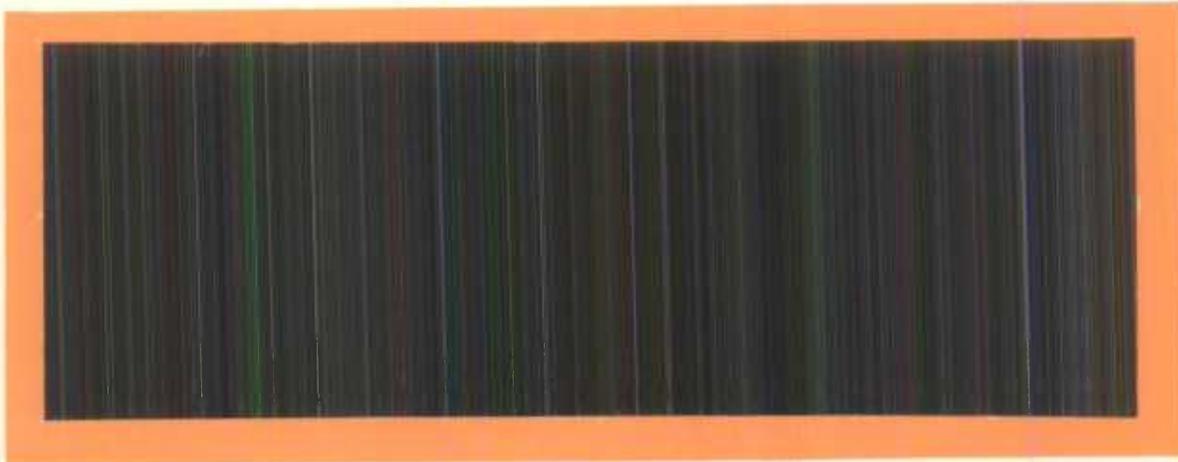




Statistics  
Canada

Statistique  
Canada



## Methodology Branch

Census & Household Survey  
Methods Division

## Direction de la méthodologie

Division des méthodes de recensement  
et d'enquêtes ménages

11-616 F  
85-68  
C2

Canada



# 50362

REMANIEMENT DE L'ENQUÊTE SUR LA POPULATION ACTIVE AU CANADA  
A PARTIR DES RESULTATS DU RECENSEMENT DE 1981



M.P. Singh, J.D. Drew, and G.H. Choudhry  
Division des méthodes de recensement et d'enquêtes-ménages  
Statistique Canada

Numéro: DMREM 85-068F



## Remaniement de l'enquête sur la population active au Canada à partir des résultats du recensement de 1981<sup>1</sup>

M.P. SINGH, J.D. DREW et G.H. CHOUDHRY<sup>2</sup>

### RÉSUMÉ

Après chaque recensement décennal de la population, l'échantillon de l'enquête sur la population active au Canada (EPAC) est remanié pour tenir compte de l'évolution des caractéristiques de la population et répondre aux nouveaux besoins en information. Le dernier programme de remaniement, qui a amené la sélection d'un nouvel échantillon au début de 1985, comportait des recherches poussées sur les moyens d'améliorer le plan de sondage, la collecte des données et les méthodes d'estimation. Les grandes lignes de ce programme sont décrites ici.

**MOTS CLÉS:** Enquête permanente; plan de sondage à plusieurs degrés; stratification; redistribution d'un échantillon; interviews téléphoniques; méthode itérative du quotient.

### 1. INTRODUCTION

L'enquête sur la population active au Canada (EPA), la plus vaste enquête mensuelle sur les ménages menée par Statistique Canada, est ordinairement remaniée après chaque recensement décennal. Dans le cadre du remaniement entamé après le recensement de 1981, un programme intensif de recherche que nous avons décrit dans une autre étude (Singh et Drew 1981a) a été mis en oeuvre pour examiner diverses méthodes d'échantillonnage, d'estimation et de collecte des données. Comme les données sur le marché du travail sont assez fiables au niveau national et provincial, les principaux objectifs fixés pour ce remaniement étaient d'améliorer la fiabilité des données infraprovinciales et d'accroître le rendement général de l'enquête. Pour augmenter le rendement, l'accent a été mis sur les moyens d'automatiser davantage les diverses phases de l'échantillonnage, une utilisation accrue des données du recensement pour mettre à jour l'échantillon au lieu d'informations recueillies indépendamment et un recours accru aux interviews téléphoniques dans le déroulement normal de l'enquête. Pour améliorer les données infraprovinciales, diverses techniques d'échantillonnage et d'estimation ont été examinées, et les résultats ont entraîné des modifications dans les méthodes utilisées auparavant et une redistribution de l'échantillon à l'intérieur des provinces.

Cet exposé résume les résultats des recherches théoriques et empiriques et des expériences sur le terrain entreprises au cours du programme de remaniement. Les sections 2 et 3 décrivent les objectifs du remaniement et l'ancien plan de sondage, après quoi les sections 4, 5 et 6 présentent les conclusions tirées des recherches qui ont justifié les modifications apportées aux méthodes d'échantillonnage et de collecte des données. La section 7 porte sur les problèmes d'estimation et la section 8 sur la redistribution de l'échantillon. Les principales répercussions du remaniement de l'échantillon sur d'autres enquêtes liées à l'EPA sont soulignées à la section 9 et, enfin, la section 10 aborde brièvement les avantages découlant des principales améliorations de l'échantillon remanié et mentionne quelques futurs projets de recherche.

<sup>1</sup> Présenté à une réunion de la Section sur les méthodes d'enquête de l'American Statistical Association à Philadelphie, août 1984.

<sup>2</sup> M.P. Singh, J.D. Drew et G.H. Choudhry, Division des méthodes de recensement et d'enquêtes-ménages, Direction de la méthodologie, Statistique Canada, 4 étage, Immeuble Jean-Talon, Tunney's Pasture, Ottawa (Ontario), Canada K1A 0T6.



## 2. DÉFINITION DES OBJECTIFS

Une des étapes fondamentales du remaniement d'une enquête régulière est la redéfinition des objectifs de l'enquête. Pour l'EPA, il a fallu réévaluer non seulement son rôle de principale source de renseignements à jour sur le marché du travail, mais également son utilisation à l'intérieur de Statistique Canada comme instrument général pour d'autres enquêtes sur les ménages (Singh et Drew 1981b).

Vers le début du programme de remaniement, il a été décidé que celui-ci devrait avant tout viser l'amélioration de l'EPA, mais qu'il était également important d'accroître la flexibilité de cette enquête pour des applications générales. À cet égard, plusieurs changements sont en cours et auront des effets positifs non seulement sur l'EPA, mais aussi sur les autres enquêtes qui y sont liées. Les objectifs concernant la fiabilité des données sur le marché du travail sont décrits plus bas et les modifications qui amélioreront les autres enquêtes sont résumées à la section 9.

Les objectifs de production de données sur le marché du travail ont été établis, lors de conférences annuelles fédérales-provinciales sur la statistique du marché du travail et de suivis bilatéraux avec les points de contact statistiques dans les dix provinces du Canada et avec les principaux ministères fédéraux qui utilisent ces données. En général, ces consultations ont permis de constater une satisfaction à l'égard de la fiabilité des données provinciales et nationales, mais un grand besoin de données infraprovinciales plus précises. Les objectifs suivants ont été fixés pour la production de données fiables à partir de l'échantillon remanié:

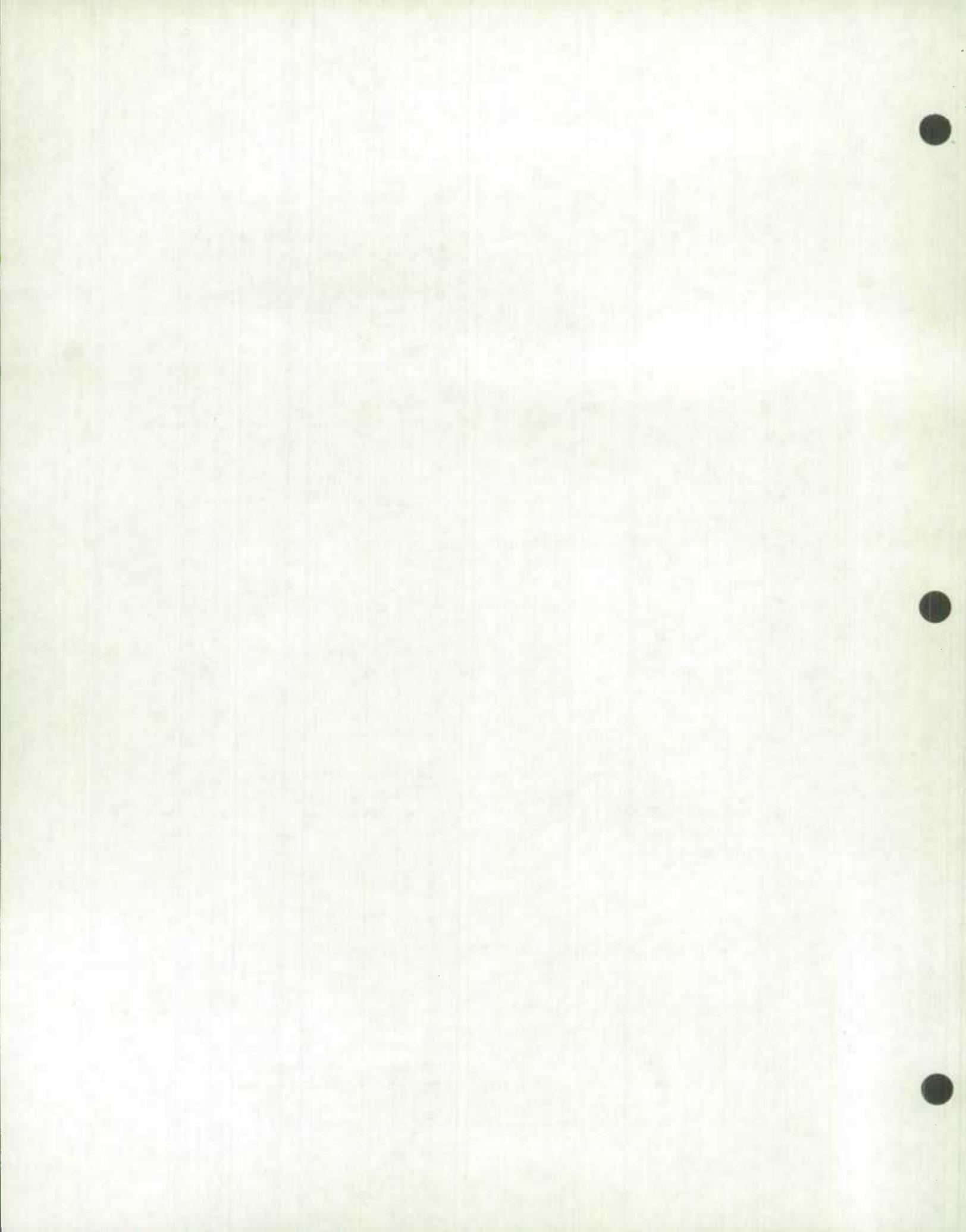
- (i) pour le Canada et chacune des dix provinces, éviter toute réduction de la fiabilité des estimations globales des niveaux et des variations mensuelles de l'emploi et du chômage;
- (ii) pour les 24 régions métropolitaines de recensement définies pour le recensement de 1981, produire des estimations du nombre de chômeurs dans les régions où le coefficient de variation (CV) est inférieur ou égal à 20%;
- (iii) pour les 66 régions économiques infraprovinciales définies en consultation avec les provinces, produire des estimations mensuelles du nombre de chômeurs dans les régions où le CV est inférieur ou égal à 25%;
- (iv) pour les villes de 60,000 habitants ou plus au Québec et en Ontario et celles de 25,000 habitants ou plus dans les autres provinces, produire des estimations trimestrielles du nombre de chômeurs dans les villes où le CV est inférieur ou égal à 25%;

Pour atteindre ces objectifs, il a fallu redistribuer une partie de l'échantillon des grandes villes et régions économiques dans certaines villes et régions économiques plus petites. Comme on voulait en plus réduire le coût de l'EPA (section 8), beaucoup d'espoirs reposaient sur les projets de recherche concernant les moyens d'accroître l'efficacité des opérations de collecte des données et de production des statistiques de l'EPA. Dans les sections suivantes, ces problèmes sont examinés du point de vue des régions autoreprésentatives (AR) et des régions non autoreprésentatives (NAR).

