

## **Plan d'échantillonnage pour l'enquête sur la santé et les limitations d'activités**

**D. DOLSON, K. McCLEAN, J.-P. MORIN, et A. THÉBERGE<sup>1</sup>**

### RÉSUMÉ

L'Enquête sur la santé et les limitations d'activités s'inscrit dans le cadre du programme visant à établir une base de données relative à la population ayant une incapacité ou un handicap au Canada. On décrit le plan d'échantillonnage utilisé pour la partie de l'enquête couvrant la population vivant hors des institutions. Les méthodes employées pour déterminer les tailles des échantillons et pour la sélection de ceux-ci sont aussi exposées.

**MOTS CLÉS:** Incapacité; échantillonnage stratifié; échantillonnage à deux degrés; allocation optimale; échantillonnage sans remise.

### 1. INTRODUCTION

Dans le cadre du programme visant à obtenir plus d'information sur la population des personnes handicapées au Canada, l'Enquête sur la santé et les limitations d'activités (ESLA) a été réalisée au cours de l'automne 1986. Elle vise à recueillir des informations sur la nature des problèmes éprouvés par cette population et, d'une façon générale, sur leurs activités quotidiennes (à la maison, au travail, à l'école, dans les déplacements, etc.). L'enquête se divise en deux parties: l'une couvre la population vivant en institution et l'autre, qui fait l'objet du présent article, couvre la population hors des institutions.

On a partitionné le Canada en 238 régions infra-provinciales (RIP). Ces régions comprennent toutes les municipalités du Québec et de l'Ontario qui regroupent plus de 125,000 habitants et toutes celles des autres provinces de plus de 75,000 habitants. Les autres régions sont composées d'agglomérations de sous-divisions de recensement respectant la contiguïté géographique et les frontières entre les provinces. Le nombre de ces régions dans chaque province est proportionnel à la racine carrée de la population moins les municipalités précédemment définies. Un des principaux objectifs de l'enquête est de produire au niveau des RIP des statistiques sur la population handicapée afin de permettre une analyse détaillée des divers besoins. En outre, on fournira des estimations pour trois groupes d'âge: enfants (moins de 15 ans), adultes (de 15 à 64 ans) et personnes âgées (65 ans et plus).

La cueillette des informations s'est effectuée en deux étapes. La première consiste en une question à plusieurs volets, incluse dans le formulaire 2B du recensement de la population du Canada de 1986, qui porte sur les limitations du répondant dans divers secteurs d'activités ainsi que sur sa propre évaluation de sa condition. Une copie de cette question (question n° 20 du formulaire du recensement de 1986) est donnée en appendice. La seconde étape a eu lieu quelque temps après le recensement. Elle est constituée d'un questionnaire de sélection et d'un suivi qui recueille des informations sur les problèmes et les activités des répondants handicapés.

La première étape vise essentiellement à répartir les répondants en deux groupes: ceux qui ont répondu "oui" à au moins une des parties de la question 20 et ceux qui ont répondu "non" à toutes les parties. Le but est d'identifier à l'avance une grande partie de la population potentiellement handicapée afin de concentrer le maximum des ressources de l'enquête sur le groupe cible. Notons toutefois que les enquêtes précédentes montrent que le groupe cible ne sera pas complètement identifié par cette question. (Voir Dolson et coll. (1984) et Dolson et coll. (1986)).

<sup>1</sup> D. Dolson, K. McClean, J.-P. Morin, et A. Théberge, Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa, Ontario, K1A 0T6.

La seconde étape est l'ESLA. On procède à des interviews individuelles pour la strate des "oui" et à des interviews téléphoniques pour la strate des "non". Du point de vue opérationnel, l'interview est formée de deux parties: le questionnaire de sélection et le suivi.

Le rôle du questionnaire de sélection est de déterminer pour quel répondant le questionnaire de suivi est pertinent. Le questionnaire destiné aux adultes comprend les 17 activités de la vie quotidiennes (AVQ) utilisées lors de l'Enquête sur la santé et les incapacités au Canada qui a eu lieu en 1983 et 1984 ainsi que la partie (a) de la question 20 du recensement et quelques questions sur les maladies et les handicaps mentaux (voir l'appendice B). Si une réponse positive est obtenue à au moins une de ces questions, l'interviewer procède au suivi, sinon l'interview prend fin. La partie (a) de la question du recensement est posée de nouveau pour tenir compte d'un éventuel changement de situation causé soit par une réponse d'un intermédiaire lors du recensement, soit par une réévaluation du répondant de son propre état.

La section de sélection sur le questionnaire destiné aux enfants comprend des questions sur les appareils spéciaux, les limitations aux activités, la fréquentation d'une école spéciale et les affections ou problèmes de santé. Une réponse positive à au moins une de ces questions entraîne l'interview de suivi. La question du recensement n'est pas posée de nouveau parce que tous les interviews relatives aux enfants nécessitent un intermédiaire et que la question sur les limitations d'activités est équivalente à la partie (a) de la question 20 du recensement.

La deuxième section du présent article décrit comment la population canadienne a été subdivisée en différentes sous-populations pour fins d'estimation. La section trois est consacrée au plan de sondage de l'ESLA. La section suivante traite du fichier des données géographiques et démographiques projetées pour 1986 qui a servi à former les bases de sondage. Enfin la dernière section décrit comment on a procédé à l'échantillonnage.

