Techniques d'enquête, décembre 1988
Vol. 14, n° 2, pp. 187–204
Statistique Canada

Redressement des chiffres du recensement de 1986 en Australie pour le sous-dénombrement

C.Y. CHOI, D.G. STEEL et T.J. SKINNER[1]

RÉSUMÉ

En Australie, les estimations démographiques sont établies à partir des chiffres du recensement; lors des trois derniers recensements (1976, 1981 et 1986), ces chiffres ont été redressés pour tenir compte du sous-dénombrement. L'opération de redressement s'inspire des résultats d'une enquête postcensitaire et d'une analyse démographique. Cet article expose les méthodes utilisées de même que les résultats du redressement des données de 1986. Les auteurs voient aussi dans l'usage formel des rapports de masculinité proposé par Wolter (1986) une amélioration par rapport au rôle qu'avaient auparavant ces ratios dans le redressement des chiffres du recensement.

MOTS CLÉS: Sous-dénombrement dans le recensement; enquête postcensitaire; estimations démographiques; rapports de masculinité.

1. INTRODUCTION

Le recensement de la population fournit les données de base qui permettent d'établir des estimations de la population pour le pays en général, chacun des huit États et les districts d'administrations locales. En Australie, ces estimations démographiques sont nécessaires pour déterminer le nombre de sièges dont disposera chaque État à la Chambre des représentants du parlement fédéral, le mode de répartition des fonds fédéraux entre les États et le niveau de financement des administrations locales. Les estimations démographiques servent aussi, comme il se doit, d'indices de la croissance et de la répartition démographiques et de dénominateurs dans divers indicateurs démographiques, sociaux et économiques. Compte tenu de l'importance de ces usages, les estimations démographiques doivent être d'un haut degré de précision.

En Australie, on reconnaît que le niveau de sous-dénombrement dans le recensement est élevé et qu'il a trait à d'importantes variables telles que le lieu de naissance, la région géographique et l'âge-sexe. C'est pourquoi les chiffres du recensement qui servent à l'établissement d'estimations démographiques sont redressés en fonction du sous-dénombrement.

Le redressement des chiffres du recensement en fonction du sous-dénombrement est une pratique récente en Australie. Avant le recensement de 1976, on se servait de chiffres non redressés pour estimer la population. On s'est rendu compte de la nécessité de redresser les chiffres du recensement lorsqu'on a constaté que les chiffres de 1976 étaient bien en-deçà des estimations démographiques qui avaient été établies pour 1976 par suite d'une mise à jour des données du recensement de 1971, et que l'enquête postcensitaire (EP) de 1976 révélait un taux de sous-dénombrement de 2.6%, comparativement à 0.5% en 1966 et à 1.3% en 1971. L'EP de 1976 révélait aussi des différences de taux de sous-dénombrement notables entre les États et les Territoires, les taux allant de 4.2%, pour le Territoire du Nord, à 1.1%, pour la Tasmanie. En 1986, le taux de sous-dénombrement a été estimé à 1.9%. Comme en 1976, on a relevé des différences notables entre les États et les Territoires. Le redressement des chiffres du

[1] C.Y. Choi, D.G. Steel et T.J. Skinner, Autralian Bureau of Statistics, P.O. Box 10, Belconnen, ACT, 2616, Australie.

recensement de 1976 et des recensements subséquents n'a pas soulevé d'opposition et en aucun temps on n'a mis en doute le bien fondé d'une telle opération ou la précision des méthodes utilisées. C'est tout le contraire de ce qui a été vécu aux États-Unis lorsqu'on a voulu redresser les chiffres du recensement de 1980 pour tenir compte du sous-dénombrement; cette opération avait alors soulevé une grande controverse.

Pour évaluer le niveau de sous-dénombrement, on se sert surtout des données d'une EP. On évalue les résultats d'une telle enquête en les comparant à des estimations établies à partir de données démographiques et d'autres données indépendantes comme celles concernant les inscriptions scolaires, les enfants dont les parents reçoivent des allocations familiales et les personnes inscrites au régime d'assurance-maladie de l'État. En Australie, les enfants de 6 à 15 ans sont obligés de fréquenter l'école et toutes les mères qui ont un enfant de moins de 17 ans reçoivent des allocations familiales; du moins, il en était ainsi avant que l'on mette en application le test de moyennes, en novembre 1987. Enfin, le régime d'assurance-maladie est obligatoire et universel. Ces sources indépendantes de données sont donc utiles pour la vérification des résultats de l'EP et des estimations démographiques.

Les estimations de la population sont les seules données du recensement qui sont redressées pour tenir compte du sous-dénombrement. Les chiffres du recensement sont publiés sans avoir été redressés.

2. L'ENQUÊTE POSTCENSITAIRE DE 1986

Pour son recensement quinquennal de la population, l'Australian Bureau of Statistics (ABS) emploie des agents recenseurs pour livrer les questionnaires à chaque ménage et les ramasser une fois qu'ils sont remplis. Les personnes sont recensées en fonction de l'endroit où elles se trouvent le soir de la journée du recensement.

Grâce à ce réseau d'agents recenseurs, l'étape de la collecte des données ne s'étend pas au-delà de deux semaines suivant la date du recensement. On peut ainsi réaliser une EP quelque temps seulement après la date du recensement - en 1986, moins de 4 ou 5 semaines après le jour du recensement. Comme l'EP renferme un certain nombre de questions qui exigent des réponses détaillées sur l'endroit où se trouvait une personne le soir du recensement, le fait qu'elle soit réalisée peu de temps après la date du recensement réduit au minimum les erreurs de mémoire et amoindrit aussi le nombre d'exclusions attribuables à des décès ou à des voyages à l'étranger.

Puisque l'EP produit les données qui servent à redresser les chiffres du recensement pour tenir compte du sous-dénombrement, il est essentiel que cette enquête soit statistiquement indépendante du recensement. Dans l'annexe nous décrivons comment on assure l'indépendance des deux enquêtes.

Dans l'EP de 1986, la méthode utilisée consistait essentiellement à prélever un échantillon de personnes indépendamment du recensement par un échantillonnage aréolaire à plusieurs degrés de logements privés. Des membres de l'équipe régulière d'intervieweurs du ABS ont été envoyés dans les ménages échantillonnés pour recueillir les données pertinentes sur chaque membre du ménage; sur place, les renseignements devaient être fournis par une personne adulte responsable. Dans une troisième étape, le personnel du Centre de transcription des données du recensement a tenté d'apparier les enregistrements du recensement et ceux de l'EP pour déterminer si chaque personne de l'échantillon devait avoir été recensée et combien de fois son nom figurait, le cas échéant, sur la liste des personnes recensées. Les méthodes utilisées sont décrites dans l'annexe.

Les résultats de l'EP permettent d'estimer le rapport entre le nombre de personnes qui auraient dû être recensées (x) et le nombre de personnes qui, selon les estimations, l'ont été réellement (y). Ce rapport est le facteur de redressement net, qui tient compte à la fois du surdénombrement et du sous-dénombrement des personnes.

Après pondération, on multiplie ce facteur de redressement par le chiffre du recensement (Y) pour obtenir une estimation de la population (X), c'est-à-dire $X = Y\ (x/y)$.

Pour tenir compte des écarts entre le rendement prévu et le rendement réel de l'échantillon dans l'EP, on a appliqué cette méthode en fonction de l'âge (groupes de cinq années), du sexe et de la région géographique (capitale/reste de l'État). Les estimations de l'EP sont produites aussi bien selon le lieu de résidence le jour du recensement que selon le lieu de résidence habituel. Le processus d'estimation comporte également un facteur de redressement pour les cas de non-contact et de non-réponse dans l'EP, qui sont peu nombreux. Par exemple, l'estimation de la population (selon le lieu de résidence habituel) pour la région géographique (s) et la cellule âge-sexe (a) est:

$$X_{sa} = Y_{sa}\, x_{sa}/y_{sa}$$

où

$$x_{sa} = \sum_{gc} \frac{D_{gc} + d_{gc}}{D_{gc}} \cdot \frac{x_{sagc}}{f_g}$$

et

$$y_{sa} = \sum_{gc} \frac{D_{gc} + d_{gc}}{D_{gc}} \cdot \frac{y_{sagc}}{f_g} \ .$$

Dans ces formules d'estimation, l'indice c désigne le code de dénombrement du ménage de l'EP pour le recensement et l'indice g désigne la région géographique où a été échantillonnée la personne pour les besoins de l'EP. D_{gc} représente le nombre de ménages répondants et d_{gc} est le nombre de ménages de la région g et de la classe de dénombrement (c) qui n'ont pu être contactés ou n'ont pas répondu à l'enquête. Le taux de sondage varie d'un État à l'autre et est désigné par f_g.