## 3. DESCRIPTION DE L'ANCIEN PLAN DE SONDAGE DE L'EPA

Une description complète de l'ancien plan de sondage de l'EPA a été présentée par Platek et Singh (1976). Cette section résume les principaux aspects de ce plan de sondage afin de rendre plus clairs les sujets traités dans les sections suivantes.

Les unités autoreprésentatives (UAR) de l'ancien plan de sondage correspondaient aux villes qui, au moment de la conception du plan de l'enquête, étaient assez grandes pour avoir comme rendement prévu un échantillon de 20 ménages, soit le nombre minimal de ménages pouvant être confié à un interviewer. La limite inférieure de la taille des UAR variait de 10,000 personnes dans la région de l'Atlantique à 25,000 personnes au Québec et en Ontario.



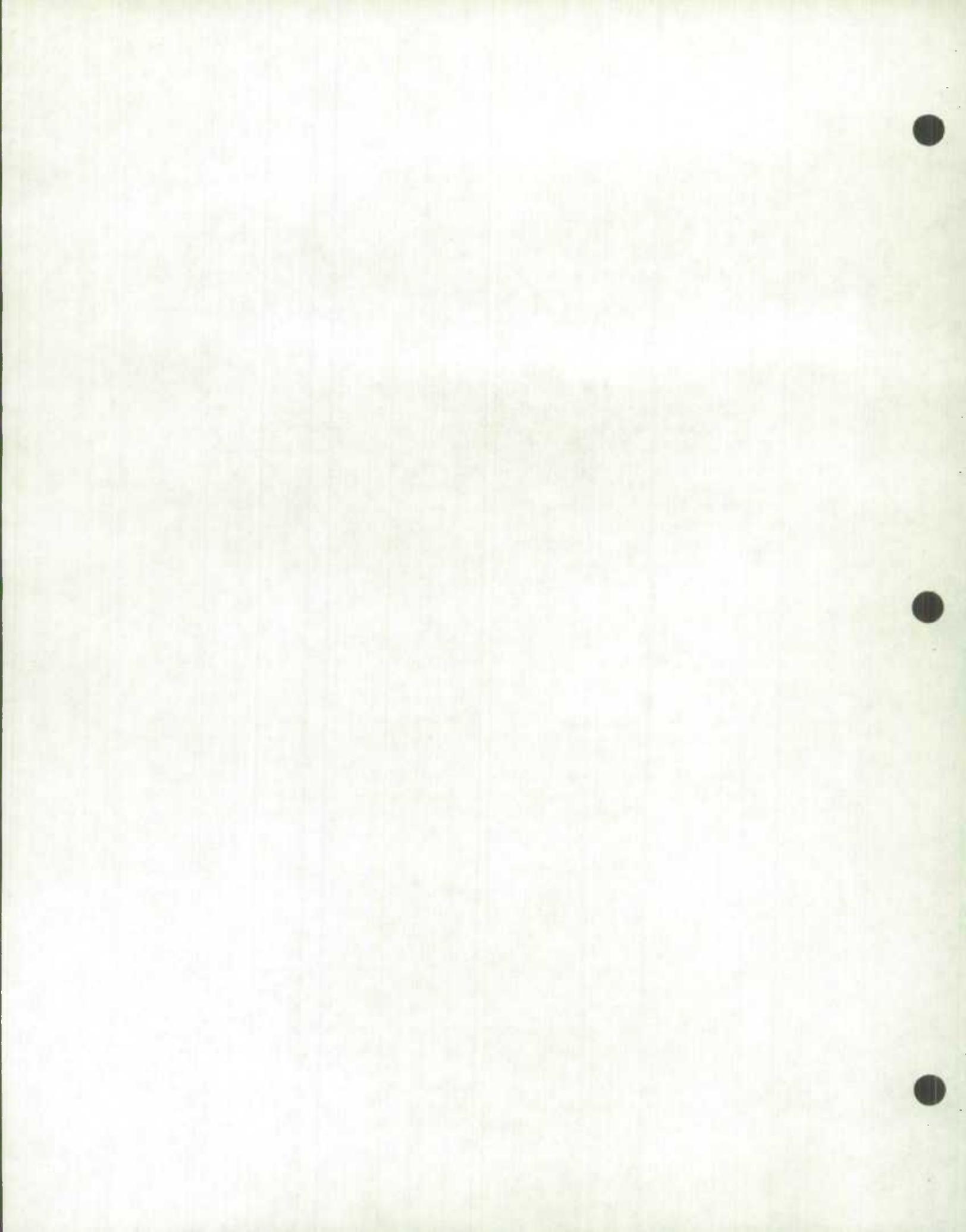
Les grandes UAR étaient stratifiées en profondeur par regroupement de secteurs de recensement contigus - régions géostatistiques comprenant de 3,000 à 5,000 habitants et dont la stabilité d'un recensement à l'autre en fait des unités opérationnelles pratiques - sans tenir compte de l'optimalité de leurs caractéristiques. Des unités primaires d'échantillonnage, désignées par le terme "grappes" et composées essentiellement d'un îlot urbain, ont été délimitées à partir de comptes relevés sur place en 1973. Un échantillon à deux degrés de grappes et de ménages était prélevé selon la méthode de Rao, Hartley et Cochran (1962) pour la sélection de groupes aléatoires avec probabilité proportionnelle à la taille (ppt). En plus de la base de sondage aréolaire, une base ouverte était tenue à jour pour les immeubles d'appartements dans les grandes villes.

La méthode de sélection de la base aréolaire comporte cet avantage particulier que sa souplesse permet de modifier la taille de l'échantillon (Singh et Drew 1977) et en facilite la mise à jour (Platek et Singh 1977, Drew, Choudhry et Gray 1978). L'échantillon des UAR doit être mis à jour périodiquement parce que les comptes utilisés pour la sélection avec ppt deviennent progressivement moins fiables au fil du temps, ce qui accroît la variance d'échantillonnage des estimations de l'enquête. Comme l'échantillonnage se fait de façon indépendante dans chaque groupe aléatoire, l'échantillon peut être mis à jour à l'aide de la méthode de Keyfitz (1957), laquelle permet de réviser les probabilités de sélection à partir de valeurs récentes des comptes de logements, de conserver autant d'unités déjà sélectionnées que possible et d'éviter d'inclure dans l'échantillon mis à jour les logements sélectionnés pour l'ancien échantillon. De 1978 jusqu'au début de la période de remaniement en 1982, des mises à jour régulières ont été faites pour les régions AR en croissance rapide, et presque la moitié de la base de sondage a été modifiée. La fréquence des mises à jour n'a pas été suffisante pour réduire les effets du *plap* de sondage aux niveaux mesurés au cours des 4 premières années de l'enquête, mais il a été possible d'empêcher de nouvelles détériorations de ces effets qui augmentaient en moyenne de 7% à 8% par année pour les estimations du nombre de chômeurs.

Les unités non autoreprésentatives (UNAR) sont des régions à l'extérieur des UAR et comprennent des régions rurales et de petits centres urbains. À l'intérieur des régions NAR, les régions économiques étaient stratifiées par catégorie d'activité économique de manière à ce que toutes les strates forment des zones de terrain contiguës. Pour chaque strate, on procédait à une sélection d'unités primaires d'échantillonnage (UPÉ) aussi représentatives de leur strate que possible en ce qui a trait au rapport population rurale/population urbaine et à diverses caractéristiques importantes de la population active. La partie rurale d'une UPÉ était constituée de secteurs de dénombrement (SD) ruraux contigus, mais la partie urbaine n'était pas toujours contiguë à cette zone rurale parce qu'il fallait respecter le rapport population rurale/population urbaine. Les plus grands centres urbains dans ces strates étaient souvent inclus dans plusieurs UPÉ à la fois.

Au remaniement de 1973, deux UPÉ ont été sélectionnées dans chaque strate selon la méthode d'échantillonnage systématique avec classement aléatoire et probabilité proportionnelle à la taille (Hartley et Rao 1962). En 1977, la taille de l'échantillon de l'EPA est passée de 33,000 ménages à 55,000 ménages par mois, et cet échantillon agrandi a été réparti de manière à améliorer la fiabilité des données provinciales. Ainsi, la taille de l'échantillon affichait une hausse proportionnelle plus marquée dans les petites provinces. L'échantillon des régions NAR a été élargi par l'addition de 1 à 4 UPÉ supplémentaires à chaque strate (Gray 1975).

Les parties rurales et urbaines des UPÉ sélectionnées étaient échantillonnées indépendamment. Dans les parties urbaines de ces UPÉ, un échantillon à deux degrés composé de grappes et de logements était prélevé, tandis qu'un échantillon à trois degrés composé d'unités secondaires d'échantillonnage (SD du recensement ou combinaisons de SD), de grappes (zones de terrain faciles à délimiter comptant au plus 20 logements) et de logements. Sauf pour la dernière étape de l'échantillonnage, la sélection des unités était fondée sur la méthode d'échantillonnage systématique avec classement aléatoire et ppt.



#### 4. REMANIEMENT DU PLAN DE SONDAGE DES RÉGIONS AUTOREPRÉSENTATIVES

La taille minimale de l'échantillon des villes classées comme UAR dans le nouveau plan de sondage a été haussée à 50 logements parce qu'une analyse a révélé que les coûts unitaires sont très élevés quand les échantillons des régions AR sont petits. La composition de l'univers des régions AR a toutefois peu changé à cause des effets compensateurs de l'élargissement de l'échantillon de l'EPA de 33,000 à 55,000 ménages à la fin des années 1970 et à cause de la nouvelle répartition de l'échantillon remanié.

Pour les raisons mentionnées plus haut, le plan de sondage des régions AR est resté essentiellement le même, et le principal objectif visé pour ces secteurs était de mettre à jour les mesures de leur taille sans engager les coûts élevés d'un dénombrement indépendant sur le terrain comme celui entrepris au dernier remaniement. Les données du recensement de 1981 ont été utilisées pour cette mise à jour, et deux méthodes différentes ont été appliquées aux villes décomposables en côtés d'îlots (grandes villes pour lesquelles on possédait des données agrégées au niveau des côtés d'îlots) et les villes non décomposables en côtés d'îlots. Les techniques de mise à jour et les unités d'échantillonnage choisies pour ces deux méthodes sont décrites plus bas. Le procédé de stratification à deux niveaux adopté pour les régions AR est expliqué à la section 5.

##### Villes décomposables en côtés d'îlots

C'est l'existence de données de recensement agrégées au niveau des côtés d'îlots dans les zones bâties des villes les plus grandes qui a été le facteur clé dans la décision de remanier au complet l'échantillon de ces villes, lesquelles forment les 2/3 de la base des régions AR. Ce remaniement a également permis d'améliorer le schéma de stratification à l'aide du procédé décrit à la section suivante.

Pour les parties de ces villes qui sont décomposables en côtés d'îlots, les îlots de recensement ont été choisis comme grappes (c'est-à-dire comme UPÉ). On a examiné les composantes de la variance dans un plan de sondage à deux degrés fondé sur la méthode des groupes aléatoires de Rao, Hartley et Cochran pour deux types de grappes—les îlots de recensement et les SD—en simulant le plan de sondage de l'EPA à partir des données du recensement de 1976 pour les UAR d'Halifax et de Saskatoon (Choudhry, Drew et Lee 1984). Les résultats de cette étude ont montré que, dans les cas où les mesures de la taille des UAR sont à jour, il y avait peu de différence entre les variances d'échantillonnage obtenues pour les SD et les îlots. Le choix des îlots comme grappes était donc justifié du point de vue opérationnel. Les îlots offraient une base de sondage presque toute faite (il a fallu diviser ou combiner des unités dans une proportion de cas variant de 5 à 10%), ce qui permettait d'élaborer un plan de sondage fortement automatisé et économique à remanier. Autre fait important, les données des recensements futurs peuvent être relevées pour les îlots stables sur le plan géostatistique (mais non pour les SD qui, en tant qu'unités opérationnelles, changent à chaque recensement), ce qui rend possible une mise à jour peu dispendieuse de l'échantillon tous les cinq ans. Les zones bâties comprennent 86% de ces villes.