## 2. POPULATIONS COUVERTES

Les personnes qui résident en permanence dans les hôpitaux généraux, les hôpitaux psychiatriques, les centres de soins spéciaux ou les établissements pour personnes âgées et malades chroniques, les centres de traitement ou établissements pour handicapés physiques et les orphelinats ou foyers pour enfants font l'objet d'une partie spéciale de l'enquête: l'ESLA-institutions. Dans le présent article, il est question de la partie de l'enquête qui couvre la population canadienne à l'exclusion des personnes couvertes par l'ESLA-institutions, des personnes qui résident dans les prisons, les camps militaires, les foyers pour les jeunes délinquants, les navires de guerre, les établissements pénitentiaires, les établissements de correction et des résidents des logements collectifs de la catégorie "autres" (par exemple, cirques, communes non-religieuses).

Chaque secteur de dénombrement (SD) dont la population n'est pas totalement exclue de l'enquête est classé dans une des cinq bases de sondage suivantes:

1. Réserves indiennes où le dénombrement avait été fait au moyen d'interviews en 1981;
2. Autres réserves indiennes;
3. SD où le dénombrement se fait habituellement par interviews;
4. SD dans les RIP de Whitehorse, Yellowknife, Pine Point, Hay River et Fort Smith;
5. Tous les autres SD.

L'ordre de priorité pour l'appartenance à une base est 1-2-4-3-5. Ainsi, un SD formé par une réserve indienne situé dans la RIP de Whitehorse est classé comme étant une réserve indienne.

Chaque SD est divisé en deux: le SD "oui", formé des personnes qui donneraient une réponse positive à la question du recensement, et le SD "non" formé des personnes qui y répondraient négativement. Un plan d'échantillonnage différent est utilisé pour chacune des cinq bases de sondage: dans la première base, tous les SD "oui" sont sélectionnés et aucun des SD "non"; dans la base deux, tous les SD "oui" et un échantillon des SD "non" sont sélectionnés; dans la troisième base, aucun des SD "non" n'est choisi et un échantillon des SD "oui" sont sélectionnés; tous les SD de la base quatre sont sélectionnés; et enfin un échantillon des SD "oui" et un échantillon des SD "non" sont choisis dans la base cinq.

### 3. PLAN DE SONDAGE

La méthode d'échantillonnage dont il est question dans cette section a été utilisée pour les bases de sondage trois et cinq. Pour des raisons d'espace, le plan de sondage relatif à la deuxième base de sondage n'est pas exposé dans le présent article. (Pour plus de renseignements sur la méthodologie de l'ESLA, voir Dolson et coll. (1986)).

#### 3.1 Plan d'échantillonnage

Chaque province est partitionnée en régions infra-provinciales (RIP) lesquelles sont elles-mêmes partitionnées en secteurs de dénombrement (SD).

Les habitants de chaque SD sont divisés en un SD "oui" et un SD "non", selon qu'ils répondraient positivement ou négativement à la question 20 du formulaire du recensement. Dans chaque RIP, on stratifie les SD "oui" en grands et petits SD suivant le critère expliqué dans la quatrième section du présent article. Les personnes appartenant à un SD "oui" sont associés à une strate et à une RIP alors que les personnes appartenant à un SD "non" ne sont associés qu'à leur SD. Dans chaque province, la population est subdivisée en trois groupes d'âges: enfants (moins de 15 ans), adultes (15 à 64 ans) et personnes âgées (65 ans et plus).

La méthode d'échantillonnage est un plan de sondage stratifié à deux degrés pour les SD "oui" de chaque RIP et un plan de sondage à deux degrés pour les SD "non" de la province. Les SD constituaient les unités primaires et les répondants les unités secondaires.

Toutes les personnes ayant répondu au formulaire 2B du recensement dans un SD "oui" sélectionné dans l'échantillon sont interviewées et un tiers d'entre elles le sont dans les SD "non" sélectionnés dans l'échantillon.

#### 3.2 Allocation de l'échantillon

Ce plan de sondage doit permettre de minimiser les coûts d'échantillonnage étant donné un coefficient de variation maximum des estimations et une variance fixée de l'estimateur  $\hat{B}$  du biais relatif  $B$ . On définit  $B$  comme étant le rapport du nombre  $T_0$  de personnes "non" qui ont une caractéristique d'intérêt dans la province, sur le nombre  $T_1$  de personnes "oui" qui ont une caractéristique d'intérêt dans la province. Par personne "non" il faut entendre une personne qui répondrait négativement à tous les volets de la question 20 du formulaire du recensement, et par personne "oui", une personne qui répondrait positivement à au moins un des volets.

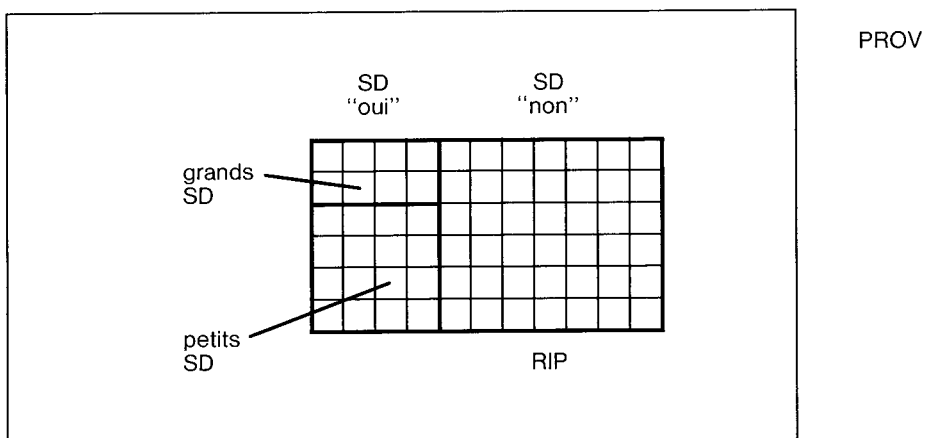


Figure 1. Illustration du plan de sondage.