Défini de cette façon, l'estimateur est un estimateur par quotient stratifié a posteriori. Si nous faisons abstraction pour l'instant du fait que des personnes peuvent être recensées incorrectement ou plus d'une fois, l'estimateur en question est celui que l'on obtient par un système dual ou une méthode de saisie-resaisie dont font état notamment Bishop, Fienberg et Holland (1975, p. 231-234). C'est ce qu'illustre le diagramme ci-dessous, où, en vertu de l'hypothèse d'indépendance, l'estimation de la population totale est $Y\ (x/y)$ qui correspond à l'estimation par quotient X.

EP

		Inclus	Exclus
Recensement	Inclus	y	Y
	Exclus		
		x	

Or, l'EP de 1986 visait à faire connaître le nombre de personnes qui avaient été oubliées dans le recensement et le nombre de personnes qui avaient été recensées en trop, c'est-à-dire celles qui avaient été recensées par erreur ou plus d'une fois. La valeur estimée X tient compte

à la fois des cas de surdénombrement et des cas de sous-dénombrement. À cet égard, la méthode utilisée diffère de la méthode de saisie-resaisie classique.

Pour estimer la variance, nous avons considéré X omme une estimation par quotient tirée d'un échantillon à plusieurs degrés. Les erreurs types relatives des estimations de l'EP figurent dans le tableau 1. D'après ce tableau et les tableaux 2 et 4, nous pouvons voir que les erreurs types sont beaucoup moins élevées que les corrections découlant des estimations nationales de l'EP selon l'âge et le sexe et selon l'État et le sexe.

Pour une description plus détaillée de l'enquête postcensitaire de 1986 et des méthodes d'estimation, voir l'annexe.

Tableau 1

Recensement de 1986: Erreurs types relatives des estimations démographiques de l'EP

Age	Hommes	Femmes	Total
	%	%	%
0- 4	0.29	0.36	0.24
5- 9	0.29	0.30	0.22
10-14	0.28	0.29	0.21
15-19	0.32	0.32	0.24
20-24	0.49	0.43	0.34
25-29	0.49	0.36	0.32
30-34	0.39	0.34	0.27
35-39	0.36	0.30	0.24
40-44	0.38	0.32	0.26
45-49	0.37	0.30	0.25
50-54	0.43	0.38	0.30
55-59	0.38	0.30	0.25
60-64	0.41	0.38	0.29
65-69	0.43	0.37	0.29
70-74	0.53	0.41	0.34
75+	0.47	0.39	0.31
Tous les âges	0.12	0.10	0.08

État	Hommes	Femmes	Total
	%	%	%
Nouvelles-Galles-du-Sud	0.21	0.18	0.14
Victoria	0.23	0.21	0.16
Queensland	0.27	0.24	0.19
Australie méridionale	0.27	0.20	0.17
Australie occidentale	0.29	0.25	0.19
Tasmanie	0.36	0.31	0.25
Territoire du Nord	1.65	1.53	1.22
Territoire fédéral de la capitale	0.61	0.74	0.55

3. ESTIMATION DU SOUS-DÉNOMBREMENT DANS LE RECENSEMENT À L'AIDE DE DONNÉES DÉMOGRAPHIQUES

On peut aussi estimer le taux de sous-dénombrement dans le recensement au moyen de données démographiques comme celles tirées de recensements antérieurs et des registres des naissances et des décès de même que les données sur la migration externe. Par exemple, on peut estimer la population à une date précise en mettant à jour les données d'un recensement antérieur à l'aide des données sur les naissances, les décès et la migration externe. Plus le recensement en question est éloigné dans le temps, plus la série chronologique des données de l'état civil et des données sur la migration nécessaires est étendue et moins on a à se fier sur la justesse des données du recensement de référence, car le nombre estimé de personnes nées après la date du recensement pertinent ne tient qu'à la fiabilité des données sur les naissances, les décès et la migration. Les données sur la migration interne en Australie ne sont pas suffisamment fiables pour que l'on puisse utiliser des méthodes démographiques dans l'estimation de taux de sous-dénombrement au niveau infra-national. Par conséquent, seuls les totaux nationaux peuvent servir à évaluer la qualité d'un recensement.

Les données sur les naissances et les décès en Australie existent sous forme de séries chronologiques qui remontent jusqu'au 19^{e} siècle et ces séries sont vraisemblablement complètes. Dans chacun de ses rapports publiés de 1911 à 1961 après chaque recensement de la population, le statisticien du Commonwealth australien affirmait que la très grande majorité des naissances et des décès avaient été consignés mais reconnaissait qu'il pouvait y avoir eu quelques omissions et qu'il y avait des retards dans les inscriptions. Le rapport du statisticien n'est plus publié depuis 1961. Cependant, il n'y a rien qui indique que les naissances et les décès sont enregistrés avec moins de régularité qu'auparavant.

L'Australie tient aussi depuis longtemps des statistiques complètes et fiables sur les courants migratoires. Ces statistiques portent sur tous les mouvements migratoires, qu'il s'agisse des mouvements permanents ou des mouvements à court terme et à long terme. Cependant, elles présentent plusieurs lacunes qui limitent leur utilité en ce qui a trait à l'évaluation de la qualité des recensements. Premièrement, il y a des périodes dans l'histoire de l'Australie où, croit-on, les courants migratoires n'auraient pas été enregistrés correctement (par exemple, durant la Seconde Guerre mondiale et la période qui a suivi immédiatement cette guerre). En deuxième lieu, comme les mouvements migratoires à court terme sont à la hausse depuis 1960, on n'utilise plus qu'un échantillon d'enregistrements (environ 1 sur 20) pour l'analyse statistique depuis 1971. En troisième lieu, il peut arriver que les voyageurs soient classés dans la mauvaise catégorie de mouvements migratoires. Pour éviter ces erreurs de classification, on compare les estimations démographiques, les chiffres du recensement et les estimations de l'EP pour le jour du recensement en fonction du lieu effectif de résidence, qui englobe les trois catégories de mouvements migratoires.

Pour évaluer le taux de sous-dénombrement dans le recensement de 1986, nous avons établi des estimations de la population selon l'âge et le sexe en date du recensement en nous servant de données sur les naissances, les décès et la migration externe qui remontent à 1921 ainsi que des résultats du recensement de 1921. Les estimations de la population de 65 ans et moins reposent donc essentiellement sur les données relatives aux naissances, aux décès et à la migration et ne sont aucunement influencées par le degré de précision du recensement de 1921.

4. VALIDATION DES ESTIMATIONS DE L'EP DE 1986

Le tableau ci-dessous donne les chiffres estimés de la population au 30 juin 1986 selon l'âge et le sexe; ces chiffres sont fondés sur les résultats de l'analyse démographique et de l'EP de 1986. Le tableau donne également le nombre de personnes inscrites au régime d'assurance-maladie selon l'âge et le sexe.