Les SD ont été choisis comme unités d'échantillonnage dans les parties de ces villes pour lesquelles on ne disposait pas de données agrégées au niveau des côtés d'îlots. La variance a été calculée seulement pour le cas où les mesures de la taille des régions sont à jour. On croyait que, puisque les SD sont des unités plus grandes que les îlots, elles seraient moins touchées par la croissance très concentrée qui peut avoir lieu à l'extérieur des zones bâties d'une ville. Le choix des SD dans ces régions permettait également de réduire beaucoup les coûts du remaniement parce qu'environ 80% des régions auxquelles cette règle a été appliquée sont situées au Québec et en Ontario où, comme les fractions de sondage sont faibles, peu d'unités ont dû être divisées.

L'étude de la variance et l'étude du temps et des coûts (Lemaître 1983) ont révélé que la variance par unité de coût est à peu près constante pour les grappes où on sélectionne de 2 à 8 logements. On a donc décidé de retenir le facteur de densité de 4 ou 5 logements dans chaque grappe, comme dans l'ancien plan de sondage, pour les strates où les grappes sont des îlots, mais de choisir plutôt de 6 à 8 logements dans les SD à cause de leur taille plus grande.



### Villes non décomposables en côtés d'îlots

Comme plus de 70% des UAR non décomposables en côtés d'îlots étaient nouvelles ou n'avaient plus les mêmes limites qu'auparavant et que la plupart des autres UAR n'avaient pas été mises à jour depuis le remaniement de 1973, on a décidé de remanier au complet l'échantillon de ces villes. La grappe définie pour ces UAR était un îlot de recensement ou une combinaison d'îlots dans les zones bâties de ces villes et les chiffres sur les nombres de logements étaient tirés directement des registres des visites et des cartes géographiques utilisés lors du recensement. À l'extérieur des zones bâties, des SD complets ou divisés ont été choisis comme grappes et des comptes relevés sur le terrain ont parfois été utilisés pour diviser des SD.

L'utilisation des registres des visites du recensement a coûté plus que le système employé pour les villes décomposables en côtés d'îlots, mais nettement moins que les comptages indépendants sur le terrain qui étaient nécessaires à l'ancien plan de sondage.

## 5. STRATIFICATION

### 5.1 Algorithme et variables de stratification

Une version modifiée d'un algorithme non hiérarchique élaboré par Friedman et Rubin (1967) a été adoptée pour la stratification des régions AR et NAR. Ce choix repose sur les résultats d'études faites par Judkins et Singh (1981) et Kostanich, Judkins, Singh et Schantz (1981), qui ont évalué plusieurs algorithmes de stratification pour l'enquête sur la Current Population Survey du U.S. Bureau of the Census. Les modifications apportées à cet algorithme permettent de former des strates géographiquement contiguës et/ou compactes et offre la possibilité de construire soit des grappes homogènes (c'est-à-dire des strates), soit des grappes hétérogènes (c'est-à-dire des unités primaires d'échantillonnage à l'intérieur des strates NAR). Une description détaillée de cette méthode et des résultats d'évaluations empiriques a été présentée par Foy, Bélanger, Drew et Joncas (1984). Cette section en donne un résumé.

L'algorithme répartit d'abord les unités au hasard en un nombre déterminé de strates. À chaque itération, toutes les unités de stratification sont examinées et ajoutées à une strate quelconque de manière à minimiser la somme pondérée des carrés de plusieurs variables à l'intérieur de la strate sans dépasser les contraintes relatives à la taille des strates. L'algorithme atteint un optimum local lorsque l'addition d'une unité augmente cette somme des carrés à l'intérieur d'une strate. Une étude de Judkins et Singh (1981) a révélé que les optimums locaux sont meilleurs quand le nombre d'origines aléatoires est assez élevé (c'est-à-dire 30).

Pour l'option de contiguïté, l'algorithme lit une matrice qui précise, pour chaque unité, toutes les autres qui y sont contiguës. Les premières strates contiguës qui satisfont aux contraintes relatives à leur taille sont construites à partir des unités choisies au hasard comme points de départ. Pendant l'optimisation, le transfert des unités se fait sous une nouvelle condition selon laquelle la contiguïté doit être maintenue. Pour que les strates soient compactes, les centroïdes de la population (longitude et latitude) sont ajoutés comme variables dans la somme pondérée des carrés qui doit être minimisée.

La stratification des régions NAR et AR repose sur les données du recensement de 1981 et utilise jusqu'à 17 variables de stratification. Parmi les variables démographiques utilisées, il y a le nombre total de personnes occupées, le revenu d'emploi, le nombre de personnes qui ont terminé leurs études secondaires, la population âgée de 15 ans et plus, de 15 à 24 ans et de 55 ans et plus, et le nombre d'actifs dans les domaines suivants: agriculture, exploitation forestière et pêche, industries minières, industries manufacturières, construction, transports et services. L'algorithme utilise aussi des variables relatives au logement comme, par exemple, le nombre total de logements, le nombre de logements loués, le nombre de ménages composés d'une seule personne et le nombre de ménages composés de deux personnes.



Toutes les variables correspondant à des branches d'activité comprenant moins de 2% de la population active d'une région en train d'être stratifiée ont été supprimées, et le poids des autres variables est modifié pour qu'elles aient une importance égale dans le processus d'optimisation. Le nombre de chômeurs n'a pas été inclus parmi les variables de stratification à cause de son instabilité. Une étude a démontré que, si on les évalue après le recensement subséquent, les strates définies sans tenir compte du nombre de chômeurs sont plus efficaces non seulement pour l'estimation d'autres caractéristiques, mais également pour le calcul du nombre de chômeurs. En revanche, l'inclusion de variables sur les différents quartiers a permis d'accroître l'efficacité des estimations du nombre de chômeurs.

## 5.2 Stratification, à deux niveaux dans les régions AR

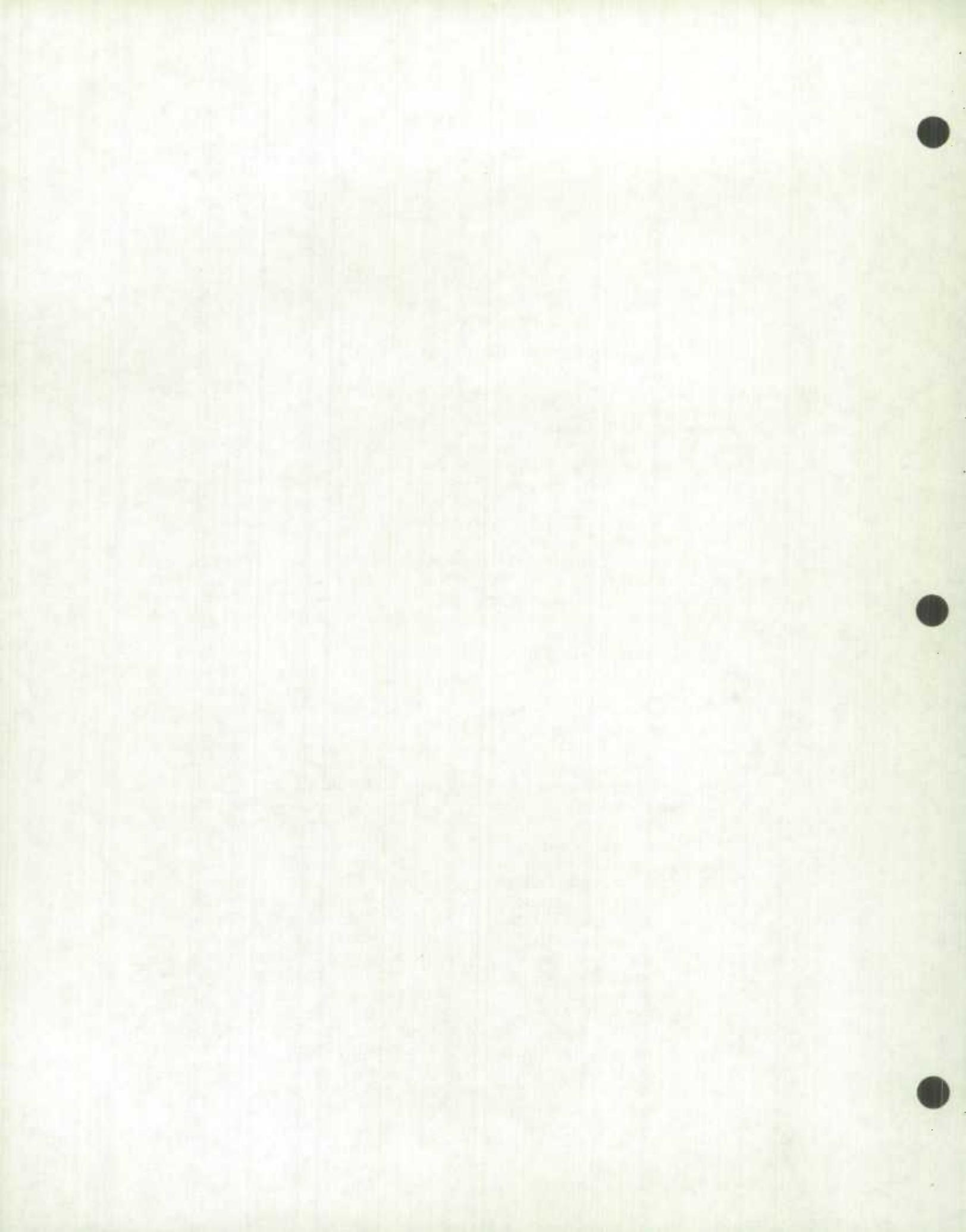
Dans les grandes UAR dont l'échantillon englobe 300 ménages ou plus, deux niveaux de stratification ont été définis. Les strates primaires, dont l'échantillon aréolaire et l'échantillon d'immeubles d'appartements comprennent de 150 à 170 ménages en tout, sont des ensembles de secteurs de recensement contigus. Les strates primaires sont définies de manière à correspondre à deux tâches d'interviewer. Trois ou quatre strates secondaires aréolaires non géographiques comprenant chacune six ou un multiple de six grappes échantillonnées sont créées dans chaque strate primaire; les secteurs de recensement servent d'unités de stratification et l'optimisation est fondée sur les données du recensement de 1981 relatives aux caractéristiques des personnes n'habitant pas un immeuble d'appartements.

L'échantillon d'appartements est prélevé séparément par échantillonnage systématique avec ppt à partir d'une base de sondage ouverte et comprend généralement une seule strate pour l'ensemble d'une UAR. Un tri des appartements existant au moment de la conception du plan de sondage, en fonction des strates primaires, a produit une stratification géographique implicite pour l'échantillon d'immeubles d'appartements.

Dans les petites UAR décomposables en côtés d'ilots où ni un échantillon séparé d'immeubles d'appartements, ni des strates géographiques primaires n'étaient justifiés, des strates aréolaires non géographiques optimales ont été construites directement. Dans les villes pour lesquelles on ne disposait pas de données agrégées au niveau des côtés d'ilots, la stratification était beaucoup moins facile; on a décidé de former des strates géographiques simples.

Les deux niveaux de stratification définis pour les grandes UAR semblaient appropriés des points de vue opérationnel et technique. Comme les contraintes géographiques sont moins strictes que celles de l'ancien plan de sondage, on a pu accroître l'optimalité des résultats; le fait de conserver la contrainte de contiguïté aux niveaux élevés permettra d'avoir une unité convenable pour la mise à jour de l'échantillon, plus tard au cours des années 1980, et facilitera la planification des tâches des interviewers. Par ailleurs, dans l'ancien plan de sondage, les strates AR étaient généralement confiées à un seul interviewer; les estimations de la variance ne tenaient donc pas compte de la composante de la variance totale attribuable à la variance de réponse corrélée. Comme les tâches d'interviewer sont généralement divisées géographiquement et que les strates secondaires sont non géographiques, cela permet d'avoir une superposition des strates et des tâches d'interviewer dans le nouveau plan de sondage, sans augmenter les coûts de la collecte des données, et de mieux prendre en considération la variance de réponse corrélative dans les estimations de la variance.

