

Estimation de la répartition par âge et par sexe de la population totale des petites régions¹

DAVID S. O'NEIL et CHRIS D. McINTOSH²

RÉSUMÉ

Ce document décrit une méthode de production d'estimations par âge et par sexe de l'état de la population des petites régions, à partir d'estimations de la population totale, de données sur les naissances et les décès et d'estimations chronologiques de la migration nette résiduelle. On y présente également une évaluation fondée sur les chiffres du recensement de 1981 concernant les divisions de recensement et les districts scolaires de la Colombie-Britannique.

MOTS CLÉS: Estimations de la population par âge et par sexe; petites régions; migration nette résiduelle.

1. INTRODUCTION

Le Bureau central de la statistique produit actuellement des estimations postcensitaires de la population pour diverses régions infraprovinciales à l'aide d'une méthode de régression (Bureau central de la statistique, 1982). En plus des estimations de la population totale par petite région, il produit des estimations par âge et par sexe.

Ce document expose la méthode permettant à partir de l'estimation de la population totale, d'obtenir des estimations par âge et par sexe pour les régions infraprovinciales de la Colombie-Britannique.

2. VUE D'ENSEMBLE

La méthode retenue pour déterminer la population des petites régions par sexe et par année d'âge comporte deux volets.

Le premier volet consiste à examiner les données chronologiques sur la migration nette résiduelle à partir des chiffres de recensement afin d'établir un certain nombre de répartitions de la migration par sexe et par année d'âge pour chaque petite région (Shryock et Siegal 1980).

Le second volet de la méthode consiste à faire vieillir la population de base pour chaque sexe, à ajouter les naissances et à retrancher les décès, ce qui donne pour chaque région une nouvelle répartition de la population que l'on désigne la population de base naturelle. En retranchant la population de base naturelle de la population totale estimée selon le sexe, on obtient un résidu qui est égal à la migration nette selon le sexe si la population dénombrée et les données de l'état civil sont exacts pour les deux périodes. Ce résidu par sexe est distribué par année d'âge au moyen d'une répartition chronologique, puis il est ajouté à la population de base naturelle, ce qui donne l'estimation de la population par âge et par sexe d'une région pour la période suivante.

À cause des courts délais de production des données d'entrée, les estimations de la population totale peuvent être calculées quatre mois après la date de référence (1^{er} juin) et les répartitions par âge et par sexe, un ou deux mois plus tard.

¹ Version abrégée du document présenté à la réunion du Comité fédéral-provincial sur la démographie, Ottawa, novembre 28-29, 1985.

² D.S. O'Neil, SRL Sociometrics Resources Ltd., et C.M. McIntosh, Intersoft Resources Ltd., Bureau central de la statistique, ministère du développement de l'industrie et de la petite entreprise de la Colombie-Britannique, 2^e étage, 1405, Douglas Street, Victoria, Colombie-Britannique, Canada, V8W 3C1.
Les opinions exprimées dans ce document sont celles des auteurs et ne représentent pas nécessairement le point de vue du gouvernement de la Colombie-Britannique.

3. RÉPARTITION CHRONOLOGIQUES DE LA MIGRATION NETTE

Pour chacun des 74 districts scolaires de la Colombie-Britannique, les estimations de la migration nette par âge et par sexe ont été calculées, par la méthode des résidus, pour les périodes de 1961-1966, 1966-1971 et 1971-1976. C'est ce qu'on appelle les répartitions chronologiques par petite région.

Si l'on examine ces répartitions de la migration nette par petite région, on découvre qu'elles sont extrêmement instables dans le temps. Pour réduire les effets de cette instabilité, un certain nombre de mesures ont été prises.

Premièrement, les répartitions de la migration par petite région ont été classées selon qu'elles ont eu lieu au cours d'une période de migration totale nette positive ou d'une période de migration totale nette négative. On a découvert que les répartitions par âge de la migration résiduelle pour de nombreuses régions diffèrent selon que la migration est positive ou négative.

On a ensuite regroupé les petites régions présentant des répartitions de migration semblables, pour ensuite calculer les répartitions de la migration nette positive et de la migration nette négative. Elles ont été désignées répartitions chronologiques par groupe. L'analyse de certains groupes d'âge (selon la méthode SPSS/PC) a permis de regrouper les répartitions chronologiques de la migration des petites régions. Cette méthode a permis de grouper la plupart des régions en trois grappes, et huit groupes étaient formés uniquement d'une région chacun.

Une fois les régions ainsi regroupées, on a déterminé les répartitions de la migration positive et les répartitions de la migration négative pour les périodes de migration positive et négative les plus récentes.