Notons  $N_0$  le nombre de SD "non" dans la province,  $N_{jk}$  le nombre de SD "oui" dans la strate  $j$  et la RIP  $k$  de la province,  $n_0$  et  $n_{jk}$  les tailles d'échantillons correspondantes et  $c_0$  et  $c_{jk}$  les coûts unitaires d'échantillonnage correspondants. Si on a  $N_p$  RIP dans la province, on veut donc minimiser

$$\sum_{k=1}^{N_p} (c_{1k} n_{1k} + c_{2k} n_{2k}) + c_0 n_0$$

étant donné

$$CV^2(y_k) \leq CV^2_*; \text{Var}(\hat{B}) = \text{Var}_*(\hat{B});$$

$$n_{jk} \leq N_{jk}; n_{2k} = \lambda_k n_{1k}; n_0 \leq N_0$$

$$(j = 1, 2; k = 1, \dots, N_p)$$

où  $\lambda_k$  est le rapport du nombre attendu de personnes handicapées dans les petits SD sur le nombre attendu de personnes handicapées dans les grand SD de la RIP  $k$ ,  $y_k$  est le nombre estimé de personnes "oui" qui ont une caractéristique d'intérêt dans la RIP  $k$  et les valeurs notées avec des \* sont des constantes.

Si la fraction de sondage dans les SD "oui" est  $f_1$ ,  $M_{ijk}$  est le nombre de personnes "oui" dans le SD  $i$  de la strate  $j$  de la RIP  $k$  de la province, et  $p_{ijk}$  est la probabilité d'une caractéristique d'intérêt pour une personne "oui" du SD  $i$  de la strate  $j$  de la RIP  $k$ , alors on a

$$E(y_k) = Y_k = \sum_{j=1}^2 \sum_{i=1}^{N_{jk}} M_{ijk} p_{ijk},$$

$$\text{Var}(y_k) = \sum_{j=1}^2 \left\{ \frac{N_{jk}^2}{n_{jk}} \left(1 - \frac{n_{jk}}{N_{jk}}\right) S_{jk}^2 + \frac{N_{jk}}{n_{jk}} \left(\frac{1-f_1}{f_1}\right) \sum_{i=1}^{N_{jk}} M_{ijk} S_{ijk}^2 \right\},$$

où

$$y_k = \sum_{j=1}^2 \sum_{i=1}^{N_{jk}} \frac{N_{jk}}{n_{jk}} M_{ijk} p_{ijk},$$

$$S_{jk}^2 = \frac{1}{N_{jk} - 1} \sum_{i=1}^{N_{jk}} \left( M_{ijk} p_{ijk} - \left( \sum_{i=1}^{N_{jk}} \frac{M_{ijk} p_{ijk}}{N_{jk}} \right) \right)^2,$$

$$S_{ijk}^2 = \frac{M_{ijk}}{M_{ijk} - 1} p_{ijk} (1 - p_{ijk}).$$

Après quelques opérations algébriques, on obtient

$$\begin{aligned} \text{Var}(y_k) = & \frac{1}{n_{1k}} \left\{ N_{1k}^2 S_{1k}^2 + \left(\frac{1-f_1}{f_1}\right) N_{1k} \sum_{i=1}^{N_{1k}} M_{i1k} S_{i1k}^2 + \frac{N_{2k}^2 S_{2k}^2}{\lambda_k} \right. \\ & \left. + \left(\frac{1-f_1}{f_1}\right) \frac{N_{2k}}{\lambda_k} \sum_{i=1}^{N_{2k}} M_{i2k} S_{i2k}^2 \right\} - \sum_{j=1}^2 N_{jk} S_{jk}^2. \end{aligned}$$

On peut donc représenter  $CV^2(y_k)$  sous la forme suivante:

$$CV^2(y_k) = \frac{\text{Var}(y_k)}{Y_k^2} = \frac{A_k}{n_{1k}} - B_k. \quad (3.1)$$

Par ailleurs,  $B$ , le biais relatif, et  $\hat{B}$ , son estimateur, sont donnés par

$$B = \frac{T_0}{T_1} = \frac{\sum_{i=1}^{N_0} M_{i0} p_{i0}}{\sum_{k=1}^{N_p} Y_k},$$

$$\hat{B} = \frac{t_0}{t_1} = \frac{\sum_{i=1}^{n_0} \frac{N_0}{n_0} M_{i0} p_{i0}}{\sum_{k=1}^{N_p} y_k},$$

où  $M_{i0}$  est le nombre de personnes "non" dans le SD  $i$  de la province, et  $p_{i0}$  est la probabilité d'une caractéristique d'intérêt pour une personne "non" du SD  $i$ .

En supposant que  $t_0$  et  $t_1$  sont indépendants, on a

$$\text{Var}(\hat{B}) = B^2 \left( \frac{\text{Var}(t_0)}{T_0^2} + \frac{\text{Var}(t_1)}{T_1^2} \right). \quad (3.2)$$

Après quelques opérations algébriques, si  $f_0$  est la fraction de sondage dans les SD "non", on obtient

$$\text{Var}(t_0) = \frac{1}{n_0} \left\{ N_0^2 S^2 + N_0 \left( \frac{1-f_0}{f_0} \right) \sum_{i=1}^{N_0} M_{i0} S_i^2 \right\} - N_0 S^2$$

où

$$S^2 = \frac{1}{N_0 - 1} \sum_{i=1}^{N_0} \left( M_{i0} p_{i0} - \left( \sum_{i=1}^{N_0} \frac{M_{i0} p_{i0}}{N_0} \right) \right)^2, \quad S_i^2 = \frac{M_{i0}}{M_{i0} - 1} p_{i0} (1 - p_{i0})$$

ce qui peut être représenté sous la forme suivante:

$$\text{Var}(t_0) = \frac{A_0}{n_0} - B_0. \quad (3.3)$$

D'autre part, en supposant les  $y_k$  indépendants, on a

$$\text{Var}(t_1) = \sum_{k=1}^{N_p} \text{Var}(y_k).$$

En considérant l'équation (3.1), on a

$$\text{Var}(t_1) = \sum_{k=1}^{N_p} \left( \frac{A_k Y_k^2}{n_{1k}} - B_k Y_k^2 \right). \quad (3.4)$$