Le tableau 1 présente les résultats d'une étude de variance pour deux UAR— Ottawa et la ville de Québec—et permet de comparer l'efficacité des strates géographiques utilisées dans l'ancien plan de sondage avec celle des strates optimales à deux niveaux pour une expérience faite à partir des données du recensement de 1971. Les chiffres calculés au moment du recensement de 1981 relativement au pourcentage de réduction de la variance du premier degré attribuable à la stratification indiquent que les plus grands changements découlent de la stratification optimale pour le revenu et les logements loués. Le fait que l'amélioration ne soit que légère pour d'autres caractéristiques telles que le nombre de personnes occupées et de chômeurs confirme la force et la robustesse de la stratification simple, mais profonde, de l'ancien plan de sondage.



**Tableau 1**  
 Pourcentage de réduction de la variance du  
 premier degré attribuable à la stratification  
 Comparaison de l'ancienne et de la nouvelle méthode

Variable	Méthode de stratification		Variable	Méthode de stratification	
	ancienne	nouvelle		ancienne	nouvelle
total, personnes occupées	9.1	12.1	agriculture <sup>1</sup>	5.9	3.9
revenu d'emploi	18.1	30.4	forêts/pêche <sup>1</sup>	3.1	2.4
personnes ayant terminé leurs études secondaires	39.4	42.1	industries minières <sup>1</sup>	4.8	3.0
population de 15 ans et plus	9.2	12.6	industries manufacturières	23.5	23.1
population de 15 à 24 ans	12.9	17.6	construction	11.9	11.2
population de 55 ans et plus	25.3	29.7	transports	4.2	6.4
total, logements	28.5	33.1	services	14.5	19.8
logements loués	20.9	28.8	nombre de chômeurs <sup>1</sup>	7.1	9.7
ménages d'une seule personne	33.7	38.4			
ménages de deux personnes	27.5	29.6			

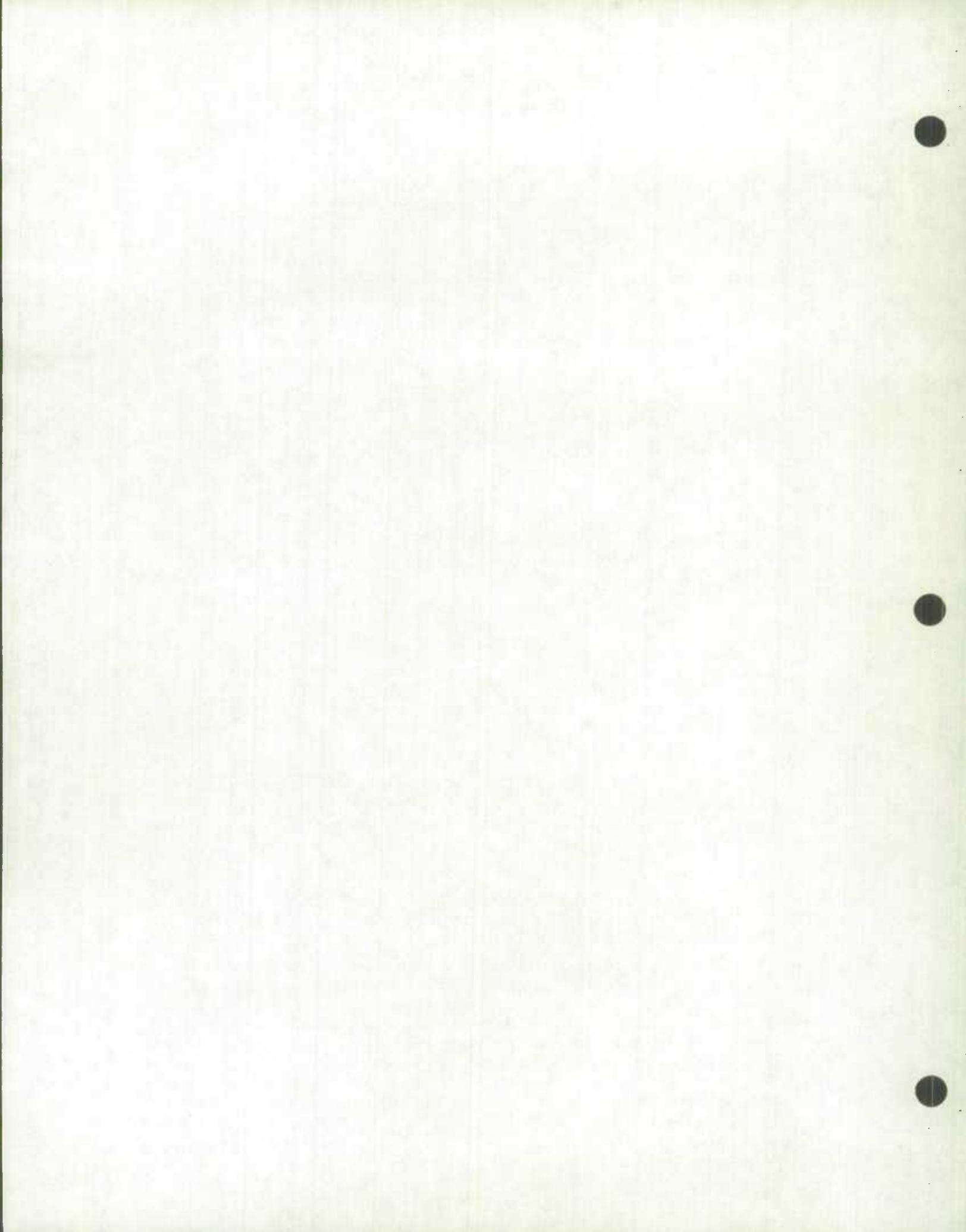
<sup>1</sup> caractéristiques non utilisées pour l'optimisation, dans la nouvelle méthode.

Dans les régions NAR, le même algorithme de classification a été appliqué à chaque région économique pour définir des strates rurales ou des strates mixtes urbaines et rurales (voir section 6.3). L'utilisation de l'algorithme de classification pour la formation des UPÉ est examinée à la section 6.5.

## 6. ASPECTS DU PLAN DE SONDAGE DES RÉGIONS NAR

### 6.1 Extension des interviews téléphoniques

Depuis le début des années 1970, la collecte des données pour la période du deuxième au sixième mois d'inclusion d'un ménage dans l'échantillon se fait par interview téléphonique, dans les régions autoreprésentatives, essentiellement pour réduire les coûts. Par contre, dans les régions NAR, les interviews sont encore effectuées par des interviewers à cause des risques que le grand nombre de lignes partagées représente pour le caractère confidentiel des renseignements demandés. Toutefois, comme les interviews téléphoniques permettent de réaliser des économies immédiates et de tirer des avantages à long terme de techniques telles que la composition de numéros au hasard et les interviews téléphoniques assistées par ordinateur, on a décidé d'examiner la possibilité d'étendre les interviews téléphoniques aux régions NAR.



On a procédé à un premier essai sur le terrain visant uniquement les régions urbaines où plus de 80% des lignes sont privées. Ce test a été appliqué à une partie de l'échantillon réel de l'EPA pour évaluer l'effet des interviews téléphoniques sur la qualité des données. Pour faciliter cette analyse, les tâches des interviewers ont été divisées en une partie qui devait être faite par téléphone et une autre qui devait être faite sur place.

Cette expérience a eu lieu de janvier 1982 à juin 1983 et a été réalisée progressivement de façon à ne pas nuire au déroulement de l'enquête. Les principaux résultats démontrent que les taux de non-réponse sont moins élevés avec les interviews téléphoniques (3.4% contre 4.3% dans l'échantillon témoin), qu'une proportion élevée de ménages ont un téléphone (96% dans toutes les provinces sauf une), que peu de ménages (1%) refusent d'être interviewés par téléphone et que les interviews téléphoniques ne produisent aucune différence notable entre les estimations des caractéristiques de la population active.

Un deuxième essai dans les régions rurales a abouti à des conclusions semblables. Comme les résultats de ces deux essais étaient positifs, il a été décidé de mener des interviews téléphoniques dans l'ensemble des régions NAR pour le reste de l'année 1983 et le début de 1984.

Les principales conséquences de cette décision sur le plan de sondage de l'échantillon des régions NAR ont été les suivantes:

- (i) Augmentation du volume des tâches des interviewers: Dans l'ancien plan de sondage, les tâches dans les régions NAR comprenaient en moyenne 50 logements. Comme les coûts unitaires sont faibles dans les tâches qui englobent beaucoup de ménages (Lemaître 1984) et que l'utilisation du téléphone réduit les déplacements, il a été possible de choisir de 55 à 60 ménages pour l'échantillon de chaque UPÉ des régions NAR.
- (ii) Attribution du même numéro de renouvellement au deuxième niveau: Dans le nouveau plan de sondage, contrairement à l'ancien plan, le même numéro de renouvellement sera attribué à tous les logements dans les unités secondaires, ce qui permettra de réduire le nombre de visites dans les unités secondaires, du deuxième au sixième mois d'inclusion d'un ménage dans l'échantillon.

## 6.2 Élimination d'un degré d'échantillonnage dans les régions rurales

Dans l'ancien plan de sondage, la sélection de l'échantillon rural à l'intérieur des UPÉ se faisait en trois étapes: unités secondaires (secteurs de dénombrement du recensement), grappes et logements. Les grappes correspondaient à des zones de terrain faciles à délimiter, qui comprenaient jusqu'à 20 logements et étaient définies à partir de chiffres relevés sur le terrain quand une nouvelle unité secondaire était ajoutée à l'échantillon. À l'intérieur des unités secondaires, on choisissait généralement 5 ou 6 grappes composées de 3 ou 4 logements.

Dès le début du programme de remaniement, on a songé à supprimer la sélection de grappes dans les régions rurales pour les raisons suivantes: (i) la variance d'échantillonnage diminuerait parce qu'il y aurait un degré de moins dans le plan de sondage et (ii) le délai de préavis nécessaire pour inclure de nouvelles unités secondaires dans l'échantillon diminuerait de 13 mois à 7 mois.

Une étude a été menée sur le terrain auprès d'un échantillon d'unités secondaires, au moment de leur inclusion dans l'échantillon de l'EPA, pour évaluer la possibilité de tenir à jour de bonnes listes des logements dans les SD ruraux et pour examiner les coûts d'un tel programme. Les résultats obtenus pour ces deux questions ont été positifs. L'effet que l'élimination de cette étape de l'échantillonnage pourrait avoir sur la variance a également été étudié. L'ancien plan de sondage et le plan de sondage modifié ont été simulés à l'aide des données du recensement de 1971 et les composantes de la variance ont été mesurées pour l'estimateur de Horwitz-Thompson sans estimation par le quotient. Le plan de sondage modifié a permis de réduire la variance totale de 20% à 25% pour les principales caractéristiques de la population active (Choudhry, Lee et Drew 1984).

Compte tenu de ces résultats, on a décidé, au début du programme de remaniement, d'éliminer l'étape de la sélection des grappes dans les régions rurales et d'examiner d'autres aspects plus généraux du plan de sondage.



### 6.3 Plan de sondage avec stratification urbaine et rurale

L'ancien plan de sondage renfermait une stratification urbaine et rurale implicite. Les UPÉ étaient définies de manière à ce que le rapport entre la population urbaine et rurale soit à peu près le même que celui de la strate et, à l'intérieur des UPÉ sélectionnées, les parties urbaines et rurales étaient échantillonnées indépendamment. Théoriquement, chaque UPÉ devait correspondre à la tâche d'un interviewer. Toutefois, en pratique, cette correspondance était floue parce que, comme il fallait atteindre le rapport voulu entre la population urbaine et rurale, il arrivait souvent que les parties urbaines et rurales des UPÉ n'étaient pas contiguës.

L'ancien plan de sondage,  $P_0$  (dans lequel les grappes rurales étaient supprimées), a été comparé avec un autre plan de sondage,  $P_1$ , qui renfermait une stratification urbaine et rurale explicite. Comme  $P_0$ ,  $P_1$  comprenait 3 étapes d'échantillonnage, tant dans les régions urbaines que dans les régions rurales. Les strates urbaines étaient ventilées en UPÉ (centres urbains distincts ou voisins), en grappes et en logements; les strates rurales étaient classées en UPÉ (groupes de SD ruraux voisins), en unités secondaires d'échantillonnage (SD) et en logements. Dans  $P_1$ , les UPÉ urbaines et les UPÉ rurales étaient définies indépendamment de manière à produire des échantillons correspondant à des tâches d'interviewer.