4. ESTIMATIONS DE LA POPULATION DES PETITES RÉGIONS SELON LE SEXE ET L'ANNÉE D'ÂGE

Comme il a été mentionné dans la section 3, les répartitions chronologiques de la migration nette calculées par la méthode des résidus varient considérablement pour certaines régions. Il semble que cela soit imputable à deux facteurs. Premièrement, de nombreuses régions à l'étude ont des économies de ressources déficientes qui affichent de grandes fluctuations, ce qui entraîne des mouvements des taux de migration. Deuxièmement, le calcul d'une distribution en percentiles d'un paramètre tel que la migration nette ayant des valeurs positives, négatives ou nulles introduit un certain degré d'instabilité.

Pour prévenir la construction d'une répartition chronologique de la migration nette qui ne soit pas représentative de la situation courante dans l'année d'estimation, on a calculé cinq répartitions chronologiques, par sexe, qui ont été ventilées par année d'âge. Une description de ces répartitions est donnée ci-après.

- 1) La première répartition de la migration qui a été choisie est la répartition chronologique par petite région de chaque petite région dont le signe est le même que la migration nette vers cette région.
- 2) La deuxième répartition est la répartition chronologique par groupe du groupe auquel la petite région appartient et dont le signe est le même que celui de la migration nette vers cette région.
- 3) On a obtenu la troisième répartition en additionnant *séparément* la migration, pour la période la plus récente, de toutes les petites régions possédant une migration nette positive et une migration nette négative, puis en calculant la répartition par âge.
- 4) La quatrième répartition est celle de la population de base naturelle pour chaque petite région.
- 5) La cinquième et dernière répartition est la répartition par âge des migrants vers la Colombie-Britannique en général. Comme la migration vers la Colombie-Britannique a été positive au cours de toutes les années visées, cette répartition est positive. On l'a néanmoins utilisée, peu importe que la migration vers une petite région ait été positive ou négative.

Il n'est pas toujours possible de calculer toutes les cinq répartitions. Ainsi, on ne peut le faire si une petite région n'a jamais eu de migration nette négative dans le passé mais si elle en a une pour l'année visée. Dans un tel cas, on se sert uniquement des répartitions que l'on peut calculer.

Des essais basés sur les données du recensement de 1981 ont montré que, des cinq répartitions de la migration nette décrites plus haut, la première (la répartition chronologique par petite région) a l'erreur absolue moyenne en pourcentage la plus faible pour tous les districts scolaires et tous les groupes d'âge; elle est suivie de la deuxième (répartition chronologique par groupe), puis de la troisième, etc. Cependant, bien que la première répartition ait l'EAMP la plus faible, elle n'a pas produit le taux d'erreur le plus bas dans chaque cas. Aussi, une technique de sélection a été appliquée pour remplacer la répartition de la population produite au numéro 1 par celle produite aux numéros 2, 3, 4 ou 5 uniquement dans les cas où la répartition de la population produite au numéro 1 était considérée comme n'étant pas fidèle à la répartition de la population pour l'année choisie. La technique de sélection suivante a été conçue à partir des résultats des tests faites sur le recensement de 1981.

Une fois qu'on a calculé toutes les répartitions possibles, on ajoute chacune à la population de base naturelle et on obtient jusqu'à cinq estimations possibles de la population d'une petite région, selon le sexe et l'année d'âge, pour la période suivante. On examine alors ces estimations de la population par âge et par sexe afin de déterminer celle qui produit le moins de changements dans la structure par âge de la petite région par rapport à l'année précédente. Pour ce faire, on calcule d'abord la différence moyenne en pourcentage entre les structures par âge de chacune des cinq populations possibles au temps $t + 1$ et celles de la population au temps t . On calcule ensuite l'écart-type des moyennes obtenues et on indique au moyen d'un code la répartition qui affiche l'écart-type le plus bas. Si l'écart-type produit à partir de la répartition chronologique par petite région est beaucoup plus grand que l'écart-type minimum (c.-à-d. la répartition signalée), cette répartition chronologique est rejetée. On applique le même procédé à la répartition chronologique par groupe, et ainsi de suite jusqu'à ce qu'une des cinq populations possibles soit retenue.

Une fois la population optimale au temps $t + 1$ calculée pour toutes les petites régions, il reste à faire deux derniers rajustements. En premier lieu, on a remplacé les chiffres relatifs aux groupes d'âge 0-14 ans par ceux provenant des allocations familiales, puis ajusté proportionnellement les populations des autres groupes d'âge afin que la population totale de chaque petite région reste constante. En second lieu, on a ajusté proportionnellement la population pour s'assurer que la répartition par âge de la somme de toutes les estimations relatives aux populations des petites régions corresponde à la répartition par âge établie par Statistique Canada pour la Colombie-Britannique.