De (3.2), (3.3) et (3.4), on obtient

$$\begin{aligned} \text{Var}(\hat{B}) &= B^2 \left\{ \left( \frac{A_0}{n_0 T_0^2} - \frac{B_0}{T_0^2} \right) + \sum_{k=1}^{N_p} \left( \frac{A_k Y_k^2}{n_{1k} T_1^2} - \frac{B_k Y_k^2}{T_1^2} \right) \right\} \\ &= \frac{1}{n_0} \left( \frac{B^2 A_0}{T_0^2} \right) + \sum_{k=1}^{N_p} \frac{1}{n_{1k}} \left( \frac{B^2 A_k Y_k^2}{T_1^2} \right) - B^2 \left( \frac{B_0}{T_0^2} + \sum_{k=1}^{N_p} \frac{B_k Y_k^2}{T_1^2} \right) \\ &= \frac{X}{n_0} + \sum_{k=1}^{N_p} \frac{W_k}{n_{1k}} - Z. \end{aligned}$$

On peut réexprimer le problème d'optimisation comme étant celui de minimiser

$$\sum_{k=0}^{N_p} c_k n_k$$

en tenant compte des contraintes

$$0 < a_k \leq n_k \leq b_k \quad (k = 0, 1, 2, \dots, N_p) \quad (3.5)$$

et

$$\sum_{k=0}^{N_p} d_k / n_k = e \quad (3.6)$$

où, pour  $k = 1, 2, \dots, N_p$ ,

$$n_k = n_{1k}, \quad c_k = c_{1k} + c_{2k} \lambda_k, \quad a_k = \frac{A_k}{CV_*^2 + B_k}, \quad b_k = \min(N_{1k}, N_{2k} / \lambda_k).$$

En pratique, plutôt que de prendre  $b_k = \min(N_{1k}, N_{2k} / \lambda_k)$  on prend

$$b_k = \frac{N_{1k} N_{2k} (1 + \lambda_k)}{\lambda_k^2 N_{1k} + N_{2k}},$$

puis si  $n_k > N_{1k}$ , alors on prend  $n_{1k} = N_{1k}$  et

$$n_{2k} = \lambda_k n_k + \frac{(n_k - N_{1k}) N_{2k}}{N_{1k} \lambda_k},$$

tandis que si  $\lambda_k n_k > N_{2k}$  alors on prend  $n_{2k} = N_{2k}$  et

$$n_{1k} = n_k + \frac{(\lambda_k n_k - N_{2k}) N_{1k} \lambda_k}{N_{2k}}.$$

On considère donc que  $N_{2k} / (N_{1k} \lambda_k)$  petits SD équivalent à un grand SD. Il y a, en moyenne, autant de personnes handicapées dans un grand SD que dans  $N_{2k} / (N_{1k} \lambda_k)$  petits.

En procédant ainsi, on ne respecte pas toujours la relation  $n_{2k} = \lambda_k n_{1k}$ . Toutefois, on évite de se contenter d'un CV supérieur à celui visé, lorsque, par exemple, il reste des petits SD à observer (même si tous les grands SD ont été sélectionnés).

Il se peut que pour certaines valeurs de  $k$ , on ait  $a_k \geq b_k$ . Dans un tel cas, on prend  $n_k = b_k$ . Soient

$$E_1 = \{k = 0, 1, 2, \dots, N_p \mid n_k = a_k\},$$

$$E_2 = \{k = 0, 1, 2, \dots, N_p \mid n_k = b_k > a_k\},$$

$$E_3 = \{k = 0, 1, 2, \dots, N_p \mid a_k < n_k < b_k\},$$

$$E_4 = \{k = 0, 1, 2, \dots, N_p \mid n_k = b_k \leq a_k\},$$

la solution existe si

$$\sum_{k=0}^{N_p} d_k / b_k \leq e,$$

et elle est de la forme

$$n_k = \begin{cases} a_k & (k \in E_1) \\ b_k & (k \in E_2 \cup E_4) \\ K (d_k / c_k)^{1/2} & (k \in E_3) \end{cases} \quad (3.7)$$

où

$$K = \frac{\sum_{k \in E_3} (d_k / c_k)^{1/2}}{e - \sum_{k \in E_1} d_k / a_k - \sum_{k \in E_2 \cup E_4} d_k / b_k}, \quad (3.8)$$

puisque les  $n_k$  ( $k \in E_3$ ) minimisent  $\sum_{k \in E_3} c_k n_k$ , sous la contrainte

$$\sum_{k \in E_3} d_k / n_k = e - \sum_{k \in E_1} d_k / a_k - \sum_{k \in E_2 \cup E_4} d_k / b_k.$$