Ces deux plans ont été évalués en fonction de la variance d'échantillonnage et du coût d'enquête (Choudhry, Lee et Drew 1984). Dans l'étude de la variance, ces deux plans ont été simulés pour des strates comprenant 2 UPÉ à l'aide de chiffres repères du recensement de 1971 et les variables étudiées étaient tirées des données du recensement de 1976.

Pour comparer les coûts de ces plans de sondage, un modèle simple a été élaboré pour l'exploitation de  $P_0$ , l'ancien plan de sondage, par des interviews téléphoniques, et les différents coûts ont été estimés à partir des résultats d'une étude détaillée des temps et des coûts (Lemaître 1983). Pour estimer le coût relatif des déplacements nécessités par les plans  $P_0$  et  $P_1$ , on a fait une étude de simulation dans laquelle la dispersion moyenne de l'échantillon correspondant à ces deux plans a été mesurée jusqu'au deuxième degré du plan de sondage à partir des centroïdes de la population des SD.

Il s'est avéré que le rendement du plan  $P_1$  était 1.09 fois celui de  $P_0$  et que, sur le plan des coûts d'enquête et de la variance d'échantillonnage,  $P_1$  était supérieur à  $P_0$ . En outre, l'efficacité relative de  $P_1$  par rapport à  $P_0$  était de 1.25 pour le nombre de personnes occupées et de 1.05 pour le nombre de chômeurs.

Compte tenu de ces résultats, le plan  $P_1$  a été appliqué dans 70% des régions économiques où la population urbaine et rurale était suffisante pour former des strates séparées. Dans les autres régions économiques, sauf à l'Île-du-Prince-Édouard, le plan  $P_0$  a été adopté (voir section 6.6).

### 6.4 Nombre d'UPÉ choisies à l'intérieur de chaque strate

Dans le plan de sondage de l'EPA, comme la taille de l'échantillon de chaque UPÉ est fixe, le nombre d'UPÉ choisies à l'intérieur de chaque strate détermine également le nombre de strates. Dans plus de deux tiers des cas, les strates urbaines, rurales ou combinées à l'intérieur des RÉ ne produisaient que l'équivalent de l'échantillon de 2 ou 3 UPÉ. On a exclu la possibilité de stratifier davantage ces régions parce qu'il devrait y avoir au moins 2 UPÉ par strate pour permettre l'estimation sans biais de la variance.

Les autres RÉ ont été stratifiées de manière à obtenir 2 ou 3 UPÉ par strate. La réduction estimée de la variance du premier degré par rapport à celle de l'ancien plan de sondage, où il y avait de 3 à 6 UPÉ par strate, atteint 14% pour le nombre de personnes occupées (Choudhry, Lee et Drew 1984). La stratification repose sur l'algorithme de classification décrit à la section 5.

### 6.5 Utilisation de l'algorithme de classification pour la formation des UPÉ des régions NAR

Dans l'ancien et le nouveau plan de sondage de l'EPA, la stratification a lieu avant la formation des unités primaires d'échantillonnage des régions NAR. Les UPÉ sont délimitées de manière à ressembler autant que possible à l'ensemble de leur strate selon les variables de stratification et à être aussi compactes que possible sur le plan géographique. Les UPÉ sont délimitées à l'aide de l'algorithme de classification résumé plus haut et reposent sur une minimisation dans le cas des variables géographiques et une maximisation dans le cas des variables non géographiques.



### 6.6 Échantillonnage à deux degrés pour l'Île-du-Prince-Édouard

Pour la plus petite province du Canada, l'Île-du-Prince-Édouard, les fractions de sondage ont été fixées à 4% pour que les estimations mensuelles sur la population active aient le degré de fiabilité nécessaire. Comme ces taux de sondage sont élevés, on a adopté un plan de sondage moins complexe fondé sur un échantillon à deux degrés composé de SD et de logments et stratifié en profondeur. On a constaté que ce plan de sondage est un peu plus coûteux que  $P_0$ , mais, en général, si on tient également compte de la variance, il est bien supérieur à  $P_0$ , l'efficacité relative des estimations du nombre de personnes occupées et de chômeurs étant de 2.21 et 1.11 respectivement par rapport à celles de  $P_0$  (Choudhry, Lee et Drew 1984).

## 7. ESTIMATION

### 7.1 Calcul d'estimations par la méthode du quotient pour le degré final

Dans l'ancienne version de l'EPA, des estimations détaillées pour les diverses tranches d'âge et sexe dans chaque province étaient calculées par la méthode du quotient pour le degré final. Depuis que Statistique Canada a élaboré des méthodes permettant de produire des estimations démographiques infraprovinciales de meilleure qualité et dans des délais plus courts (Verma, Basavarajappa et Bender 1983), une étude a été entreprise sur une étape intermédiaire d'estimation par le quotient dans laquelle des estimations d'enquête sur la population âgée de 15 ans et plus sont corrigées par la méthode du quotient à partir d'estimations externes, avant le calcul des estimations finales par la méthode du quotient. On a constaté que cette étape intermédiaire n'avait pas d'effet sur les variances des données relatives aux provinces, mais permettait de réduire la variance des estimations relatives aux régions infraprovinciales de près de 20% pour le nombre de personnes occupées et de 7% pour le nombre de chômeurs (Earwaker et Bélanger 1981). En pratique, une technique d'estimation itérative par la méthode du quotient, où deux étapes d'estimation par le quotient sont répétées jusqu'à ce que les deux conditions de contrôle soient satisfaites, a été appliquée au début de 1983.

### 7.2 Amélioration des estimations relatives aux ménages et aux familles

Paul et Lawes (1982) ont utilisé les fichiers de données longitudinales de l'EPA, qui englobent les ménages pour les six mois de leur inclusion dans l'échantillon, pour démontrer que les taux de non-réponse les plus élevés sont observés chez les ménages composés d'un petit nombre de membres. Dans l'ancienne EPA, la non-réponse était compensée par une nouvelle pondération des données sur des régions. Cette correction était faite sans tenir compte de la taille des ménages, de sorte que les estimations de la taille des ménages et des familles renfermaient des biais de 1% à 3%. Il y avait également un autre problème qui entraînait un manque de convergence dans les statistiques sur les familles et les personnes (Macredie 1983). La Division de la démographie de Statistique Canada est actuellement en train d'établir des estimations démographiques de la taille des familles et, une fois que ces résultats seront connus, ils seront intégrés au calcul des estimations par la méthode itérative du quotient pour le degré final, afin de résoudre ces deux problèmes.

Provisoirement, une étude a été entreprise pour estimer les distributions de la taille des ménages répondants et non répondants à l'aide des données longitudinales de l'EPA, avant le calcul de l'estimation finale par la méthode itérative du quotient (Ghangurde 1984).

### 7.3 Estimations pour les petites régions

Les demandes d'estimations de la population active dans les petites régions (domaines) telles que les circonscriptions électorales fédérales (CÉF) et les divisions de recensement (DR), qui représentent chacune plus de 250 unités d'un bout à l'autre du Canada, ont augmenté au cours des dernières années. Comme il n'a pas été possible d'utiliser exactement les mêmes limites que celles de ces régions dans le plan de l'enquête, diverses autres techniques d'estimation pour les petites



régions ont été évaluées. Une des solutions proposées était un estimateur dépendant de l'échantillon, qui combine un estimateur pour domaines stratifiés a posteriori et un estimateur synthétique. Cet estimateur devient exclusivement un estimateur pour domaines stratifiés a posteriori, si la taille de l'échantillon prélevé dans un domaine est suffisante selon certains critères, mais ajoute une composante synthétique dont le poids relatif dépend des lacunes de l'échantillon qui représente un domaine. Selon les résultats d'une étude, il a été recommandé qu'un estimateur dépendant de l'échantillon soit élaboré pour la production d'estimations moyennes annuelles ou pluriannuelles relatives à des régions telles que les CÉF et les DR (Drew, Singh et Choudhry 1982). D'autres travaux d'application, de recherche et de mise au point sont actuellement en cours à Statistique Canada dans le cadre du programme de données sur les petites régions.

#### 7.4 Estimation de la variance

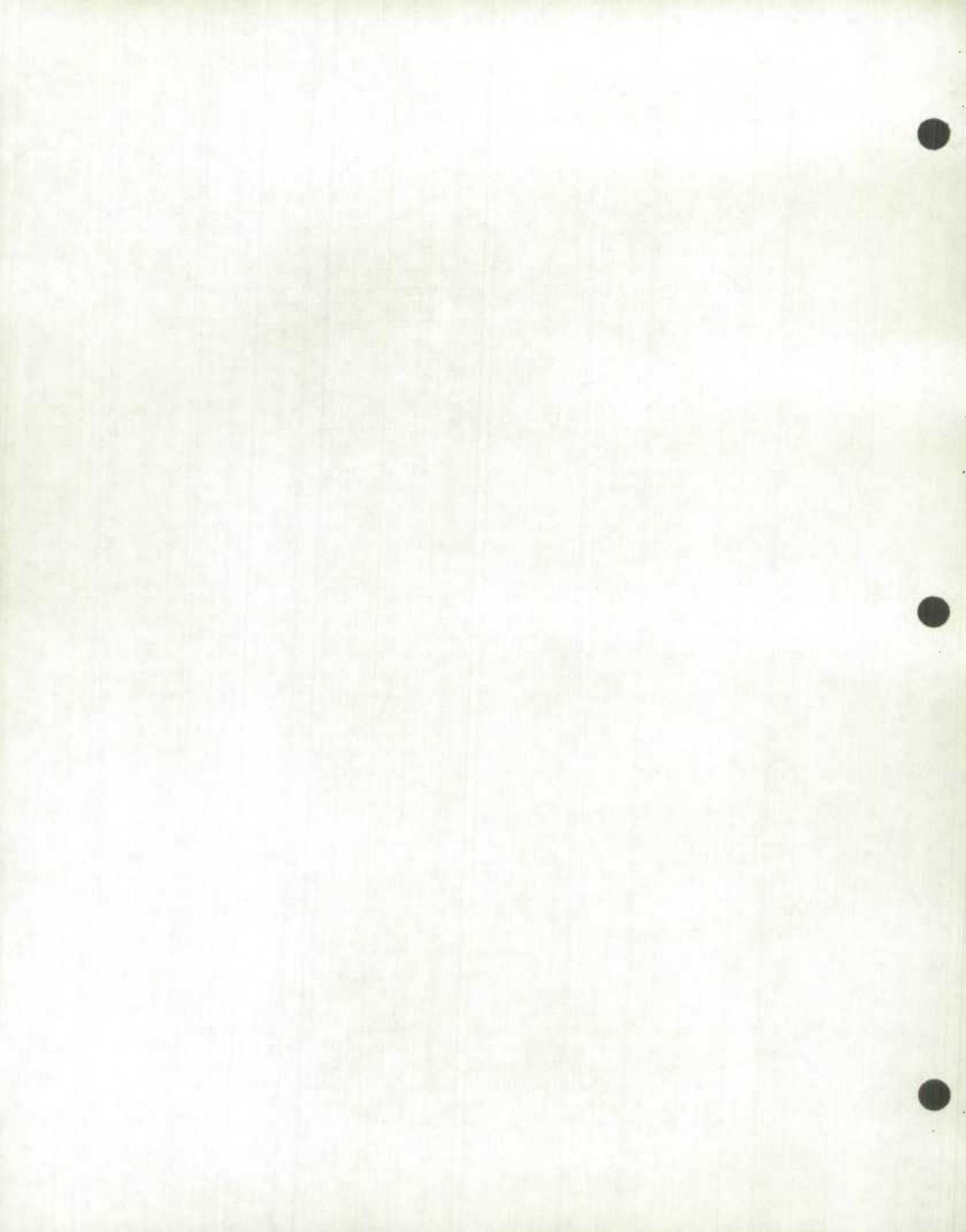
La méthode utilisée pour l'estimation de la variance dans l'échantillon remanié sera de nouveau celle de Keyfitz (1957), quoiqu'elle sera modifiée de façon à produire une estimation à deux étapes par la méthode du quotient pour le degré final, c'est-à-dire une estimation calculée en une seule itération de la méthode itérative du quotient. Comme les itérations subséquentes n'ont qu'un très petit effet sur les estimations, on n'en tient pas compte dans l'estimation de la variance. D'autres modifications de la technique actuelle d'estimation de la variance sont également à l'étude. Ainsi, on songe à utiliser les grappes comme échantillons répétés dans les UPÉ, au lieu de regrouper les grappes en deux pseudo-échantillons répétés, comme à l'heure actuelle.