Tableau 2

Chiffres estimés de la population de 1986 selon l'âge et le sexe (fondés sur les résultats de l'EP de 1986 et de l'analyse démographique) et nombre de personnes inscrites au régime d'assurance-maladie

Âge	Hommes (milliers)									
	Population				Écart par rapport au recensement			Taux de sous-dénombrement (en pourcentage)		
	Recensement (a)	EP(a)	AD(b)	Régime d'assurance-maladie	EP	AD	Régime d'assurance-maladie	EP	AD	Régime d'assurance-maladie
0- 4	608.3	616.4	612.8	611.4	8.0	4.5	3.1	1.3	0.7	0.5
0- 5	594.9	602.4	603.0	612.3	7.5	8.1	17.4	1.2	1.3	2.8
10-14	660.8	670.4	668.4	674.3	9.6	7.6	13.5	1.4	1.1	2.0
15-19	673.1	688.4	687.7	693.1	15.3	14.6	20.0	2.2	2.1	2.9
20-24	648.5	679.5	681.3	681.3	31.0	32.8	32.8	4.6	4.8	4.8
25-29	649.2	677.7	675.0	688.5	28.5	25.8	39.3	4.2	3.8	5.7
30-34	615.5	630.0	650.1	647.5	14.5	34.6	32.0	2.3	5.3	4.9
35-39	622.2	634.2	632.7	646.2	12.0	10.5	24.0	1.9	1.7	3.7
40-44	504.2	512.6	517.0	522.3	8.4	12.8	18.1	1.6	2.5	3.5
45-49	419.8	427.0	416.5	436.8	7.2	-3.3	17.0	1.7	-0.8	3.9
50-54	363.7	371.2	371.4	377.9	7.5	7.7	14.2	2.0	2.1	3.8
55-59	373.4	379.5	384.9	386.6	6.1	11.5	13.2	1.6	3.0	3.4
60-64	341.1	347.0	348.1	350.6	5.9	7.0	9.5	1.7	2.0	2.7
65-69	259.6	263.6	251.8	265.5	4.0	-7.8	5.9	1.5	-3.1	2.2
70-74	204.2	208.2	200.8	213.0	4.0	-3.4	8.8	1.9	-1.7	4.1
75+	229.5	233.0	181.5	250.1	3.5	-48.0	20.6	1.5	-26.4	8.2
Total	7768.3	7941.0	7883.1	8057.3	172.7	114.8	289.0	2.2	1.5	3.6

Âge	Femmes (milliers)									
	Population				Écart par rapport au recensement			Taux de sous-dénombrement (en pourcentage)		
	Recensement (a)	EP(a)	AD(b)	Régime d'assurance-maladie	EP	AD	Régime d'assurance-maladie	EP	AD	Régime d'assurance-maladie
0- 4	579.7	591.0	583.8	580.9	11.3	4.1	1.2	1.9	0.7	0.2
5- 9	565.1	572.4	565.5	582.1	7.3	0.4	17.0	1.3	0.1	2.9
10-14	628.0	636.8	630.2	641.8	8.8	2.2	13.8	1.4	0.3	2.2
15-19	644.1	657.4	651.4	666.3	13.3	7.3	22.2	2.0	1.1	3.3
20-24	633.1	652.5	644.4	670.4	19.4	11.3	37.3	3.0	1.8	5.6
25-29	648.7	660.7	665.4	684.1	12.0	16.7	35.4	1.8	2.5	5.2
30-34	618.1	627.8	631.2	643.9	9.7	13.1	25.8	1.5	2.1	4.0
35-39	612.1	619.1	600.2	626.3	7.0	-11.9	14.2	1.1	-2.0	2.3
40-44	482.6	488.6	489.6	495.4	6.0	7.0	12.8	1.2	1.4	2.6
45-49	399.1	403.0	397.9	411.6	3.9	-1.2	12.5	1.0	-0.3	3.0
50-54	349.1	354.6	343.9	358.6	5.5	-5.2	9.5	1.6	-1.5	2.6
55-59	362.6	366.5	362.4	372.4	3.9	-0.2	9.8	1.1	-0.1	2.6
60-64	358.2	364.4	351.3	365.3	6.2	-6.9	7.1	1.7	-2.0	1.9
65-69	298.2	302.2	301.9	306.7	4.0	3.7	8.5	1.3	1.2	2.8
70-74	259.0	262.9	262.2	269.7	3.9	3.2	10.7	1.5	1.2	4.0
75+	396.2	404.7	385.0	434.1	8.5	-11.2	37.9	2.1	-2.9	8.7
Total	7833.8	7964.6	7866.2	8109.6	130.8	32.4	275.8	1.6	0.4	3.4

(a) Selon le lieu effectif de résidence
(b) Estimations démographiques fondées sur le recensement de la population de 1921 et des événements démographiques survenus par la suite.

Il existe une très forte correspondance entre les estimations de l'EP et les estimations démographiques concernant la population de sexe masculin, plus particulièrement en ce qui a trait aux hommes de moins de 30 ans. On note toutefois un écart appréciable en ce qui concerne les hommes de 30 à 34 ans, les estimations démographiques excédant de 20,000 les estimations de l'EP. Cet écart peut trouver son explication dans les mouvements migratoires à court terme qui ont eu lieu en Australie entre 1981 et 1986 et qui ont amené une forte augmentation nette du nombre d'hommes appartenant à ce groupe d'âge. Comme on n'observe pas d'augmentation de cet ordre dans les groupes d'âge voisins, on pourrait croire qu'il s'est glissé des erreurs dans les données sur les courants migratoires. Les mouvements migratoires étant très nombreux en Australie (plus de 6 millions en 1986), une petite erreur dans la déclaration de l'âge ou dans le traitement peut fausser dans une assez grande mesure les estimations démographiques en chiffres absolus nets. La possibilité d'erreurs dans les estimations démographiques est d'autant plus réelle que le taux de sous-dénombrement pour le groupe des 30 à 34 ans (5.3%) est très élevé par rapport aux taux pour les groupes d'âge voisins.

De fait, il est très possible que l'EP ait mal mesuré le taux de sous-dénombrement des visiteurs d'outre-mer. Quoi qu'il en soit, les erreurs d'estimation concernant la portion de la population qui est de passage en Australie ne devraient pas influer sur le degré de précision des estimations officielles de la population puisque celles-ci sont fondées sur le principe du lieu de résidence habituel et ne comprennent pas les visiteurs.

En ce qui concerne les femmes, le degré de correspondance entre les résultats de l'EP et les estimations démographiques pour les personnes de moins de 35 ans est satisfaisant. Toutefois pour certains groupes d'âge, les estimations démographiques sont beaucoup moins élevées que les estimations de l'EP, et pour les personnes de 35 à 39 ans et de 45 à 64 ans, les estimations démographiques sont inférieures aux chiffres du recensement bruts. De fait, les estimations démographiques pour ces groupes paraissent franchement trop faibles, ce qui tend à confirmer que les estimations démographiques ne sont pas suffisamment précises pour produire des estimations de la population et devraient servir uniquement à évaluer les résultats des EP.

Les taux de sous-dénombrement établis à l'aide des données de l'EP pour les divers groupes d'âge sont beaucoup plus uniformes que les taux établis à l'aide de l'analyse démographique. Les taux établis par l'EP pour les personnes de 20 à 29 ans sont plus élevés que pour les personnes des autres groupes d'âge et cela est normal étant donné les taux de mobilité plus élevés chez les jeunes adultes, particulièrement les hommes.

Le nombre de personnes inscrites au régime d'assurance-maladie est beaucoup plus élevé que les estimations de l'EP et les estimations démographiques, sauf en ce qui concerne le groupe des 0 à 4 ans, où c'est le contraire. Selon certaines études, cet écart s'explique par le délai qui s'écoule entre la naissance d'un enfant et le moment où celui-ci est inscrit au régime d'assurance-maladie; de même, pour ce qui est des autres groupes d'âge, l'écart s'explique par le délai qui s'écoule entre le moment où une personne meurt ou émigre et le moment où son nom est rayé du registre de l'assurance-maladie.

En comparant les estimations de l'EP à des estimations tirées du registre des allocations familiales et du registre des inscriptions scolaires pour certains groupes d'âge, on constate un degré de correspondance acceptable. Ces résultats tendent à confirmer la précision des estimations de l'EP en ce qui concerne les premiers groupes d'âge.