Quels sont les ensembles  $E_1$ ,  $E_2$ ,  $E_3$  et  $E_4$  qui correspondent à la solution? L'ensemble  $E_4$  est facile à déterminer. Par ailleurs, on doit avoir

$$a_k < (d_k / c_k)^{1/2} K < b_k \quad (k \in E_3), \quad (d_k / c_k)^{1/2} K \geq b_k \quad (k \in E_2),$$

$$(d_k / c_k)^{1/2} K \leq a_k \quad (k \in E_1). \quad (3.9)$$

Il s'agirait donc d'essayer chacune des possibilités pour  $E_1$ ,  $E_2$  et  $E_3$  jusqu'à ce qu'on obtienne une valeur de  $k$  qui satisfasse (3.9). Pour réduire le nombre de possibilités à examiner, remarquons que si pour  $k' \geq k$ ,

$$b'_k (c'_k / d'_k)^{1/2} \geq b_k (c_k / d_k)^{1/2} \quad (k, k' \in \{0, 1, \dots, N_p\}), \quad (3.10)$$

alors il existe  $k^*$  tel que  $E_2 = \{0, 1, 2, \dots, k^*\}$ , ou bien  $E_2 = \{ \}$ . Tandis que si pour  $k' \geq k$ ,

$$a'_k (c'_k / d'_k)^{1/2} \geq a_k (c_k / d_k)^{1/2} \quad (k, k' \in \{0, 1, \dots, N_p\}), \quad (3.11)$$

alors il existe  $k^{**}$  tel que  $E_1 = \{k^{**}, k^{**} + 1, \dots, N_p\}$  ou bien  $E_1 = \{ \}$ .

### 3.3 Estimation des paramètres

Afin d'effectuer le calcul de l'allocation optimale de l'échantillon, les quantités suivantes doivent être évaluées:

- $P_1$  = proportion des individus ayant été sélectionnés lors de l'ESLA et qui ont répondu "oui" à la question 20 du recensement;
- $P_2$  = proportion des individus n'ayant pas été sélectionnés lors de l'ESLA et qui ont répondu "oui" à la question 20 du recensement;
- $P_3$  = proportion des individus ayant été sélectionnés lors de l'ESLA et qui ont répondu "non" à la question 20 du recensement.

Puisque ces paramètres ne peuvent être calculés directement à partir des données de l'Enquête sur la santé et les incapacités au Canada, un test qu'on a appelé "l'étude de calibration" a été effectué en septembre et octobre 1985.

La question 20 du recensement fut posée intégralement comme question supplémentaire lors de l'Enquête sur la population active (EPA) de septembre. Elle s'adressait à un échantillon d'environ 36,000 individus. Les questions sur les 17 AVQ et une question sur les handicaps mentaux constituaient un supplément à l'EPA d'octobre et elles étaient demandées aux mêmes individus.

Pour chacun des groupes d'âges de cinq ans, on s'est servi des valeurs pondérées de l'étude de calibration afin d'estimer la probabilité d'une réponse positive  $P$  (*oui*) à la question 20 du recensement. Le questionnaire de sélection de l'ESLA diffère de celui de l'étude de calibration. Dans l'ESLA, il y a plus de questions sur les troubles mentaux ou psychologiques et la partie (a) de la question 20 du recensement est posée de nouveau. Donc, on ne s'est pas appuyé seulement sur l'étude de calibration pour calculer les paramètres.



## 4. FICHER GÉOGRAPHIQUE ET DÉMOGRAPHIQUE POUR 1986

### 4.1 Description des informations disponibles

Lors de l'allocation de l'échantillon qui a eu lieu au printemps de 1986, on disposait des données suivantes pour établir les projections de la population par groupe d'âge et par SD:

1. projection de la population par groupe d'âge et par province en 1986;
2. estimé de la population par groupe d'âge et par DR en 1984;
3. population par groupe d'âge et par SD en 1981;
4. un fichier de conversion pour établir la correspondance entre les SD de 1981 et ceux de 1986;
5. estimé du nombre de logements par SD en 1986.

La façon dont est construit le fichier de conversion repose sur le concept d'ensembles correspondants. Chaque ensemble correspondant est la plus petite région formée de SD, dont les frontières sont demeurées inchangées. Par exemple, si trois SD de 1981 ont servi à former deux SD de 1986, ce groupe de trois SD de 1981 (ou deux SD de 1986) forme un ensemble correspondant.

Les quatre méthodes décrites dans la sous-section qui suit visent à établir des projections de la population par groupe d'âge et par ensemble correspondant en 1986. Si un ensemble correspondant est formé de plusieurs SD de 1986, la population projetée pour cet ensemble correspondant peut être répartie entre ces SD de façon proportionnelle en utilisant l'estimé du nombre de logements par SD en 1986.

### 4.2 Méthodes d'estimation

Pour la province  $p$ , notons

$$EC_{l,k} = \text{le } l\text{-ième ensemble correspondant de la } k\text{-ième DR } (l = 1, 2, \dots, N_k; \\ k = 1, 2, \dots, N_p),$$

$$EC_{l,k;81}(j) = \text{population de } EC_{l,k} \text{ dans le } j\text{-ième groupe d'âge en 1981 } (j = 1, 2, \dots, 16),$$

$$DR_{k;84}(j) = \text{estimé de la population de la } k\text{-ième DR dans le } j\text{-ième groupe d'âge en 1984,}$$

$$\hat{P}_{86}(j) = \text{projection de la population dans le } j\text{-ième groupe d'âge dans la province en 1986.}$$

Pour les trois méthodes qui suivent, la première étape consiste à calculer  $\hat{DR}_{k;86}(j)$ , c'est-à-dire la projection de la population de la  $k$ -ième DR dans le  $j$ -ième groupe d'âge en 1986. On suppose qu'il existe des  $K'_j$  ( $j = 1, 2, \dots, 16$ ) tels que

$$\hat{DR}_{k;86}(j) = K'_j(\hat{DR}_{k;84}(j)) \quad (k = 1, 2, \dots, N_p; j = 1, 2, \dots, 16),$$

$$\sum_{k=1}^{N_p} \hat{DR}_{k;86}(j) = \hat{P}_{86}(j) \quad (j = 1, 2, \dots, 16).$$

Cela implique que

$$\hat{DR}_{k;86} = \frac{\hat{P}_{86}(j) DR_{k;84}(j)}{\sum_{k=1}^{N_p} DR_{k;84}(j)}.$$