On a également songé à remplacer la méthode actuelle d'estimation de la variance, par celles décrites par Rao, Hartley et Cochran (1962) et Rao (1975) pour les UPÉ, où le plan de sondage de Rao, Hartley et Cochran est utilisé (Choudhry, Lee et Sida 1984). La méthode actuelle ainsi que les autres ont été examinées avec ou sans correction par la méthode du quotient. On a constaté que, sans correction par la méthode du quotient, la méthode actuelle exagère la variance de certaines caractéristiques (par exemple, la variance du nombre de personnes occupées est surestimée de 20%), mais, avec l'estimation par le quotient, les biais sont négligeables. Les biais estimés étaient également négligeables dans le cas des autres estimateurs comparés. Le principal avantage des ces autres estimateurs est qu'ils sont plus stables. Toutefois, on a retenu la méthode actuelle parce qu'elle est simple, alors qu'il est complexe d'estimer les variances des variations ou des moyennes avec les autres méthodes.

#### 7.5 Estimateurs composites

Dans l'EPA, il existe des corrélations modérées ou fortes entre la plupart des caractéristiques observées au cours de deux mois consécutifs, parce que 5/6 de l'échantillon demeure le même. Kumar et Lee (1983) ont étudié divers estimateurs composites qui mettent à profit ces corrélations en utilisant des données sur les échantillons précédents pour améliorer les estimations du mois courant. Leurs expériences ont porté sur une classe d'estimateurs composites AK examinés récemment par Huang et Ernst (1981) et sur d'autres estimateurs employés dans la Current Population Survey du U.S. Bureau of the Census.

En supposant que l'estimateur de la méthode du quotient est sans biais, Kumar et Lee ont évalué l'erreur quadratique moyenne et constaté qu'un compromis entre les poids A et K produisait un gain d'efficacité de 5% au plus pour les estimations mensuelles du niveau de l'emploi et du chômage, et de 5% à 16% pour les estimations correspondantes des variations mensuelles. Avant qu'une décision soit prise concernant l'utilisation d'estimateurs composites dans l'EPA, d'autres études doivent être faites sur ces estimateurs pour évaluer l'effet des variations dans le biais de renouvellement dues aux modifications apportées aux rectifications pour tenir compte de la non-réponse et au calcul des estimations par le quotient et on se penchera également sur les répercussions possibles de ces estimateurs sur le déroulement de l'enquête.



### 7.6 Nouvelle règle pour l'arrondissement et la publication des chiffres

Dans l'ancienne EPA, les estimations de niveaux étaient arrondies au <sup>centième</sup> millièmes et publiées si elles dépassaient 4,000. Cette règle était appliquée uniformément à toutes les provinces et à toutes les estimations, de manière à ce que les données publiées aient un coefficient de variation de 33.3% ou moins.

Des normes plus rigoureuses pour l'arrondissement et la publication des chiffres de chaque province ont été établies pour l'échantillon remanié. Dorénavant, le CV des estimations non arrondies doit être inférieur ou égal à 33.3% et l'erreur d'arrondissement ne doit pas dépasser 20% de l'erreur-type d'une estimation non arrondie. On a constaté que la norme de publication pouvait être réduite à 2,000 ou 3,000 pour toutes les provinces, sauf le Québec et l'Ontario, et que les estimations infraprovinciales devaient être arrondies au centième plutôt qu'au millième (Kumar 1982).

## 8. REDISTRIBUTION DE L'ÉCHANTILLON

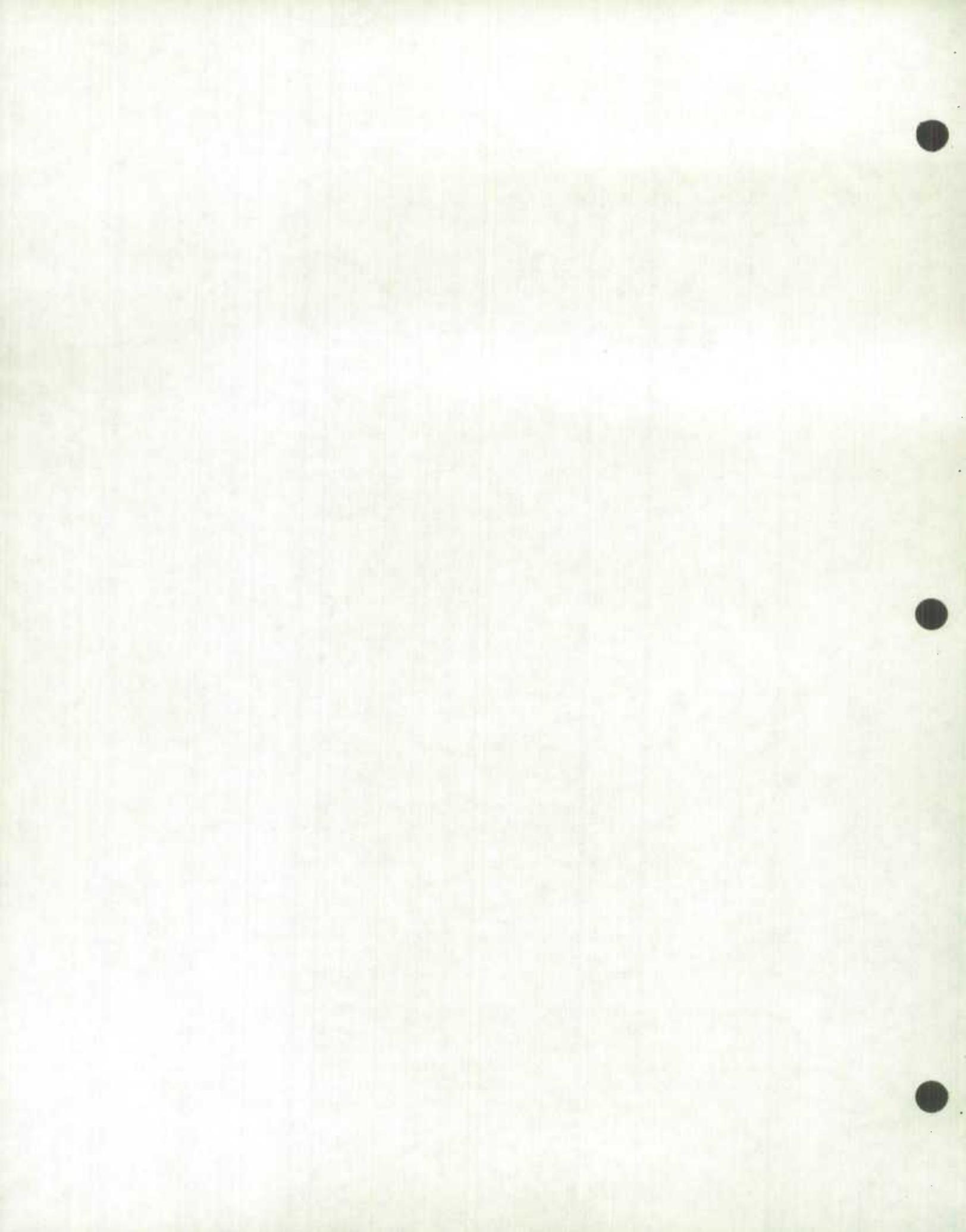
Les objectifs du remaniement visant l'amélioration des données infraprovinciales ont été précisés à la section 2. Les modifications apportées aux techniques et aux méthodes utilisées dans l'EPA favorisent une amélioration générale de la précision des données, cependant, il a également fallu songer à une nouvelle répartition de l'échantillon à l'intérieur des provinces pour atteindre les objectifs (ii), (iii) et (iv). Il était nécessaire d'élargir l'échantillon dans 13 des 66 régions économiques, dans 6 des 24 RMR et dans 27 des 42 villes autres que des RMR. Les CV du nombre de chômeurs ont pu être réduits de 28% en moyenne pour ces régions. Par ailleurs, pour les 30 RÉ où les CV se situaient dans l'intervalle de 15% à 25%, la nouvelle répartition a produit une baisse de 12% dans les CV. Pour des raisons d'ordre pratique, dans l'échantillon remanié, les données mensuelles relatives aux RÉ et aux RMR et les données trimestrielles relatives aux autres villes seront fondées sur des échantillons comprenant un minimum de 300 et 120 ménages respectivement par mois.

Il convient de souligner que deux utilisations importantes des données de l'EPA par les ministères fédéraux n'ont pas été prises en considération directement dans les objectifs du remaniement. Il s'agit de la production de moyennes mobiles du taux de chômage de 3 mois consécutifs, dans les régions infraprovinciales de l'assurance-chômage (AC), pour le calcul du nombre de semaines qu'une personne dans chaque région différente doit avoir travaillées pour toucher des prestations d'AC, et de la production de moyennes de 3 ans, dans quelques 180 à 200 régions composées de divisions de recensement distinctes ou combinées, pour la répartition de l'aide financière fédérale aux nouvelles initiatives industrielles. Toutefois, la redistribution de l'échantillon comporte des avantages indirects pour ces deux opérations.

Pour déterminer la taille des échantillons qui convient aux objectifs visés, on a utilisé les taux de chômage moyens de la période 1980-1982, puisque les prévisions à moyen terme indiquent que le chômage demeurera élevé au cours des années 1980.

Une des conséquences générales de la redistribution de l'échantillon a été un déplacement d'une bonne partie de l'échantillon des grandes RMR et des régions économiques vers les petites. Ce transfert s'est répercuté négativement sur les estimations provinciales et nationales parce l'échantillon n'était plus réparti proportionnellement. Cette diminution de la précision des estimations globales a cependant été plus que compensée par les gains de la fiabilité dus aux améliorations structurelles que les travaux de recherche ont permis d'apporter aux techniques et aux méthodes de l'enquête.

Une étude a également été menée à l'aide du modèle coût-variance proposé par Fellegi, Platteau et Gray (1967) pour déterminer les taux de sondage optimaux dans les régions NAR et AR. Le résultat a été un déplacement de l'échantillon des régions NAR vers les régions AR. Ce transfert a été particulièrement important au Québec et en Ontario, où la proportion de l'échantillon dans les régions AR a augmenté de 0.60 à 0.72 (0.78 de la base de sondage), et a permis d'améliorer les estimations provinciales du nombre de chômeurs avec l'équivalent d'une réduction de 5% de la variance pour un échantillon de taille fixe. Cette optimisation a également aidé à atteindre les objectifs (ii) et (iv) et a eu des retombées positives pour l'enquête sur les finances des consommateurs et l'enquête sur les loyers.



Les améliorations structurelles du plan de sondage découlant de facteurs comme les changements dans la stratification, l'élimination d'un degré d'échantillonnage dans les régions NAR, l'utilisation de chiffres repères infraprovinciaux dans le calcul des estimations et la nouvelle répartition de l'échantillon, ont permis non seulement d'accroître la précision des estimations nationales et provinciales, mais également d'atteindre les objectifs visés pour la production de données infraprovinciales. Il devenait alors possible de réduire la taille globale de l'échantillon de l'EPA afin d'affecter des ressources financières à la collecte occasionnelle de données sur certaines autres variables socio-économiques. La taille de l'échantillon a donc été fixée à 51,500 ménages par mois au lieu de 55,000. Cette baisse globale d'entre 6% et 7% a été réalisée par une diminution uniforme de la taille de l'échantillon dans toutes les provinces sauf l'Île-du-Prince-Édouard. En outre, les coûts unitaires de la collecte des données diminueront à cause de l'augmentation du nombre d'interviews téléphoniques.

### 9. EFFETS DU REMANIEMENT SUR LES ENQUÊTES LIÉES À L'EPA

Pour la plupart des enquêtes-ménages de Statistique Canada, la base de sondage, le plan d'échantillonnage et les systèmes de collecte et de traitement des données de l'EPA représentent des instruments qui permettent de recueillir des données plus rapidement, plus économiquement et avec plus de précision que par des sondages indépendants. Le plan et les opérations de ces enquêtes sont intégrés, à divers degrés, à ceux de l'EPA.