Même s'il existe un degré de correspondance acceptable entre les estimations de l'EP et d'autres estimations de la population, il subsiste deux problèmes que nous devons tenter de résoudre avant de pouvoir accepter les estimations de l'EP. Le premier problème vient d'une analyse des estimations de l'EP concernant les taux de sous-dénombrement dans le recensement selon l'âge et le sexe. Ces taux figurent dans le tableau 2.

Sauf pour les groupes de 0 à 4 ans et de 75 ans et plus, les taux de sous-dénombrement sont généralement plus élevés pour les hommes que pour les femmes. Tandis que les taux de sous-dénombrement pour les personnes de 75 ans et plus peuvent être influencés par une faible taille d'échantillon, le taux pour les filles de 0 à 4 ans semble exagéré (1.9% par rapport à 1.3% pour

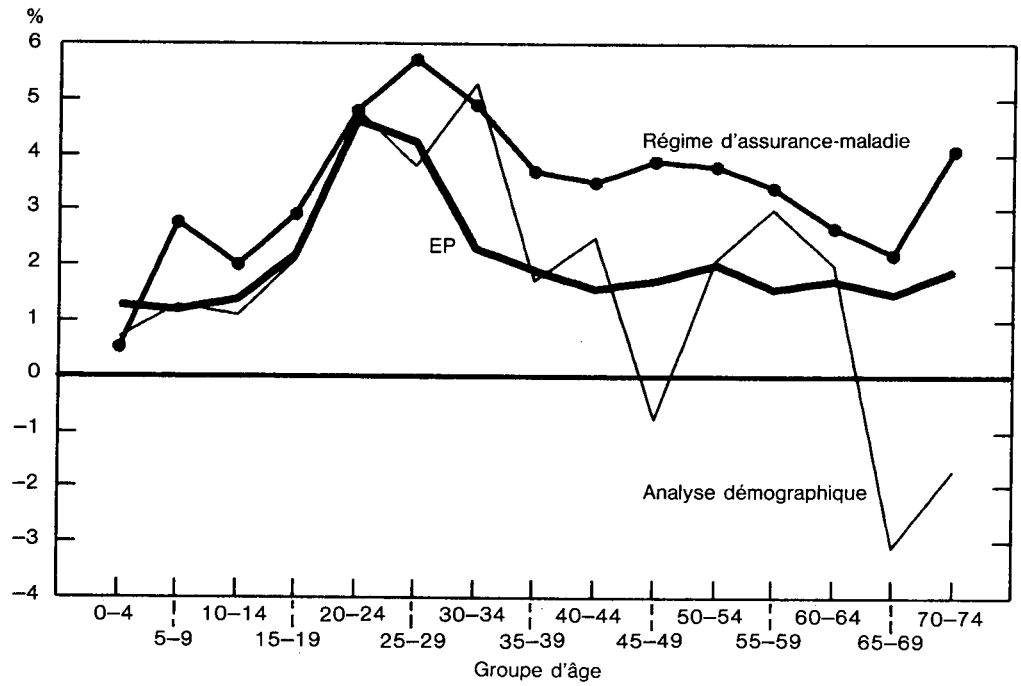

Figure 1. Taux de sous-dénombrement (en pourcentage) dans le recensement de 1986: enquête postcensitaire, analyse démographique et registre du régime d'assurance-maladie - HOMMES

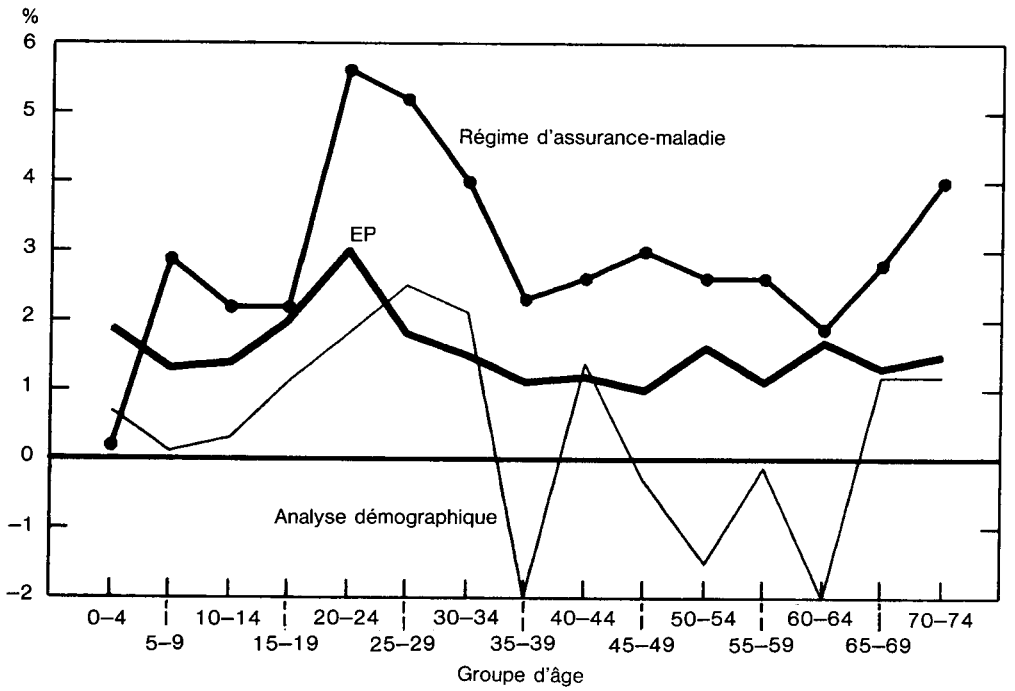

Figure 2. Taux de sous-dénombrement (en pourcentage) dans le recensement de 1986: enquête postcensitaire, analyse démopgraphique et registre du régime d'assurance-maladie - FEMMES.

les garçons du même âge et les filles de 5 à 9 ans). L'EP a permis d'estimer à 11,300 le nombre de filles de 0 à 4 ans qui ont été oubliées dans le recensement, comparativement à environ 7,000 pour les filles de 5 à 9 ans. Cette différence sensible n'existe pas du côté des garçons.

Le rapport de masculinité établi par l'EP pour le groupe d'âge 0-4 est de 104.3 garçons pour 100 filles, ce qui est inférieur au rapport de masculinité établi à l'aide des chiffres du recensement (104.9) et au rapport de masculinité établi à l'aide des données démographiques (105.0 garçons pour 100 filles).

Les chiffres ci-dessus donnent à penser que l'EP a surestimé le nombre de filles de 0 à 4 ans bien que nous voyions difficilement comment cette enquête aurait pu surestimer l'effectif de ce groupe plutôt que l'effectif de tout autre groupe.

Tableau 3

Comparaison entre les résultats de l'EP de 1986 et des estimations indépendantes

	Estimations de l'EP	Estimations démographiques	Allocations familiales	Inscriptions scolaires
		Personnes (milliers)		
0- 4	1207.3	1196.6	1204.8(a)	-
5- 9	1174.8	1168.5	1177.0	-(b)
10-14	1307.2	1298.6	1304.2	1289.6

(a) Le nombre d'enfants d'âge 0 inscrits au programme d'allocations familiales est sous-estimé à cause du délai qui s'écoule entre le jour de la naissance et le moment où l'enfant est inscrit au programme d'allocations familiales. Pour redresser le chiffre en question, on a remplacé le nombre d'enfants d'âge 0 inscrits au programme par une estimation tirée de l'analyse démographique.

(b) Les enfants de 5 ans ne sont pas obligés de fréquenter l'école.