5. ÉVALUATION DE LA MÉTHODE

Les tableaux présentés plus loin indiquent les taux d'erreur relatifs aux estimations de la population au 1^{er} juin 1981 et répartis en groupes d'âge quinquennaux de 0 à 70 ans et plus, pour 74 districts scolaires de la Colombie-Britannique et 29 divisions de recensement. On a obtenu les estimations de la population par âge et par sexe pour les divisions de recensement en regroupant les estimations de la population pour les districts scolaires.

Pour établir l'exactitude des estimations de la population des petites régions par âge et par sexe obtenues par la méthode décrite plus haut, on a produit des estimations de la population de 1981 par sexe et par groupe d'âge quinquennal de 0 à 70 ans et plus pour 74 districts scolaires, puis on a comparé les résultats obtenus avec les chiffres du recensement de 1981. On a eu recours à deux mesures sommaires pour évaluer l'efficacité des estimations de la population par âge et par sexe. Il s'agit de l'erreur absolue moyenne en pourcentage (EAMP) et l'indice de mauvaise distribution (IMD). L'EAMP est défini comme suit:

$$EAMP = 100 \times \left[\sum_{i=1}^N \left| (P_{Ei} - P_{Ai}) / P_{Ai} \right| \right] / N$$

où P_{Ei} est la population estimée pour le groupe d'âge i , P_{Ai} est la population d'après le recensement pour le groupe d'âge i et N est le nombre de cellules. L'IMD est défini comme suit:

$$IMD = 100 \times \frac{1}{2} \left[\frac{\sum_{i=1}^N (|P_{Ai} - P_{Ei}|)}{\sum_{i=1}^N P_{Ai}} \right]$$

où P_{Ai} est la population réelle pour le groupe d'âge i et P_{Ei} est la population estimée pour le groupe d'âge i .

Comme l'indique le tableau 1, l'erreur absolue moyenne en pourcentage pour tous les groupes d'âge et toutes les régions est de 6.20% et l'IMD, de 1.95%. L'erreur moyenne en pourcentage pour les hommes est très semblable à celle pour les femmes (EAMP de 7% dans les deux cas et IMD de 2.15% pour les hommes et de 2.08% pour les femmes).

Du point de vue de l'âge, les taux d'erreur les plus élevés figurent dans les groupes des 20-29 ans et des 60-69 ans. Il faut également noter que la répartition par âge des erreurs est différente chez les hommes et chez les femmes. Chez les hommes, les taux d'erreur les plus élevés semblent se trouver dans les groupes d'âge supérieurs, tandis que chez les femmes ils sont dans les groupes très mobiles des 20-29 ans.

Tableau 1
Pourcentage d'erreurs entre les estimations et les chiffres du recensement de 1981, selon le groupe d'âge, pour tous les districts scolaires

Âge	Total		Hommes		Femmes	
	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)
0-4	3.33	.96	3.94	1.21	3.62	1.04
5-9	2.80	.76	3.28	0.88	3.62	1.02
10-14	2.33	.64	3.54	0.84	2.88	0.87
15-19	5.20	2.01	5.68	2.01	6.18	2.24
20-24	13.32	4.77	13.50	4.62	14.54	5.12
25-29	8.31	4.07	8.42	3.70	9.41	4.65
30-34	5.02	2.12	5.42	2.45	5.72	2.06
35-39	4.88	1.33	5.73	1.62	5.38	1.34
40-44	4.52	1.33	5.84	1.51	4.67	1.52
45-49	3.60	1.22	4.47	1.37	4.78	1.49
50-54	5.66	1.33	5.86	1.48	6.68	1.54
55-59	6.11	1.72	6.19	1.78	7.82	1.97
60-64	8.86	2.44	10.35	2.95	8.91	2.17
65-69	10.60	2.66	12.53	3.52	11.44	2.30
70+	8.49	1.95	10.19	2.35	9.33	1.94
Moyenne	6.20	1.95	7.00	2.15	7.00	2.08

Comme l'indique le tableau 2, les régions peu peuplées ont en général un pourcentage d'erreur plus élevé. Le pourcentage élevé d'erreurs dans les petites régions peut être attribué à l'instabilité des économies de ressources, laquelle se reflète dans la répartition de la migration nette.