La première méthode pour estimer  $EC_{l,k;86}(j)$  consiste à supposer l'existence de  $K_j$  ( $j = 1, \dots, 16$ ) tels que

$$\widehat{EC}_{l,k;86}(j) = K_j EC_{l,k;81}(j) \quad (l = 1, \dots, N_k; j = 1, \dots, 16),$$

$$\sum_{l=1}^{N_k} \widehat{EC}_{l,k;86}(j) = \widehat{DR}_{k;86}(j) \quad (j = 1, 2, \dots, 16).$$

On dira que cette méthode utilise le modèle simple. On obtiendra

$$\widehat{EC}_{l,k;86}(j) = \frac{\widehat{DR}_{k;86}(j) EC_{l,k;81}(j)}{\sum_{l=1}^{N_k} EC_{l,k;81}(j)} \quad (l = 1, \dots, N_k; j = 1, \dots, 16).$$

Avec ce modèle simple, l'estimation de la population totale de  $EC_{l,k}$  en 1986 est

$$\sum_{j=1}^{16} \frac{\widehat{DR}_{k;86}(j) EC_{l,k;81}(j)}{\sum_{l=1}^{N_k} EC_{l,k;81}(j)}.$$

Si on croit être en mesure de fournir une meilleure estimation,  $\widehat{EC}_{l,k;86}(tot)$  de cette quantité par des moyens indépendants (par exemple, à partir de l'estimé du nombre de logements dans  $EC_{l,k}$  en 1986), alors on peut utiliser des modèles plus élaborés pour estimer  $EC_{l,k;86}(j)$ . Le modèle multiplicatif est précisé par les équations suivantes:

$$\widehat{EC}_{l,k;86}(j) = K_j (EC_{l,k;81}(j)) + e_l' \quad (l = 1, \dots, N_k; j = 1, \dots, 16),$$

$$\sum_{l=1}^{N_k} \widehat{EC}_{l,k;86}(j) = K (\widehat{DR}_{k;86}(j)) \quad (j = 1, \dots, 16),$$

$$\sum_{l=1}^{N_k} e_l' = 0,$$

$$\sum_{j=1}^{16} \widehat{EC}_{l,k;86}(j) = \widehat{EC}_{l,k;86}(tot) \quad (l = 1, \dots, N_k).$$

On peut interpréter  $e_l$  comme étant la migration intra-DR nette du  $l$ -ième ensemble correspondant.

Le troisième modèle utilisé, appelé modèle additif, est donné par les équations suivantes:

$$\widehat{EC}_{l,k;86}(j) = EC_{l,k;81}(j) + e_l + f_j \quad (l = 1, \dots, N_k; j = 1, \dots, 16),$$

$$\sum_{l=1}^{N_k} \widehat{EC}_{l,k;86}(j) = \widehat{DR}_{k;86}(j) + D \quad (j = 1, \dots, 16),$$

$$\sum_{l=1}^{N_k} e_l = D,$$

$$\sum_{j=1}^{16} \widehat{EC}_{l,k;86}(j) = \widehat{EC}_{l,k;86}(tot) \quad (l = 1, \dots, N_k).$$

On suppose donc que les accroissements (ou diminutions) de population pour chaque groupe d'âge des ensembles correspondants de la DR peuvent se décomposer en deux termes: un qui ne dépend que de l'ensemble correspondant et non de l'âge ( $e_l$ ), et un qui ne dépend que de l'âge et pas de l'ensemble correspondant ( $f_j$ ).

Un dernier modèle, trivial celui-là, serait simplement de poser

$$\widehat{EC}_{l,k;86}(j) = EC_{l,k;81}(j) \quad (l = 1, 2, \dots, N_k; j = 1, \dots, 16).$$

### 4.3 Évaluation des méthodes d'estimation

Les quatre méthodes ont été évaluées en utilisant des données pour la période allant de 1976 à 1981. On a utilisé la projection établie en 1976 de la population par groupe d'âge et par province en 1981 ( $\widehat{P}_{81}(j)$ ), la population par groupe d'âge et par SD en 1976, un fichier de conversion 1976-1981 et l'estimé pré-censitaire du nombre de logements par SD en 1981. Des estimations pour la population par groupe d'âge et par DR en 1979 (l'équivalent de  $DR_{k;84}(j)$ ) n'existant pas, on a posé

$$\widehat{DR}_{k;81} = \frac{\widehat{P}_{81}(j) \widehat{DR}_{k;84}(j)}{\sum_{k=1}^{N_p} \widehat{DR}_{k;84}(j)}.$$

Pour  $\widehat{EC}_{l,k;81}(tot)$ , nécessaire pour les modèles multiplicatif et additif, on a utilisé

$$\widehat{EC}_{l,k;81}(tot) = \frac{\sum_{j=1}^{16} EC_{l,k;76}(j) \sum_{j=1}^{16} \widehat{DR}_{k;81}(j)}{\sum_{l=1}^{N_k} \sum_{j=1}^{16} EC_{l,k;76}(j)}.$$

**Tableau 1**  
Comparaison des quatres modèles

| Prov.   | $EFF_S$ | $EFF_M$ | $EFF_A$ |
|---------|---------|---------|---------|
| T.-N.   | 0.890   | 0.891   | 0.887   |
| Î.P.-É. | 0.903   | 0.914   | 0.919   |
| N.-É.   | 0.960   | 0.972   | 0.912   |
| N.-B.   | 0.869   | 0.868   | 0.884   |
| Qué.    | 0.778   | 0.764   | 0.818   |
| Ont.    | 0.932   | 0.930   | 0.916   |
| Man.    | 0.892   | 0.904   | 0.912   |
| Sask.   | 0.732   | 0.749   | 0.801   |
| Alb.    | 0.818   | 0.827   | 0.860   |
| C.-B.   | 0.713   | 0.716   | 0.775   |
| Yukon   | 0.770   | 0.768   | 0.840   |
| T.N.-O. | 1.252   | 1.246   | 1.157   |