Tableau 4

Taux de sous-dénombrement (%) selon l'état ou le territoire (établi à l'aide des données de l'EP)

	Suivant le lieu de résidence effectif	Suivant le lieu de résidence habituel
Nouvelles-Galles-du-Sud	1.54	1.51
Victoria	1.59	1.77
Queensland	2.68	2.43
Australie méridionale	1.54	1.59
Australie occidentale	2.32	2.26
Tasmanie	1.32	1.16
Territoire du Nord	9.97	6.45
Territoire fédéral de la capital	1.95	1.61
Australie	1.91	1.84

Tableau 5

Chiffres du recensement de 1986 redressés en fonction du sous-dénombrement, selon l'âge et le sexe

Âge	Selon le lieu de résidence effectif					
	Hommes		Femmes		Total	
	Nombre (milliers)	Taux de sous-dénombrement	Nombre (milliers)	Taux de sous-dénombrement	Nombre (milliers)	Taux de sous-dénombrement
0- 4	616.3	1.30	586.6	1.17	1202.9	1.24
5- 9	602.4	1.24	572.4	1.27	1174.8	1.26
10-14	670.1	1.39	636.8	1.38	1306.9	1.39
15-19	688.3	2.19	657.3	2.02	1345.6	2.11
20-24	679.4	4.54	652.4	2.95	1331.8	3.76
25-29	677.5	4.17	660.7	1.81	1338.2	3.00
30-34	629.9	2.29	627.8	1.55	1257.7	1.92
35-39	634.0	1.87	618.9	1.11	1252.9	1.49
40-44	512.6	1.64	488.5	1.21	1001.1	1.43
45-49	426.9	1.66	403.0	0.98	829.9	1.33
50-54	371.2	2.04	354.6	1.56	725.8	1.80
55-59	379.5	1.62	366.5	1.06	746.0	1.34
60-64	347.0	1.70	364.4	1.70	711.4	1.70
65-69	263.6	1.52	302.3	1.35	565.9	1.43
70-74	208.2	1.92	262.9	1.47	471.1	1.67
75+	233.0	1.49	404.7	2.08	637.7	1.86
Tous les âges	7940.1	2.16	7959.7	1.58	15899.8	1.87

Âge	Selon le lieu de résidence habituel					
	Hommes		Femmes		Total	
	Nombre (milliers)	Taux de sous-dénombrement	Nombre (milliers)	Taux de sous-dénombrement	Nombre (milliers)	Taux de sous-dénombrement
0- 4	615.3	1.29	585.9	1.22	1201.2	1.26
5- 9	601.3	1.23	571.2	1.22	1172.5	1.23
10-14	668.5	1.29	635.7	1.36	1304.2	1.33
15-19	685.6	2.11	654.3	1.97	1339.9	2.04
20-24	673.1	4.33	646.9	2.83	1320.0	3.59
25-29	672.6	4.02	657.2	1.80	1329.8	2.92
30-34	626.6	2.21	625.6	1.53	1252.2	1.87
35-39	630.9	1.78	616.7	1.05	1247.6	1.41
40-44	510.3	1.59	487.0	1.19	997.3	1.39
45-49	424.7	1.52	401.7	0.98	826.4	1.26
50-54	369.6	1.97	353.0	1.52	722.6	1.75
55-59	377.7	1.52	364.0	0.92	741.8	1.22
60-64	345.6	1.74	361.6	1.61	707.3	1.67
65-69	262.1	1.47	300.2	1.31	562.3	1.38
70-74	207.2	1.89	261.3	1.46	468.5	1.65
75+	232.4	1.52	403.3	2.01	635.7	1.83
Tous les âges	7903.6	2.08	7925.5	1.54	15829.1	1.81

Le second problème a trait au taux de sous-dénombrement estimé très élevé qui a été établi par l'EP pour le Territoire du Nord. Comme l'indique le tableau 4, ce taux est de 9.97% lorsqu'on tient compte du lieu de résidence effectif et de 6.45% lorsqu'on tient compte du lieu de résidence habituel. Le Territoire du Nord est un région peu peuplée (selon le recensement de 1986, 154,800 habitants répartis sur 1.3 million de kilomètres carrés) qui est caractérisée par une population très mobile. La population estimée du Territoire du Nord selon l'EP est considérablement plus élevée que la population estimée selon le recensement de 1981. En comparant les estimations de l'EP pour le Territoire du Nord avec des estimations indépendantes comme le nombre d'enfants inscrits au registre des allocations familiales et le nombre d'enfants inscrits à l'école, on se rend compte également que les estimations de l'EP sont élevées. Même en reconnaissant que les estimations indépendantes peuvent être entachées d'erreur, nous pouvons affirmer avec assez de certitude que l'EP a surestimé le taux de sous-dénombrement pour le TN.

Après avoir vérifié les questionnaires de l'EP pour le Territoire du Nord, on a jugé qu'ils étaient acceptables sauf pour un district de collecte, où des problèmes de contrôle des adresses et de déplacement dans des régions peu accessibles ont fait ressortir des lacunes dans les instructions concernant les opérations sur le terrain et ont engendré des problèmes d'appariement.

On a jugé bon de rajuster à la baisse le nombre estimé de filles de 0 à 4 ans et la population estimée du TN en redressant les résultats de l'EP. En ce qui a trait aux filles de 0 à 4 ans, on a multiplié le rapport de masculinité établi à l'aide des estimations démographiques par le nombre estimé de garçons de 0 à 4 ans selon l'EP. Cela revenait essentiellement à remplacer l'estimation établie par l'EP pour les filles de 0 à 4 ans par une meilleure estimation en ayant recours au rapport de masculinité et au nombre estimé de garçons selon l'EP. On a pu ainsi réduire de 4,000 le nombre estimé de filles de 0 à 4 ans, le nouvel effectif estimé étant de 587,000.

En ce qui concerne la population du TN, on a décidé de ne pas se servir des données du district de collecte qui posait problème. Cela a eu pour effet de ramener le taux de sous-dénombrement pour le Territoire du Nord à 9.1% (calculé selon le lieu de résidence effectif) et à 5.5% (selon le lieu de résidence habituel).

Par suite de ce double redressement, le taux de sous-dénombrement pour l'ensemble du pays est passé de 1.91% à 1.87% (selon le lieu de résidence effectif) ou de 1.84% à 1.81% (selon le lieu de résidence habituel). Le tableau 5 donne les estimations de l'EP selon l'âge et le sexe après que les estimations relatives à la population du TN et aux filles de 0 à 4 ans ont été redressées.

5. ESTIMATIONS DÉMOGRAPHIQUES INFRA-NATIONALES

Les données sur la migration interne ne sont pas assez fiables pour que les estimations démographiques infra-nationales servent à évaluer le taux de sous-dénombrement dans le recensement. Néanmoins, on a comparé pour chaque État et chaque Territoire le nombre estimé d'enfants de 1 à 15 ans selon l'EP de 1986 avec le nombre d'enfants inscrits au registre des allocations familiales. De façon générale, on observe une concordance entre les deux séries de données sauf pour ce qui a trait au Territoire du Nord, où l'écart est supérieur à 2%.

Compte tenu de cette concordance entre les deux séries de données et du fait que l'on ne peut comparer les estimations de l'EP à des données indépendantes fiables touchant les groupes d'âge supérieurs à cause de l'absence de telles données, les estimations (redressées) de l'EP concernant la population des États et des Territoires ont été acceptées.

Les estimations de la population des États et des Territoires selon l'âge et le sexe et de la population des districts d'administrations locales n'ont pas été tirées directement de l'EP. L'EP de 1986 était une enquête par sondage et ses résultats sont exposés à l'erreur d'échantillonnage. Lorsqu'il s'agit d'estimations selon l'âge et le sexe au niveau de l'État ou du Territoire ou encore au niveau du district d'administration locale, les erreurs d'échantillonnage sont élevées, souvent

exagérément élevées, par rapport aux corrections qu'il faut faire pour tenir compte du sous-dénombrement. On a donc procédé de façon indirecte pour estimer la population des États et des Territoires selon l'âge et le sexe en appliquant une méthode d'ajustement proportionnel itératif (API) aux estimations de l'EP se rattachant à un niveau supérieur et comportant une faible erreur d'échantillonnage. Pour avoir une description de l'API, voir Purcell et Kish (1979). L'API consiste à redresser les chiffres du recensement selon l'âge et l'État ou Territoire en fonction des estimations nationales de la population selon l'âge et le sexe et des estimations de la population des États ou Territoires selon le sexe.