Au niveau des divisions de recensement, on constate qu'il se dégage une tendance semblable. Comme le montre le tableau 3, l'erreur absolue moyenne en pourcentage pour toutes les régions et tous les groupes d'âge est de 4.83%, (5.19% pour les hommes et 5.60% pour les femmes). L'IMD est de 1.27% au total, de 1.41% pour les hommes et de 1.35% pour les femmes. L'erreur est encore une fois bimodale, atteignant des sommets dans les groupes des 20-29 ans et des 60-69 ans. En outre, les femmes affichent des taux d'erreur plus élevés que les hommes dans les groupes des 20-29 ans, et des taux d'erreur plus bas dans les groupes des 60-69 ans.

Tableau 2
Pourcentage d'erreurs pour tous les districts scolaires
selon la taille de la population

Taille de population	Hommes		Femmes		Total	
	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)
0-9,999	8.87	3.16	10.14	3.89	10.27	3.65
10,000-24,999	6.07	2.47	6.92	2.96	6.62	2.58
25,000 +	3.66	1.67	3.92	1.78	4.09	1.78
Moyenne pour les districts scolaires	6.20	1.95	7.00	2.15	7.00	2.08

Tableau 3
Pourcentage d'erreurs entre les estimations et les chiffres du recensement de 1981, selon le groupe d'âge, pour toutes les divisions de recensement

Groupe d'âge	Total		Hommes		Femmes	
	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)
0-4	2.37	0.54	3.20	0.76	2.28	0.58
5-9	1.52	0.50	1.71	0.55	2.13	0.68
10-14	1.69	0.39	2.75	0.57	2.50	0.60
15-19	3.81	1.39	3.79	1.30	4.68	1.63
20-24	9.83	3.07	9.30	2.91	10.90	3.41
25-29	7.02	3.04	7.30	2.87	8.09	3.37
30-34	3.28	1.29	3.31	1.43	3.85	1.25
35-39	3.34	0.66	3.06	0.57	4.21	0.88
40-44	3.86	0.88	4.29	1.01	4.16	0.90
45-49	2.91	0.70	3.20	0.75	3.75	0.83
50-54	4.82	0.64	4.41	0.75	6.10	0.86
55-59	5.49	1.34	5.36	1.55	6.94	1.30
60-64	7.88	1.95	8.37	2.29	7.94	1.74
65-69	8.48	1.89	10.30	2.67	9.79	1.43
70 +	6.16	0.81	7.46	1.20	6.73	0.71
Moyenne	4.83	1.27	5.19	1.41	5.60	1.35

Le tableau 4 (pourcentage d'erreurs dans les divisions de recensement selon la taille de la population) indique qu'à la suite de l'agrégation au niveau de régions infraprovinciales plus grandes, on obtient une amélioration des taux d'erreur. Le tableau 7 montre qu'il existe une relation négative entre les taux d'erreur et la taille de la population, à l'échelon de la division de recensement.

La comparaison des tableaux 5 et 6 confirme l'amélioration des taux d'erreurs lorsqu'on agrège en fonction de cellules par âge et sexe plus grandes. Bien qu'il faille être prudent dans l'utilisation d'estimations régionales par âge et sexe, les auteurs estiment qu'elles sont fiables.

Tableau 4
Pourcentage d'erreurs pour toutes les divisions de recensement
selon la taille de la population

Taille de population	Total		Hommes		Femmes		N
	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	
0-39,000	7.22	1.94	7.55	2.13	8.79	2.29	10
40,000-59,999	4.32	1.82	5.03	2.14	4.91	1.83	10
60,000+	2.51	0.87	2.73	0.98	2.84	0.90	9
Moyenne des divisions de recensement	4.83	1.27	5.19	1.41	5.60	1.35	29

Tableau 5
Districts scolaires
Nombre d'estimations par intervalle d'erreur

	Intervalle d'erreur absolue moyenne en pourcentage				Total
	< 5	5 à 10	10 à 15	15 +	
N° de cellules	674	239	101	96	1110
Pourcentage	61%	22%	9%	9%	100%

Tableau 6
Divisions de recensement
Nombre d'estimations par intervalle d'erreur

	Intervalle d'erreur absolue moyenne en pourcentage				Total
	< 5	5 à 10	10 à 15	15 +	
N° de cellules	306	77	25	27	435
Pourcentage	70%	18%	6%	6%	100%

6. CONCLUSIONS

La méthode décrite dans ce document présente des avantages dans le cas des régions pour lesquelles il existe de bonnes sources de données chronologiques sur la population et de statistiques de l'état civil. On estime que l'application d'une méthode comportant des estimations de la migration nette est relativement simple, qu'elle produit des taux d'erreur acceptables et permet la production d'estimations par âge et sexe rapidement après la date de référence. Certes, il serait idéal d'établir des estimations de la migration interne et externe, mais on ne dispose pas actuellement d'assez d'information sur les mouvements migratoires dans les petites régions de la Colombie-Britannique. Une autre amélioration qui est envisagée a trait à l'utilisation des chiffres relatifs à la sécurité de la vieillesse pour accroître la stabilité et la précision des estimations concernant les groupes d'âge supérieurs.