On a ensuite calculé pour chaque province  $p$ , une mesure d'efficacité des modèles simple, multiplicatif et additif par rapport au modèle trivial:

$$EFF_m = \frac{\sum_{k=1}^{N_p} \sum_{l=1}^{N_k} \sum_{j=1}^{16} \left( (\widehat{EC}_{l,k;81}^{(m)}(j) - EC_{l,k;81}(j)) \right)^2}{\sum_{k=1}^{N_p} \sum_{l=1}^{N_k} \sum_{j=1}^{16} \left( (\widehat{EC}_{l,k;81}^{(T)}(j) - EC_{l,k;81}(j)) \right)^2} \quad (m = S, M, A),$$

où  $\widehat{EC}_{l,k;81}^{(m)}(j)$  avec  $m = S, M, A$  et  $T$  sont les projections obtenues à l'aide des modèles simple, multiplicatif, additif et trivial respectivement. Le tableau 1 présente des valeurs obtenues.

Le modèle simple donne les pires résultats pour une province et un territoire, le modèle multiplicatif pour deux provinces, et le modèle additif pour sept provinces et un territoire.

Le modèle simple se révèle le meilleur pour cinq provinces, le modèle multiplicatif pour deux provinces et un territoire, et le modèle additif pour trois provinces et un territoire.

Le modèle simple ayant aussi, comme son nom l'indique, l'avantage d'être simple, c'est le modèle qui a été retenu.

#### 4.4 Méthode de stratification selon la taille des secteurs de dénombrement

Si on faisait un tirage aléatoire simple pour sélectionner les SD à l'intérieur de chaque région infra-provinciale (RIP), alors les personnes handicapées faisant partie d'un SD avec de nombreuses personnes handicapées auraient moins de chance d'être sélectionnées que celles faisant partie d'un petit SD, c'est-à-dire un SD avec peu de personnes handicapées. Pour éviter de trop grandes différences au niveau des probabilités de sélection, dans chaque RIP on stratifie la population des SD selon le nombre de personnes handicapées qui s'y trouve, puis on utilise l'allocation proportionnelle, c'est-à-dire que dans chaque strate, le nombre de SD choisis est proportionnel au nombre de personnes handicapées.

À partir des résultats d'enquêtes précédentes, on a établi un lien entre la pyramide des âges de la population d'un SD et le nombre de personnes handicapées qu'on peut s'attendre à trouver dans ce SD. Le nombre de personnes handicapées étant inconnu, la variable utilisée pour la stratification et l'allocation de l'échantillon est le nombre attendu de personnes handicapées.

Dans le cas présent, on n'a que deux strates: celle des grands SD et celle des petits SD. Puisqu'on utilise l'allocation proportionnelle, pour déterminer d'une manière optimale la taille frontière au-delà de laquelle un SD est grand et en deçà de laquelle il est petit, on s'est servi d'un critère qu'on peut trouver dans Raj (1968). La taille frontière doit être égale à la moyenne de la taille moyenne des petits SD et de la taille moyenne des grands SD.

## 5. SÉLECTION DE L'ÉCHANTILLON

Il fallait choisir des échantillons pour les trois populations (enfants, adultes et personnes âgées) parmi les grands et les petits SD "oui" de chaque RIP, tant pour la base trois que pour la base cinq, et parmi les SD "non" de chaque province pour la base cinq. Lorsque, dans une RIP, il y avait moins de deux grands SD ou moins de deux petits SD, on sélectionnait tous les SD de cette RIP pour les trois populations. Les échantillons "oui" et "non" étaient créés indépendamment en utilisant l'algorithme à un passage décrit par Bebbington (1975). Les échantillons des trois populations, pour la composante "oui" et pour la composante "non", étaient emboîtés afin de minimiser le nombre total de SD choisis.

Au tableau suivant figurent les tailles des échantillons par province pour chaque groupe d'âge.

**Tableau 2**  
Tailles d'échantillons par province et par groupe d'âge

| Province | Enfants                         |                                 | Adultes                         |                                 | Personnes âgées                 |                                 |
|----------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
|          | Nombre de SD "oui" sélectionnés | Nombre de SD "non" sélectionnés | Nombre de SD "oui" sélectionnés | Nombre de SD "non" sélectionnés | Nombre de SD "oui" sélectionnés | Nombre de SD "non" sélectionnés |
| T.-N.    | 880                             | 136                             | 405                             | 154                             | 476                             | 173                             |
| Î.P.-É.  | 242                             | 242                             | 111                             | 217                             | 82                              | 166                             |
| N.-É.    | 1257                            | 157                             | 434                             | 130                             | 438                             | 115                             |
| N.B.     | 1142                            | 162                             | 459                             | 146                             | 453                             | 138                             |
| Qué.     | 4749                            | 153                             | 1070                            | 114                             | 1488                            | 133                             |
| Ont.     | 6085                            | 158                             | 1304                            | 116                             | 1542                            | 120                             |
| Man.     | 1082                            | 203                             | 457                             | 169                             | 367                             | 144                             |
| Sask.    | 2291                            | 265                             | 942                             | 241                             | 921                             | 193                             |
| Alb.     | 2762                            | 190                             | 909                             | 176                             | 1389                            | 222                             |
| C.-B.    | 3117                            | 170                             | 752                             | 125                             | 948                             | 119                             |

## 6. DISCUSSION

L'enquête post-censitaire est une méthode de sondage relativement nouvelle qui est appelée à connaître de nombreux développements au cours des prochaines années. Ce type d'enquête permet beaucoup de flexibilité dans la collecte des données et l'utilisation d'échantillons de grande taille dispersés partout à travers le pays avec des coûts et des délais convenables. L'Enquête sur la santé et les limitations d'activités constitue une première expérience dans ce domaine au Canada pour une enquête de cette importance.