L'API comporte les cycles suivants $n = 0, 1, \ldots$.

$$X_{gas}^{(2n+1)} = X_{gas}^{(2n)} \frac{X_{as}}{X_{as}} \quad (2n)$$

$$X_{gas}^{(2n+2)} = X_{gas}^{(2n+1)} \frac{X_{gs}}{X_{gs}^{(2n+1)}}$$

et $X_{gas}^{(o)} = Y_{gas}$ le chiffre du recensement pour l'État g, le groupe d'âge a et le sexe s. Cette méthode converge vers une solution unique. L'utilisation de l'API suppose évidemment que la relation entre les variables dans la structure d'association est valide et qu'elle est préservée.

En ce qui concerne les estimations de population pour les districts d'administrations locales, la question de l'erreur d'échantillonnage est plus délicate et les résultats de l'EP ne sont pas assez fiables pour que l'on puisse estimer directement le taux de sous-dénombrement qui s'applique à chaque district d'administration locale. Compte tenu de l'hypothèse que le taux de sous-dénombrement varie selon l'âge et le sexe ou le lieu de naissance (né en Australie ou né à l'étranger) et qu'il diffère suivant l'État ou Territoire et selon qu'il s'agit de la capitale ou du reste de l'État, le redressement des estimations touchant les districts d'administrations locales a été exécuté de manière à tenir compte de tous ces critères.

6. PROBLÈMES LIÉS AU PROCESSUS D'ESTIMATION DE L'EP

Comme le souligne Bailar (1985) par exemple, le biais et la convergence des estimations de l'EP sont influencés par les erreurs d'appariement, par l'existence d'une corrélation entre le fait qu'une personne a été oubliée dans le recensement et le fait qu'elle a été aussi oubliée dans l'EP et par l'existence d'enregistrements erronés dans le recensement ou l'EP. C'est précisément à cause des effets probables de ces facteurs que l'on évalue les résultats de l'EP en utilisant de la façon décrite ci-dessus des données démographiques et administratives.

Les erreurs d'appariement auront pour effet d'introduire un biais dans les estimations de l'EP. Le non-appariement d'enregistrements qui devraient normalement pouvoir être appariés aura pour effet de gonfler le nombre de personnes oubliées dans le recensement et l'EP produira alors des valeurs surestimées. Les cas de "fausse concordance" auront l'effet contraire.

La présence d'enregistrements erronés dans le recensement ou l'EP aura pour effet de gonfler la valeur de Y ou de x et, par voie de conséquence, l'estimation établie par l'EP. Le U.S. Bureau of the Census prélève un échantillon spécial (appelé échantillon D) parmi les enregistrements du recensement afin d'estimer le nombre d'enregistrements erronés dans le recensement; on supprime l'incidence de ces enregistrements sur les estimations officielles en redressant le chiffre du recensement Y. Pour une description de l'échantillon D, voir Fay, Passel et Robinson (1988). Les méthodes d'appariement et d'estimation utilisées par l'ABS visent à corriger en partie l'effet des enregistrements erronés en vérifiant non seulement si une personne a été recensée mais encore si elle aurait dû l'être ou si elle l'a été plus d'une fois. Par exemple, l'EP de 1986 a permis de déterminer que 250 personnes avaient été recensées deux fois et 4 personnes l'avaient été trois

fois. Elle a aussi permis de découvrir que des personnes avaient été recensées par erreur. De cette façon, si on assimile le processus d'estimation de l'EP à l'estimation par quotient plutôt qu'à l'estimation de système dual, on est plus en mesure d'établir le nombre d'enregistrements erronés.

L'estimation de système dual suppose que le fait qu'une personne soit oubliée ou non dans l'EP n'a rien à voir avec le fait qu'elle ait été recensée ou non. Bien que l'on ait pris toutes les mesures nécessaires pour séparer parfaitement les deux systèmes de collecte et de traitement, il peut toujours exister une certaine corrélation. Une corrélation positive signifiera que la valeur estimée de l'EP fondée sur l'hypothèse d'indépendance donne une sous-estimation; une corrélation négative signifiera que la valeur estimée de l'EP donne une surestimation. Il y a corrélation négative si le fait qu'une personne a été recensée implique qu'il sera difficile de la dénombrer dans l'EP; cependant, il n'y a pas de donnée qui permette d'affirmer cela sans contredit; le taux de réponse final pour l'EP (95%) est conforme à ceux observés dans d'autres enquêtes-ménages de l'ABS. Une corrélation positive semble plus probable; du reste, l'existence d'une telle corrélation aurait été constatée dans le recensement de 1981. Si, de fait, il y a corrélation positive, les opérations de redressement fondées sur l'EP n'auront pas donné des résultats tout à fait satisfaisants mais auront été néanmoins valables.

7. AUTRES MÉTHODES D'ESTIMATION (WOLTER 1986)

La combinaison de données de l'EP et de rapports de masculinité établis à l'aide de données démographiques ou d'autres sources est à la base des méthodes proposées par Wolter (1986). Wolter propose en effet plusieurs modèles et méthodes connexes qui combinent formellement des rapports de masculinité avec des estimations de l'EP. Ces méthodes visent à assouplir l'hypothèse d'indépendance qui est caractéristique des méthodes d'estimation de l'EP.

Tableau 6

Rapports de masculinité: Hommes pour 100 femmes

Âge	Autre	EP
0- 4	105.0	104.3
5- 9	105.2	105.2
10-14	105.2	105.3
15-19	104.7	104.7
20-24	104.1	104.1
25-29	102.6	102.6
30-34	100.3	100.3
35-39	102.4	102.4
40-44	104.5	104.9
45-49	105.2	106.0
50-54	104.2	104.7
55-59	103.0	103.5
60-64	95.2	95.2
65-69	87.1	87.2
70-74	78.8	79.2
75+	57.9	57.6

Wolter considère deux modèles. Dans le premier, il suppose que le degré de corrélation entre les probabilités de sous-dénombrement dans le recensement et dans l'EP (lequel degré est mesuré par les rapports de produits croisés établis à l'aide de tables comme le diagramme exposé dans les pages précédentes) est le même pour les hommes et les femmes dans chaque groupe d'âge. Dans le second modèle, Wolter pose l'hypothèse d'indépendance pour les femmes et utilise un rapport de masculinité déterminé à l'aide de données indépendantes pour calculer le degré de corrélation pour les hommes. Il est alors possible de déduire les rapports de produits croisés pour les hommes.

Selon une première étude, où l'on a appliqué ces méthodes à des données de l'Australie, on a observé que le premier modèle produisait des estimations de rapports de produits croisés très irrégulières, la moitié environ des valeurs estimées étant négatives. Le second modèle produisait beaucoup moins de valeurs estimées négatives mais il en restait un certain nombre; ces valeurs négatives allaient être ramenées à zéro dans un modèle modifié. Wolter (1986, p. 7) souligne aussi le problème des rapports de produits croisés négatifs. Une fois modifié, le second modèle a été appliqué à des données de 1986. Les rapports de masculinité tirés des données de l'EP pour les groupes d'âge 5-9 à 35-39 étaient conformes aux prévisions et leur application donnait exactement la valeur estimée de l'EP. En ce qui a trait au groupe d'âge 0-4, on a utilisé un rapport de masculinité établi à l'aide d'estimations démographiques et pour les groupes d'âge 40-44 et suivants, on s'est servi de rapports de masculinité estimés à l'aide des chiffres du recensement. Les rapports de masculinité sont reproduits dans le tableau 6.

Comme le rapport de masculinité utilisé n'est pas très différent des rapports de masculinité de l'EP, l'application du second modèle de Wolter a pour effet de modifier très légèrement les estimations de l'EP. Pour ce qui est des groupes d'âge 0-4 et 75 + , le nombre estimé de personnes de sexe masculin augmente de 0.7% et de 0.5% respectivement. En ce qui a trait aux groupes d'âge 45-49 et 70-74, les estimations sont réduites d'environ 0.6%. Cette analyse donne à penser que l'action combinée des problèmes évoqués ci-dessus créerait des différences de biais relativement faibles entre les sexes dans la méthode d'estimation de l'EP. Il se pourrait que des biais soient à peu près identiques pour les hommes et pour les femmes, de sorte que les rapports de masculinité de l'EP seraient généralement acceptables.