REMERCIEMENTS

Les auteurs tiennent à remercier Don McRae, Steve Miller, Ravi Verma, Garnett Picot et Paul Knapp ainsi que tous leurs prédécesseurs pour leur apport et leur soutien à la construction du système de ventilation des estimations.

Tableau 7

Pourcentage d'erreurs entre les estimations et les chiffres du recensement de 1981, selon la division de recensement, pour tous les groupes d'âge

Division du recensement	Population totale	Total		Hommes		Femmes	
		EAMP (%)	IMD (%)	EAMP (%)	IMD (%)	EAMP (%)	IMD (%)
1000 East Kootenay	53,725	4.24	2.04	5.24	2.29	3.88	2.15
3000 Central Kootenay	52,045	4.00	2.18	4.03	2.13	5.06	1.69
5000 Kootenay-Boundary	33,235	2.32	1.23	2.34	1.18	3.21	1.68
7000 Okanagan-Similkameen	57,185	5.04	2.64	6.02	3.08	4.72	2.49
9000 Fraser-Cheem	56,930	3.12	1.60	3.33	1.78	4.15	2.08
11000 Central Fraser Valley	115,015	3.14	1.43	3.46	1.52	3.65	1.81
13000 Dowdney-Alouette	62,000	2.10	1.15	2.56	1.23	2.26	1.32
15000 Greater Vancouver	1,168,700	1.63	0.94	1.68	0.93	1.67	0.98
17000 Capital	249,475	1.64	0.87	2.31	1.21	1.18	0.61
19000 Cowichan Valley	45,315	3.09	1.66	3.36	1.69	3.85	2.08
21000 Nanaimo	84,815	3.07	1.58	3.40	1.74	3.22	1.66
23000 Alberni-Clayoquot	32,560	2.75	1.36	2.88	1.27	3.27	1.68
25000 Comox-Strathcona	68,620	1.44	0.80	1.85	0.87	2.85	1.50
27000 Powell River	19,050	5.36	2.58	5.06	2.44	6.18	3.03
29000 Sunshine Coast	16,625	4.84	2.57	6.79	3.58	5.65	2.81
31000 Squamish-Lillooet	18,925	1.82	0.99	2.56	1.37	3.10	1.58
33000 Thompson-Nicola	102,430	2.13	1.10	2.07	0.10	2.65	1.37
35000 Central Okanagan	85,235	3.96	1.93	3.91	1.88	4.32	2.14
37000 North Okanagan	69,033	5.26	2.52	6.44	3.06	5.05	2.50
39000 Columbia-Shuswap	45,425	3.04	1.63	3.56	1.84	2.99	1.66
41000 Cariboo	58,810	3.18	1.93	3.90	2.18	3.42	2.06
43000 Mount Waddington	14,675	8.96	3.04	5.13	1.59	17.77	5.49
45000 Central Coast	3,050	17.99	7.62	21.62	8.86	14.92	7.34
47000 Skeena-Queen Charlotte	24,030	4.82	2.09	5.70	2.58	4.61	1.84
49000 Kitimat-Stikina	41,790	6.26	1.99	4.99	1.66	8.59	2.78
51000 Bulkley-Nechako	38,310	6.23	2.31	5.76	2.10	6.83	2.57
53000 Fraser-Fort George	89,430	3.50	1.41	3.39	1.25	3.72	1.68
55000 Peace River-Liard	55,340	8.00	2.95	9.43	3.65	7.34	2.83
57000 Stikine	2,685	17.15	6.89	17.89	6.88	22.39	8.35
Erreur moyenne		4.83	2.17	5.19	2.31	5.60	2.51

BIBLIOGRAPHIE

- BUREAU CENTRAL DE LA STATISTIQUE (1982). British Columbia Municipal Population Estimation Methodology. Document non publié. Ministère du Développement de l'industrie et de la petite entreprise, Victoria.
- SPSS INC. (1984). *Statistical Package for the Social Sciences/PC*. Chicago, B265-B280.
- SHRYOCK, H.S., et SIEGAL, J.S. (1980). *The Methods and Materials of Demography*. 2, U.S. Bureau of the Census, 628-630.