, il convient de procéder à une évaluation des techniques de vérification et d'imputation du recensement, pour étudier les distorsions possibles dans la répartition selon l'âge et le sexe.



Au Canada, comme ailleurs, on reconnaît l'importance pour les statisticiens et les démographes d'unir leurs efforts dans le cadre des recherches sur la couverture du recensement. Les méthodes démographiques peuvent aider à évaluer la CVD, en fournissant des estimations indépendantes de la population, selon l'âge et le sexe. Dans l'ensemble, l'avantage des méthodes démographiques réside dans la cohérence qu'elles offrent d'un recensement à un autre au niveau de cohortes particulières, tandis que l'utilité des chiffres du recensement corrigés en fonction des études de cou-

verture tient aux totaux ainsi générés, tant au niveau national que des provinces. Par ailleurs, le fait d'obtenir des résultats similaires par le biais de différentes méthodes ne fait qu'ajouter à la crédibilité des estimations officielles de la population. Compte tenu du coût raisonnable de l'analyse démographique - qui fait déjà partie intégrante du programme d'estimations démographiques de Statistique Canada - il semble justifié d'affecter davantage de ressources au développement d'indicateurs démographiques des problèmes de couverture du recensement.



Références

- Australian Bureau of Statistics (1995). Population Estimates and Concepts, Sources and Methods, n° 3228-0 au catalogue.
- Bagnall, A.R. (1983). Methods and Procedures in the Compilation of Estimated Resident Population 1981 and in the Construction of the 1971-81 Time Series. Document technique, Australian Bureau of Statistics.
- Begeot, F., L. Smith et D.Pearce. (1993). «First Results from Western European Censuses» dans Population Trends, 74:18-24.
- Bell, W.R. (1993). «Using Information from Demographic Analysis in Post-Enumeration Survey Estimation», dans Journal of the American Statistical Association, 88(423):1106-1118.
- Bender, R. (1992). Estimations démographiques du sous-dénombrement net du recensement de 1991. Document de travail, Division de la démographie, Statistique Canada.
- Burgess, R.D. (1988). «Evaluation of Reverse Record Check Estimates of Undercoverage in the Canadian Census of Population», dans Techniques d'enquête, 14(2):137-156.
- Castles, I. (1989). 1991 Census of Population and Housing: Content and Procedures. Australian Bureau of Statistics. Document d'information., ABS, n° 2613.0 au catalogue.
- Charles, E. (1940). The Changing Size of Family in Canada, Census Monograph No. 1, 1941.
- Choi, C.Y., D.G. Steel et T.K. Skinner (1988). «Adjusting the 1986 Australian Census Count for Underenumeration» dans Techniques d'enquête, 14(2):173-190.
- Coale, A. J. (1955). «The Population of the United States in 1950 Classified by Age, Sex and Color», Journal of the American Statistical Association, 50(1):16-54.
- Coale, A.J. et N.W. Rives (1973). «A Statistical Reconstruction of the Black Population of the United States, 1880-1970: Estimates of True Numbers by Age and Sex, Birth Rates and Total Fertility» dans Population Index, 39: 3-36.
- Coale, A.J. et M. Zelnik (1963). New Estimates of Fertility and Population in the United States: A Study of Annual White Births from 1855 to 1960 and of Completeness of Enumeration in the Censuses from 1880 to 1960, Princeton, N.J.: Princeton University Press.
- Dale, A. et C. Marsh, eds. (1993). The 1991 Census User's Guide, London: HMSO.
- Das Gupta, P. (1991). Confidence Intervals for Sex Ratios in Demographic Analysis. Note de recherche, Population Division, U.S. Bureau of the Census.



- Das Gupta, P. (1991). Demographic Analysis Evaluation Project D11: Models for Assessing Errors in Undercount Rates Based on Demographic Analysis, Preliminary Research and Evaluation Memorandum No. 84 (PREM), U.S. Bureau of the Census.
- Diamond, I. (1994). Where and Who are the Missing Million? Measuring Census of Population Undercount, Department of Social Statistics, University of Southampton.
- Dick, P. (1996). Memorandum: Comments on the paper «Alternate Methodologies in the Evaluation of Census Coverage», Division des méthodes d'enquêtes sociales.
- Dionne, C. (1995). «Les relations intergénérationnelles et quelques modèles qui en découlent» dans Revue européenne de démographie, 11:85-101.
- Dionne, C. et D. Kerr (1995). «Différentes techniques pour évaluer le dénombrement du recensement: propositions méthodologiques et applications». Document de travail, Division de la démographie, Statistique Canada.
- Dionne, C. et D. Kerr (1995b). «Utilisation du modèle intergénérationnel pour ajuster le dénombrement du recensement». Document de travail, Division de la démographie, Statistique Canada.
- Fay, R., J.S. Passel et J.G. Robinson (1988). The Coverage of Population in the 1980 Census. Evaluation and Research Report, PHC 80-E4, United States Bureau of the Census, Washington, D.C.
- Fortier, C. et R. Raby (1989). Évaluation de la qualité des données par âge et sexe du recensement de 1986 au niveau national, Cahier québécois de démographie, Vol 18 (2):285-320.
- Hall, A.R. (1969). Australian Demographic Data Bank, Department of Economics, RSSH, Australian National University.
- Heady, P., S. Smith et V. Avery (1993). 1991 Census Validation Survey: Coverage Report, Office of Population Censuses and Surveys, London: HMSO.
- Himes, C.L. et C. Clogg (1992). «An Overview of Demographic Analysis as a Method for Evaluating Census Coverage in the United States» dans Population Index, 58: 587-607.
- Hogan, H. (1993). Journal of the American Statistical Association, Vol. 88 (423): 1044-1149.
- Lapierre-Amadcyk, E. (1970). Estimation of Net Census Underenumeration by Age and Sex using Demographic Analysis Techniques. Document de travail, Division de la démographie, Statistique Canada.
- Maddan, R. (1995). Census of Population and Housing, 6 August 1991, Census 91: Data Quality - Undercount, Australian Bureau of Statistics, n° 2940.0 au catalogue.