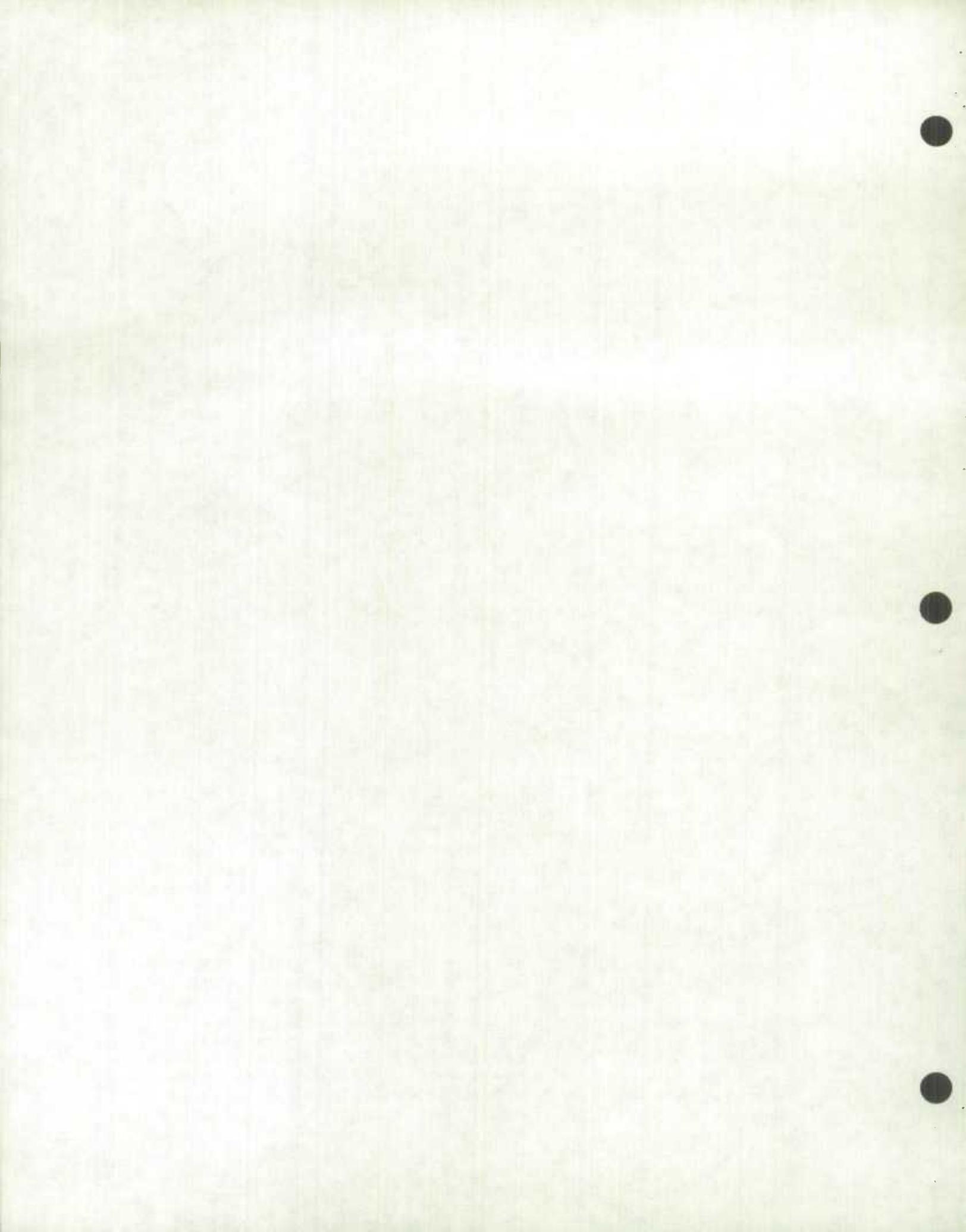
La plupart de ces projets prennent la forme d'enquêtes supplémentaires pour lesquelles des questions sont ajoutées au questionnaire de l'EPA et les seuls frais sont les coûts additionnels de ces nouvelles questions. Les enquêtes qui peuvent nuire à l'EPA, soit parce qu'elles portent sur un sujet délicat, soit parce que leur questionnaire est trop long, ne sont pas menées de cette manière. Normalement, les données pour ce genre d'enquête sont recueillies par les interviewers de l'EPA, auprès d'un échantillon séparé de ménages, dans les régions où l'EPA est menée. D'autres enquêtes sont encore moins étroitement liées à l'EPA parce qu'elles n'ont pas lieu dans les mêmes régions que l'EPA, mais sont élaborées à partir du plan de sondage de l'EPA et soumises à un contrôle pour éviter un chevauchement avec l'échantillon de l'EPA.

Tel qu'il a été mentionné plus haut, le programme de remaniement vise avant tout l'EPA, mais on a également essayé d'accroître l'utilité générale de l'EPA pour d'autres activités statistiques. Les modifications apportées à l'EPA à cette fin sont résumées brièvement ci-dessous.

La redistribution d'une partie de l'échantillon des régions NAR vers les régions AR produira une répartition plus robuste en général et, en particulier, améliorera l'estimation des variations du revenu et des loyers dans l'enquête sur les finances des consommateurs et l'enquête sur les loyers. Par ailleurs, la taille minimale de l'échantillon des RMR est maintenant de 300 ménages, ce qui aura un effet positif sur les enquêtes qui produisent des estimations relatives aux RMR.

Le nouveau schéma général de stratification multidimensionnelle en fonction de 15 variables (dans toutes les régions NAR et AR) sera également mieux pour les autres enquêtes que la méthode actuelle de stratification par activité économique ou par unité géographique.

Trois changements constituent un avantage direct pour les enquêtes menées auprès de différents ensembles de ménages: (i) l'élimination d'un degré d'échantillonnage dans les régions rurales entraînera une réduction énorme du délai de préavis, qui était de 13 mois auparavant, mais de 7 mois dans la nouvelle EPA, (ii) les UPÉ seront plus compactes géographiquement parce qu'il y aura dorénavant des strates rurales et urbaines explicites, ce qui facilitera les petites enquêtes où les UPÉ doivent bien correspondre aux tâches des interviewers et (iii) la fiabilité découlant de l'amélioration du programme de stabilisation de l'échantillon permettra de tirer des sous-échantillons de pratiquement n'importe quelle taille, à l'échelle nationale, provinciale ou infraprovinciale, pour les enquêtes élaborées à partir du plan de sondage de l'EPA.



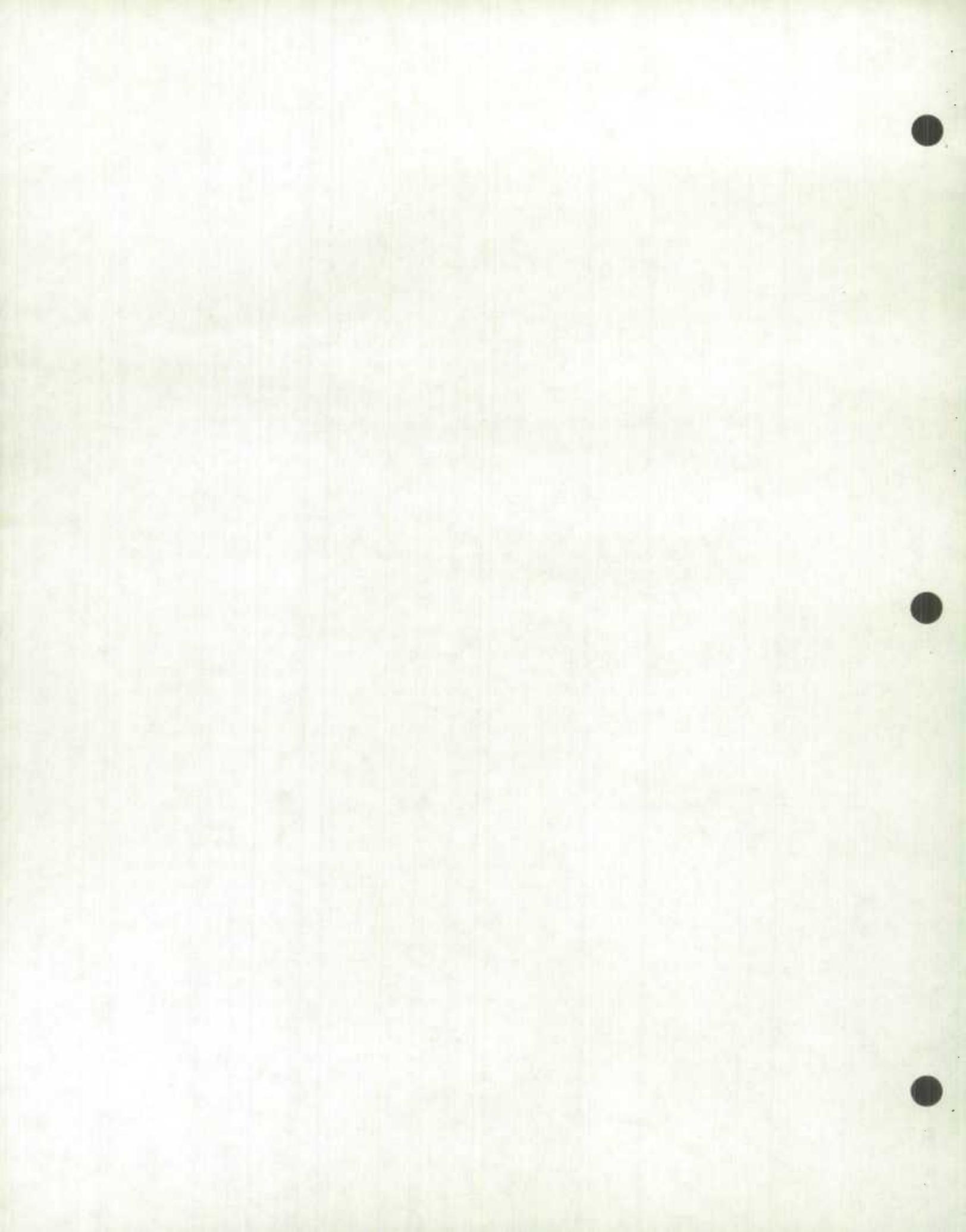
Enfin, le fait que la méthode d'estimation choisie pour le remaniement a été modifiée de façon à ce qu'une distribution de la taille des ménages soit incluse à titre provisoire dans la rectification par la méthode du quotient (des estimations démographiques de la taille des familles seront incluses plus tard comme dimension supplémentaire dans le calcul des estimations par la méthode itérative du quotient pour le dernier degré) augmentera la convergence des statistiques sur le marché du travail relatives aux familles et aux personnes, et améliorera les statistiques sur les dépenses et le revenu des familles.

## 10. SOMMAIRE DES CHANGEMENTS ET DES FUTURS PROJETS DE RECHERCHE

La plupart des travaux de recherche pour le remaniement de l'EPA à partir des résultats du recensement de 1981 sont terminés, la mise en oeuvre du nouveau plan de sondage est en cours et certains aspects de la recherche sur les méthodes d'estimation doivent se poursuivre. Fait intéressant, beaucoup d'études faites pour ce programme ont confirmé la validité des techniques et des méthodes employées jusqu'à présent dans l'EPA comme, par exemple, la sélection de deux UPÉ par strate dans les régions NAR, l'application de la méthode de Rao, Hartley et Cochran et d'un facteur de densité (4 ou 5 ménages par grappe) aux régions AR et le maintien du plan de renouvellement étalé sur six mois. Toutefois, plusieurs recherches ont également permis d'améliorer le programme de renouvellement et le nouveau plan de sondage; les principaux facteurs qui ont favorisé ces changements sont l'importance accordée aux objectifs concernant la fiabilité des données (voir section 2), la production d'informations plus précises et facilement accessibles et les progrès technologiques.

Parmi les améliorations du programme de remaniement, mentionnons l'utilisation de données du recensement de 1981 au lieu de données recueillies indépendamment sur le terrain pour la mise à jour de l'échantillon des régions AR, la simplification de la délimitation des degrés du plan de sondage et l'automatisation de la stratification et de la formation des UPÉ. Il sera également possible d'éviter des dépenses par l'introduction progressive d'une bonne partie de l'échantillon remanié. Ainsi, un groupe de renouvellement choisi selon le nouveau plan de sondage remplacera un groupe de renouvellement de l'ancien échantillon, une fois tous les mois au cours d'une période de 6 mois, alors que, dans le passé, l'ancien échantillon continuait d'être renouvelé pendant 3 ou 4 mois en même temps que le nouvel échantillon croissait (Mayda, Drew et Lindeyer 1984). Il a été estimé que le remaniement actuel coûtera \$1.8 million de moins (en dollars de 1983-1984) que le remaniement précédent.