Les recensements de 1981 et de 1986 ont prouvé la nécessité des rapports de masculinité dans l'évaluation des mesures de sous-dénombrement et nous croyons que la méthode de Wolter est un moyen utile de produire des estimations auxquelles peuvent être comparés les chiffres du recensement et les estimations de l'EP. Le fait que les rapports de masculinité de l'EP de 1986 ont été généralement acceptables signifie que l'utilisation de la méthode de Wolter a changé peu de choses. Sauf en ce qui concerne le groupe d'âge 0-4, la situation est tout à fait différente de celle qu'on a connue en 1981, lorsqu'on a jugé nécessaire de redresser les estimations de l'EP pour certains groupes d'âge en se fondant sur des rapports de masculinité indépendants. Cette différence de situation entre 1981 et 1986 reflète peut-être une diminution du degré de corrélation entre les probabilités de sous-dénombrement dans le recensement et dans l'EP en 1986.

8. CONCLUSION

Tandis que l'ABS a redressé les chiffres des trois derniers recensements pour tenir compte du sous-dénombrement, la fiabilité des estimations de l'EP tient au fait que cette enquête produit des résultats comparables à ceux d'autres sources. On ne prévoit pas une modification majeure de la méthode pour le recensement de 1991. Cependant, nous croyons nécessaire d'approfondir les causes possibles de biais en vérifiant notamment la conformité des méthodes d'appariement manuelles et en analysant les méthodes qui visent à éliminer le biais de corrélation. On prévoit aussi étudier la possibilité de mettre sur pied une banque de données démographiques selon le lieu de résidence habituel de manière à éliminer ou à réduire les effets des nombreux mouvements migratoires à court terme.

REMERCIEMENT

Nous tenons à remercier les personnes qui ont révisé cet article pour leurs commentaires et leurs propositions concernant des sujets de recherche.

ANNEXE 1

L'ENQUÊTE POSTCENSITAIRE DE 1986

Généralités

L'EP de 1986 a été réalisée dans la quatrième et cinquième semaine qui ont suivi le jour du recensement. À cet occasion, on a interviewé un échantillon de la population vivant dans environ 35,000 logements privés (2/3 de un pour cent des logements) répartis dans toute l'Australie et abritant environ 100,000 personnes. Le taux de sondage variait selon l'État ou le Territoire, les États ou Territoires moins étendus ayant des taux de sondage plus élevés. Les intervieweurs recueillaient des données personnelles comme le nom, l'âge, le sexe, l'état matrimonial et le lieu de naissance en vue de les apparier avec les données contenues dans les questionnaires du recensement. Pour chaque personne visée par l'enquête, l'intervieweur inscrivait le lieu de résidence habituel, le lieu où la personne se trouvait le soir du recensement, l'adresse de la personne avant et après le jour du recensement et toute autre adresse à laquelle la personne aurait pu être recensée. À chaque adresse donnée, les renseignements personnels étaient comparés aux données contenues dans les questionnaires du recensement afin de déterminer si la personne en question avait été recensée ou non ou de définir le nombre de fois qu'elle avait été recensée si elle l'avait été plus d'une fois.

Champ de l'EP et structure de l'échantillon

Hormis les cas particuliers mentionnés ci-dessous, le champ de l'EP comprenait toutes les personnes qui devaient avoir été recensées, sauf celles qui étaient à l'étranger ou qui sont décédées entre le jour du recensement et le jour de l'EP. Les agents diplomatiques et les personnes en résidence diplomatique n'ont pas été recensés; ces gens ont donc été exclus du champ de l'enquête, comme les enfants nés après le recensement. Pour les personnes qui étaient à l'étranger le jour du recensement mais qui étaient incluses dans le champ de l'EP, on a tenté un appariement avec les questionnaires du recensement pour voir si elles n'avaient pas été recensées par erreur.

Pour des raisons pratiques, les régions très peu peuplées ont été exclues du champ de l'enquête. Lors du recensement, des procédures spéciales avaient été appliquées pour rejoindre et recenser la population aborigène, les travailleurs des camps miniers et des fermes d'élevage, etc. Pour réaliser l'EP dans ces régions, il aurait fallu appliquer les mêmes procédures qu'au recensement, ce qui n'aurait pas permis un calcul juste et indépendant du taux de sous-dénombrement. C'est pourquoi ces régions ont été exclues du champ de l'EP.

Les logements non privés ou spéciaux comme les hôpitaux, les hôtels et les motels ont aussi été exclus du champ de l'enquête. La grande majorité des personnes qui se trouvent dans ces logements y sont habituellement pour une période relativement courte et selon les règles normales du ABS en matière d'enquête, ces personnes seraient susceptibles d'être enregistrées à leur lieu de résidence habituel, où seraient fournis les renseignements pertinents. Par cette règle, on s'est donc trouvé à exclure du champ de l'enquête un nombre relativement restreint de personnes qui demeurent en permanence dans ces logements. Pour les besoins de l'estimation, on a supposé que le taux de sous-dénombrement pour la population ne faisant pas partie du champ d'enquête était le taux moyen pour la capitale ou une autre ville, selon le cas, pour chaque État et le taux moyen pour le Territoire dans le cas des deux Territoires.

Comme les personnes demeurant en permanence dans des logements non privés ou spéciaux et celles vivant dans les régions peu peuplées représentent globalement moins de 3% de la population totale, une différence de taux de sous-dénombrement entre ces groupes et les groupes inclus dans le champ de l'EP a peu de chances d'influer beaucoup sur le taux de sous-dénombrement estimé pour l'État ou le pays en général.

Lien entre le recensement et l'EP

Il est essentiel que l'EP soit aussi indépendante que possible du recensement, sinon les facteurs qui ont favorisé un sous-dénombrement ou un surdénombrement lors du recensement peuvent se retrouver dans l'EP, ce qui aurait pour effet d'introduire un biais dans l'estimation du taux de sous-dénombrement. De plus, le fait de connaître les régions ou secteurs qui doivent être inclus dans le champ de l'EP pourrait conditionner le comportement des agents recenseurs dans ces secteurs, de sorte que l'échantillon de l'EP ne refléterait pas fidèlement le taux de sous-dénombrement dans le recensement. C'est pourquoi on a tenu à mobiliser deux équipes distinctes pour le recensement et l'EP, tant sur le terrain qu'au bureau central. Les intervieweurs de l'EP n'avaient pas agi comme agents recenseurs ou chefs d'équipe et le personnel affecté aux opérations sur le terrain ne connaissait pas les régions ou secteurs qui allaient être inclus dans le champ de l'EP.

L'indépendance des enquêtes était assurée de deux autres façons : par la création de deux systèmes de collecte qui fonctionnent indépendamment l'un de l'autre et par l'adoption de procédures spéciales concernant les questionnaires du recensement reçus par la poste après que les opérations de l'EP ont été engagées.

Pour assurer l'indépendance des opérations des deux enquêtes, on a débuté l'EP après que tous les questionnaires du recensement dûment remplis ont été ramassés. Ainsi, les agents recenseurs n'étaient pas sur le terrain au même moment que les intervieweurs de l'EP et il n'y avait aucune possibilité qu'ils s'influencent les uns les autres, même involontairement.