- Michalowski, M. (1992). Estimations provisoires du sous-dénombrement net du recensement de 1991 basées sur les bases de données administratives. Document de travail, Division de la démographie, Statistique Canada.
- Michalowski, M. (1993). Révisions postcensitaires et estimations intercensitaires: Canada, provinces et territoires, 1971-1993 - Méthodologie pour la dérivation des estimations annuelles de population par âge et sexe, Section des estimations de la population, Division de la démographie, Statistique Canada.
- Mulry and Spencer (1991). «Total Error in PES Estimates of Population» dans Journal of the American Statistical Association, 86(416):839-863.
- Office of Population Censuses and Surveys (1993). «How Complete was the 1991 Census» dans Population Trends, 71 (spring):22-26.
- Office of Population Censuses and Surveys (1993b). «Rebasing the Annual Population Estimates» dans Population Trends, 71 (fall):27-32.
- Office of Population Censuses and Surveys (1995). 1991 Census, General Report, Great Britain, London: HMSO.
- Purcell, N.J. et L. Kish (1989). «Estimation for Small Domains» dans Biometrics, 35, 365-384.
- Robinson, J.G. (1991a). Demographic Analysis Evaluation Project D1: Error in the Birth Registration Completeness Estimates, Preliminary Research and Evaluation Memorandum No. 74 (PREM), U.S. Bureau of the Census.
- Robinson, J.G. (1991b). Demographic Analysis Evaluation Project D3: Uncertainty Intervals for Estimated White Births, 1915-1934, Preliminary Research and Evaluation Memorandum No. 76 (PREM), U.S. Bureau of the Census.
- Robinson, J.G. (1991c). Demographic Analysis Evaluation Project D4: Uncertainty Intervals for Estimated Black Births, 1915-1934, Preliminary Research and Evaluation Memorandum No. 77 (PREM), U.S. Bureau of the Census.
- Robinson, J.G. (1991d). Demographic Analysis Evaluation Project D6: Robustness of Estimates of the Population Aged 65 and Over, Preliminary Research and Evaluation Memorandum No. 79 (PREM), U.S. Bureau of the Census.
- Robinson, J.G. et E.L. Kobilarcik (1995). Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach, travail présenté au «Annual Meetings of the Population Association of America», 6-8 avril 1995.
- Robinson, J.G. et S. Lapham (1991). Demographic Analysis Evaluation Project D9: Inconsistencies in Racial Classifications of the Demographic Estimates and the Census, Preliminary Research and Evaluation Memorandum No. 82 (PREM), U.S. Bureau of the Census.



- Robinson, J.G., K.A. Woodrow et B. Ahmed (1991). Demographic Analysis Evaluation Project D7: Uncertainty Measure for Other Components, Preliminary Research and Evaluation Memorandum No. 80 (PREM), U.S. Bureau of the Census.
- Robinson, J.G., D.L. Ward et G.S. Spencer (1991). Demographic Analysis Evaluation Project D8: Uncertainty for Models to Translate 1990 Census Concepts into Historical Racial Classification, Preliminary Research and Evaluation Memorandum No. 81 (PREM), U.S. Bureau of the Census.
- Romaniuc, A. (1988) «Une approche démographique à l'évaluation du recensement de 1986 et des estimations de population pour le Canada» dans Techniques d'enquête, 14(2):157-172.
- Siegel, J.S. (1974). «Estimates of Coverage of the Population by Sex, Race, and Age in the 1970 Census» dans Demography, 11:1-24.
- Siegel, J.S. et M. Zelnik (1966). «An Evaluation of Coverage in the 1960 Census of Population by Techniques of Demographic Analysis and by Composite Methods» dans Proceedings of the Social Statistics Section, American Statistical Association, pp. 71-85.
- Statistique Canada (1993). Rapports techniques du recensement de 1991, n° 92-341F au catalogue.
- Statistique Canada (1994). Estimations intercensitaires révisées de la population et des familles au 1^{er} juillet, 1971-1991, n° 91-537 au catalogue.
- U.S. Bureau of the Census (1996). The Plan for Census 2000, U.S. Department of Commerce.
- Whelpton (1950). «Birth and Birth Rates in the Entire United States, 1909-1948» dans Vital Statistics Special Reports, 33, 137-162.
- Wiggins, R.D. (1993). «The Validation of Census Data, Post-Enumeration Survey Approaches» dans A. Dale et C. Marsh (eds), The 1991 Census User's Guide, London: HMSO.
- Woodrow, K.A. (1991a). Demographic Analysis Evaluation Project D2: Preliminary Estimates of Undocumented Residents in 1990, Preliminary Research and Evaluation Memorandum No. 75 (PREM), U.S. Bureau of the Census.
- Woodrow, K.A. (1991b). Demographic Analysis Evaluation Project D5: Preliminary Estimates of Emigration Component, Preliminary Research and Evaluation Memorandum No. 78 (PREM), U.S. Bureau of the Census.