Les principales améliorations qui ont rendu le plan de sondage de l'EPA plus économique qu'auparavant sont l'extension des interviews téléphoniques à la collecte des données pour la période du deuxième au sixième mois d'inclusion dans l'échantillon à l'intérieur des régions NAR, la révision du plan de sondage des régions NAR pour définir des strates urbaines et rurales explicites, l'élimination d'un degré d'échantillonnage dans les régions rurales, l'élaboration d'un schéma général de stratification pour les régions AR et NAR, l'utilisation de chiffres de population repères dans le calcul des estimations et la mise au point d'une meilleure répartition de l'échantillon. Ces améliorations ont suffi pour accroître la précision des données infraprovinciales sans nuire à celle des données relatives aux provinces, même si la taille globale de l'échantillon a diminué d'entre 6% et 7%. La fiabilité a augmenté de 14% en moyenne pour les coefficients de variation du nombre de chômeurs dans la moitié des régions économiques et des RMR où les estimations étaient les moins fiables dans l'ancien plan de sondage, et les changements constatés dans les autres régions étaient faibles ou nuls. Les gains d'efficacité seront encore plus importants pour les estimations infraprovinciales du nombre de personnes occupées. En ce qui a trait au coût de l'enquête, la diminution de la taille de l'échantillon et la réduction des frais entraînée par l'extension des interviews téléphoniques permettront d'économiser environ \$0.7 million par année (en dollars de 1983-1984).



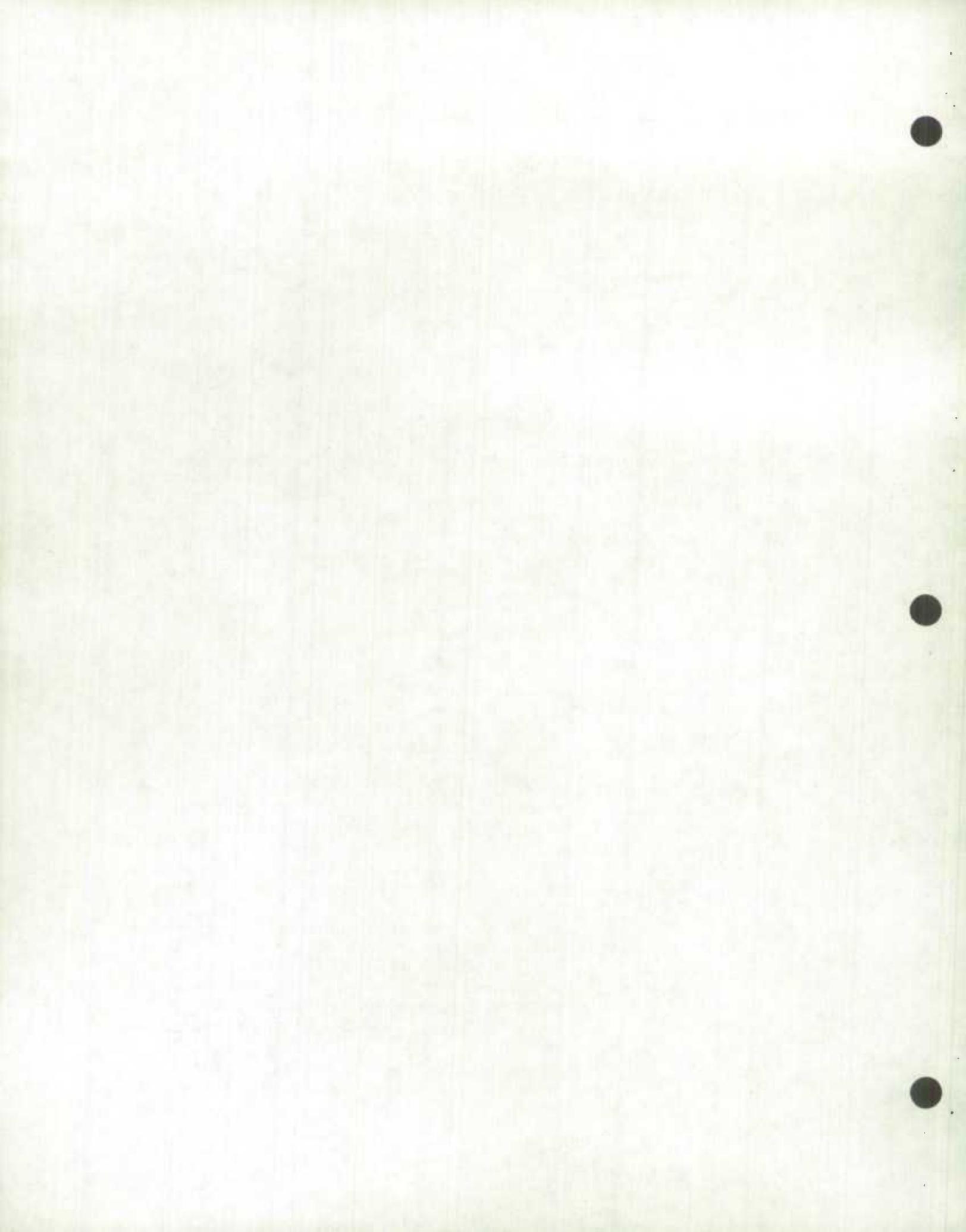
Une fois que le remaniement de l'échantillon sera terminé, une des principales questions examinées dans les recherches futures sur la méthodologie de l'EPA sera l'élaboration d'un plan de sondage double dans lequel une partie de l'échantillon sera convertie en une base de sondage téléphonique à l'aide des techniques de composition de numéros au hasard (CNH). Dans le cadre d'un nouveau programme d'élaboration d'enquêtes téléphoniques (Hofmann, Drew, Catlin et Mayda 1984), des statisticiens planifient à l'heure actuelle un programme pluriannuel qui comprendra des expériences avec la méthode de CNH pour étudier les incidences d'une augmentation du taux de non-réponse sur l'échantillon choisi pour les interviews téléphoniques, des recherches sur les méthodes d'estimation applicables aux plans de sondage doubles et une évaluation des avantages et des désavantages de la centralisation et de la décentralisation des interviews téléphoniques. Un autre projet d'étude portera sur les moyens de mettre à jour économiquement l'échantillon aréolaire des régions AR au cours de la période intercensitaire.

### REMERCIEMENTS

Les auteurs remercient les membres du comité chargé du remaniement de l'échantillon de l'EPA pour leur appui et leur conseils pendant le déroulement du programme de remaniement et, plus particulièrement, I.P. Fellegi, D.B. Petrie, G.J. Brackstone, R. Platek, I. Macredie, M. Levine, M. Brochu et F. Mayda. Les auteurs remercient également tous les membres de l'équipe de méthodologie, qui ont participé aux travaux de recherche et de mise en application de ce projet et ont rendu possibles les améliorations décrites plus haut. Les auteurs remercient également l'arbitre pour ses observations utiles.

### BIBLIOGRAPHIE

- CHOUHRY, G.H., LEE, H. et DREW, J.D. (1984). Cost Variance Optimizations for the Canadian Labour Force Survey. Document technique interne (en voie de rédaction), Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- CHOUHRY, G.H., LEE, H. et SIDA, R. (1984). Variance Estimation for the Redesigned Labour Force Survey. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- DREW, J.D., CHOUHRY, G.H. et GRAY, G.B. (1978). Some Methods for Updating Sample Survey Frames and Their Effects on Estimation. *Techniques d'enquête*, 4, 225-263.
- DREW, J.D., SINGH, M.P. et CHOUHRY, G.H. (1982). Évaluation des techniques d'estimation pour les petites régions dans l'enquête sur la population active au Canada. *Techniques d'enquête*, 8, 19-52.
- EARWAKER, S. et BÉLANGER, Y. (1981). Ratio Estimation at the Subprovincial Level for the Labour Force Survey. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- FELLEGI, I.P., GRAY, G.B. et PLATEK, R. (1967). The New Design of the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 62, 421-453.
- FOY, P., BÉLANGER, Y., DREW, J.D. et JONCAS, M. (1984). Multivariate Clustering Algorithm for Stratifications and its Application to the Canadian Labour Force Survey. Document technique (en voie de rédaction), Division des méthodes de recensement et d'enquête-ménages, Statistique Canada.
- FRIEDMAN, H.P. et RUBIN, J. (1967). On some Invariant Criteria for Grouping data. *Journal of the American Statistical Association*, 62, 1159-1178.
- GHANGURDE, P.D. (1984). Evaluation of LFS Non-Response Adjustment in Household Size Cells. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- GRAY, G.B. (1973). On Increasing the Sample Size (No. of PSUs). Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.



- HARTLEY, H.O. et RAO, J.N.K. (1962). Sampling with Unequal Probabilities without Replacement. *Annals of Mathematical Statistics*, 33, 350-374.
- HOFMANN, H., DREW, D., CATLIN, G. et MAYDA, F. (1984). A Proposal for a Telephone Survey Development Program. Document interne, Statistique Canada.
- HUANG, E. et ERNST, L. (1981). Comparison of an Alternate Estimator to the Current Composite Estimator in CPS. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 303-308.
- JUDKINS, D.R. et SINGH, R.P. (1981). Using Clustering Algorithms to Stratify Primary Sampling Units. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings*, 274-284.
- KEYFITZ, N. (1951). Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities. *Journal of the American Statistical Association*, 46, 105-109.
- KEYFITZ, N. (1957). Estimates of Sampling Variance where two Units are Selected from each Stratum. *Journal of the American Statistical Association*, 52, 503-510.
- KOSTANICH, D., JUDKIN, D., SINGH, R. et SCHANTZ, M. (1981). Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings*.
- KUMAR, S. (1982). Investigation of the Labour Force Survey Rounding and Release Criteria. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- KUMAR, S. et LEE, H. (1983). Évaluation de l'application d'estimateurs composites à l'enquête sur la population active au Canada. *Techniques d'enquête*, 9, 196-221.
- LEMAÎTRE, G. (1983). Some Results from the Time and Cost Study. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- MACREDIE, I. (1983). Family Oriented Measures of Employment and Unemployment. Exposé présenté au comité de travail de l'OCDE sur les statistiques de l'emploi et du chômage.
- MAYDA, F., DREW, D. et LINDEYER, J. (1984). Phase-in of the Redesigned Labour Force Survey Sample. Document technique, (en voie de rédaction), Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- PAUL, E.C. et LAWES, M. (1982). Caractéristiques des ménages répondants et non répondants dans l'enquête sur la population active du Canada. *Techniques d'enquête*, 8, 53-93.
- PLATEK, R. et SINGH, M.P. (1976). Méthodologie de l'enquête sur la population active du Canada. N 71-526 au catalogue, Statistique Canada.
- PLATEK, R. et SINGH, M.P. (1977). A Strategy for Updating Continuous Surveys. *Metrika*, 25, 1-7.
- RAO, J.N.K. (1975). Unbiased Variance Estimation for Multi Stage Designs. *Sankhya*, série C, 37, 133-139.
- RAO, J.N.K., HARTLEY, H.O. et COCHRAN, W.G. (1962). On a Simple Procedure of Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society*, série B, 24, 482-490.
- SINGH, M.P. et DREW, J.D. (1977). Sample Expansion in Self Representing Units of the Canadian Labour Force Survey. Document technique, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.
- SINGH, M.P. et DREW, J.D. (1981a). Research Plans for the Canadian Labour Force Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings*.
- SINGH, M.P. et DREW, J.D. (1981b). Redesigning Continuous Surveys in a Changing Environment. *Techniques d'enquête*, 7, 44-73.
- VERMA, R.B.P., BASAVARAJAPPA, K.G. et BENDER, R.K. (1983). The Regression Estimates of Population for Sub-Provincial Areas in Canada. *Techniques d'enquête*, 9, 242-266.

008

STATISTICS CANADA LIBRARY  
BIBLIOTHEQUE STATISTIQUE CANADA



1010148858