Il fallait prévoir des procédures spéciales pour les questionnaires du recensement reçus après le début de l'EP pour annuler l'effet que pouvaient avoir les opérations de l'EP sur les membres responsables du ménage, qui n'avaient pas remis leur questionnaire à temps. À certaines adresses, les intervieweurs de l'EP ont découvert des questionnaires du recensement qui n'avaient pas encore été ramassés. Cela s'expliquait par le fait que les personnes en question préféraient retourner leur questionnaire par la poste et ne l'avaient pas encore fait ou que l'agent recenseur n'avait pu entrer en contact avec ces personnes pour ramasser le questionnaire. Si ce n'avait été de l'EP, quelques-unes de ces personnes n'auraient peut-être jamais retourné leur questionnaire du recensement. Pour éliminer tout risque de biais attribuable à ce comportement, les questionnaires du recensement retournés par la poste après le lundi, 20 juillet 1986 (jour où a débuté l'EP) étaient considérés en retard. Des procédures spéciales pour le traitement des questionnaires remis en retard sont énoncées plus loin dans cette annexe.

Procédures d'appariement de l'EP

Les opérations d'appariement visant à déterminer si une personne a été oubliée, recensée une seule fois ou recensée plus d'une fois, se sont déroulées en deux étapes. Celles-ci consistaient en des procédés administratifs appliqués par les employés du Centre de transcription des données du recensement.

La première étape consistait à repérer les questionnaires du recensement correspondant aux adresses des ménages inclus dans l'échantillon de l'EP. Les opérations de traitement du recensement de 1986 étaient centralisées à Sydney. Les employés du Centre de transcription des données du recensement devaient comparer l'adresse indiquée sur la page frontispice de la formule d'interview de l'EP avec toutes les adresses qui figuraient dans le registre de l'agent recenseur responsable du district de collecte (DC) où demeurait le ménage en question. Le registre précité était un moyen de contrôle pour la livraison et le ramassage des questionnaires du recensement et contenait des renseignements comme le nom et l'adresse des membres de tous les ménages du DC de même que le nombre de personnes dans chacun de ces ménages.

Lorsque l'adresse du ménage était imprécise, par exemple dans les régions rurales, les préposés du traitement devaient utiliser aussi le nom du membre responsable du ménage, le nom de la propriété, etc. pour mieux identifier le ménage. De plus, les préposés devaient passer en revue toutes les adresses inscrites dans le registre de manière à déceler les enregistrements répétés. Ils vérifiaient aussi les adresses figurant dans les registres des DC voisins lorsque l'adresse du ménage échantillonné était aux limites du DC.

La deuxième étape consistait à apparier les enregistrements et on se servait pour cela du nom et des caractéristiques démographiques des personnes enregistrées lors du recensement et de l'EP. Durant le processus d'appariement, une formule de recherche était produite chaque fois que l'adresse indiquée pour un membre du ménage échantillonné différait de l'adresse du logement correspondant. La formule de recherche était traitée comme la formule d'interview de l'EP et on tentait de repérer le questionnaire du recensement qui correspondait à l'adresse indiquée sur la formule de recherche.

En règle générale, le processus d'appariement se déroulait sans difficulté. Dans certains cas toutefois, des erreurs d'orthographe et des adresses incomplètes rendaient plus difficile l'appariement de noms. Pour déterminer alors si la personne en question avait été recensée ou non, on se fondait sur d'autres renseignements comme l'âge, le sexe, l'état matrimonial, le lieu de naissance et le lien avec les autres membres du ménage. Dans l'incertitude, les préposés du traitement devaient consulter leur superviseur.

Durant l'interview, le répondant devait aussi indiquer si le nom de chaque membre du ménage figurait dans un questionnaire du recensement. Lorsque l'appariement échouait, faute de renseignements pertinents, la déclaration du répondant était néanmoins tenue pour vrai. Il y a eu quelques cas où ce genre de renseignements n'ont pas été fournis par le répondant; on a alors considéré les personnes en question comme n'ayant pas été recensées.

Après l'appariement, les données ont été enregistrées sur bande, puis contrôlées et restructurées de manière à obtenir un fichier d'enregistrements unitaires épuré, qui indique le nombre de fois qu'un personne incluse dans l'échantillon de l'EP a été enregistrée au cours du même recensement.

Traitement des questionnaires du recensement remis en retard et des «faux» questionnaires

En définissant la formule d'estimation

$X = Y\,(x/y)$, où
X = chiffre du recensement estimé, redressé en fonction du sous-dénombrement
Y = chiffre brut du recensement, non redressé
x = nombre estimé de personnes (selon l'EP) qui auraient dû être recensées
y = nombre estimé de personnes (selon l'EP) qui ont été recensées,

on a considéré les «faux» questionnaires du recensement et les questionnaires remis en retard comme des cas d'omission dans le recensement.

De «faux» questionnaires du recensement ont été créés sur le terrain pour les logements que l'on savait habités mais dont les ménages occupants n'avaient pas retourné leur questionnaire du recensement et n'avaient pu être rejoints. Les agents recenseurs devaient exercer beaucoup de prudence en créant ces «faux» questionnaires et devaient s'assurer, hors de tout doute, que les logements en question étaient occupés le soir du recensement. Ils devaient aussi recueillir autant de renseignements que possible sur le nombre de membres de ces ménages et leurs caractéristiques démographiques.

Lorsqu'on réussissait à apparier une adresse relevée durant l'EP à un «faux» questionnaire du recensement, il était impossible d'exécuter l'appariement de façon satisfaisante à cause de l'absence de noms et de caractéristiques personnelles fiables dans le questionnaire du recensement.

De même, les questionnaires remis en retard ne doivent pas être traités comme les questionnaires qui ont été remis dans les délais prescrits. Comme la visite d'un intervieweur de l'EP peut avoir incité des ménages à remplir leur questionnaire du recensement, bien que tardivement, l'inclusion de ces questionnaires pourrait avoir pour conséquence d'introduire un biais dans l'estimation du taux de sous-dénombrement.

Dans le recensement de 1986, les «faux» questionnaires et les questionnaires remis en retard contenaient les noms de 115,000 personnes, soit 0.7% de la population. Ce nombre n'est pas inclus dans le chiffre du recensement brut (Y) ni dans le nombre estimé de personnes (selon l'EP) qui ont été recensées (y), mais il l'est dans le nombre estimé de personnes (selon l'EP) qui auraient dû être recensées (x). Autrement dit, les personnes dont le nom figurait dans de «faux» questionnaires ou des questionnaires remis en retard ont été considérées comme oubliées et le redressement pertinent a été effectué à l'aide de (x). Le facteur de redressement (x/y) est gonflé du fait que les «faux» questionnaires et les questionnaires remis en retard sont exclus de (y); en revanche, les deux séries de questionnaires sont aussi exclus du chiffre du recensement brut (Y), ce qui compense.

Méthode d'estimation

L'estimation s'est faite selon l'âge, le sexe et la région géographique (division statistique de la capitaleêste de l'État). Des facteurs de redressement ont été inclus dans les formules d'estimation pour tenir compte partiellement des ménages qui n'avaient pas répondu ou qui n'avaient pu être contactés. Ces facteurs permettent de redresser les deux principales valeurs estimées, x et y, en imputant effectivement, pour chaque ménage qui a refusé de répondre ou qui n'a pas été contacté, le nombre moyen de membres d'un ménage et, pour chaque personne ainsi imputée, le taux de sous-dénombrement moyen pour la combinaison âge x sexe x région géographique appropriée. Pour réduire le biais attribuable à l'utilisation de tels facteurs, ceux-ci ont été calculés pour divers sous-groupes de ménages selon le code de dénombrement attribué au recensement (par exemple logement occupé, questionnaire remis en retard). Ce code de dénombrement était réputé avoir un rapport avec le taux de non-réponse observé dans l'EP.

BIBLIOGRAPHIE

BAILAR, B.A. (1985). Comments on "Estimating the population in a Census Year: 1980 and beyond" par E.P. Ericksen et J.B. Kadane, *Journal of the American Statistical Association,* 80, 109-114.

BISHOP, Y.M.M., FIENBERG, S.A., et HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: MIT Press.

FAY, R., PASSEL, J.S., et ROBINSON, J.G. (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.

PURCELL, N.J., et KISH, L. (1979). Estimation for small domains. *Biometrics*, 35, 365-384.

WOLTER, K.M. (1986). Capture-recapture estimation in the presence of known sex ratio. SRD Research Report, United States Bureau of the Census, Washington, D.C.