Le plan d'échantillonnage exposé dans le présent article représente une tentative pour maximiser l'emploi des possibilités offertes par l'approche post-censitaire avec une utilisation optimale des ressources disponibles. Le contrôle de la taille de l'échantillon demeure une des difficultés majeures inhérentes à la méthode proposée. La détermination de l'allocation de l'échantillon étant effectuée avant le recensement, tous les calculs doivent être faits à partir de projections basées sur le recensement antérieur. Dans ce contexte, la taille d'un échantillon constitué d'un ensemble de petites régions choisies suivant les résultats de ces projections risque de connaître des variations considérables une fois concrétisée dans le véritable recensement.

On se trouve alors placé devant, d'une part, la possibilité d'obtenir un échantillon insuffisant pour les exigences de qualité des estimations et, d'autre part, l'éventualité de dépasser les ressources allouées à la collecte des données. Afin d'obvier à ces difficultés, la stratégie suivante a été mise en oeuvre. Un nombre cible d'interviews pour chaque population a été calculé pour l'échantillon "oui". Ce nombre était basé sur la taille d'échantillon requise pour produire des estimés répondant à nos critères de qualité. Cependant, un nombre de SD supérieur à celui nécessaire pour obtenir le nombre cible d'interviews a été sélectionné, et cela pour les raisons mentionnées précédemment. Si le nombre réel d'interviews à faire, tel que calculé sur le terrain, est supérieur au nombre cible, alors, pour des raisons de coût, un sous-échantillon de SD est exclu de l'enquête. Ce n'est que pour le bureau régional de Halifax (recouvrant les provinces du Nouveau-Brunswick, de la Nouvelle-Écosse et de l'Île-du-Prince-Édouard) que le nombre d'interviews de l'échantillon "oui" était passablement plus élevé que le nombre cible. Il fut donc décidé d'exclure certains SD de cette partie de l'échantillon. Pour savoir quels SD seraient exclus, il fallait connaître le nombre cible et le nombre réel d'interviews pour chaque SD. Pour 40% des SD, on a dû imputer le nombre réel d'interviews car il n'était pas disponible à temps.

Pour ce faire, on connaissait le total, par district de commissaire au recensement, des nombres réels d'interviews; la portion de ce total qui n'avait pas déjà été utilisée a été distribuée parmi les SD qui requéraient une imputation, proportionnellement au nombre cible d'interviews.

On a ensuite calculé la différence, pour chaque population, entre le nombre réel et le nombre cible d'interviews pour chacune des deux strates de chaque RIP. Une différence positive (réel-cible) indiquait une population pour laquelle des SD pouvaient être exclus de l'enquête. Dans chaque strate, les SD ont été classifiés en trois groupes (1, 2 et 3) selon qu'ils étaient sélectionnés pour trois, deux ou une seule des populations respectivement. Le fichier des SD a ensuite été trié par strate et par groupe de façon ascendante, l'ordre des SD à l'intérieur d'un groupe étant aléatoire. Chaque SD a été considéré successivement et était supprimé pour les trois populations si:

- 1) une différence positive demeurait non négative après suppression du SD;
- 2) une différence négative n'était pas réduite davantage.

De cette façon, chaque différence positive a été réduite à un nombre aussi près que possible de zéro compte tenu de l'ordre aléatoire des SD.

## REMERCIEMENTS

Les auteurs remercient D.A. Binder pour sa contribution à l'élaboration du plan de sondage et P. Reed pour la partie informatique de l'allocation de l'échantillon. Les auteurs aimeraient également remercier les arbitres pour leurs observations utiles.

## APPENDICE

### Question n° 20 du formulaire 2B du recensement

20. a) Êtes-vous limité(e) dans vos activités à cause d'une incapacité physique, d'une incapacité mentale ou d'un problème de santé chronique: (consultez le guide)

À la maison?

non, je ne suis pas limité(e)

oui, je suis limité(e)

À l'école ou au travail?

non, je ne suis pas limité(e)

oui, je suis limité(e)

sans objet

Dans d'autres activités, par ex., dans vos trajets entre la maison et votre lieu de travail ou dans vos loisirs?

non, je ne suis pas limité(e)

oui, je suis limité(e)

b) Avez-vous des incapacités ou handicaps à long terme?

non

oui

### Questions de sélection pour l'ESLA (questionnaire des adultes)

1. Éprouvez-vous des difficultés à entendre ce qui se dit au cours d'une conversation normale avec une autre personne?
2. Éprouvez-vous des difficultés à entendre ce qui se dit au cours d'une conversation en groupe avec au moins trois autres personnes?
4. Éprouvez-vous des difficultés à lire les caractères ordinaires d'un journal (avec des verres si vous en portez habituellement)?
5. Éprouvez-vous des difficultés à voir clairement la figure de quelqu'un à 12 pieds/4 mètres (par ex., d'un bout à l'autre d'une pièce) avec des verres si vous en portez habituellement?
7. Éprouvez-vous des difficultés à parler et être compris(e)?
8. Éprouvez-vous des difficultés à marcher sur une distance de 400 verges/mètres sans vous reposer (environ trois pâtés de maisons)?
9. Éprouvez-vous des difficultés à monter et descendre un escalier (environ 12 marches)?
10. Éprouvez-vous des difficultés à transporter un objet de 10 livres sur une distance de 30 pieds/5 kilogrammes sur 10 mètres (par ex., un sac d'épicerie)?
11. Éprouvez-vous des difficultés à vous déplacer d'une pièce à une autre?
12. Éprouvez-vous des difficultés à vous tenir debout pendant de longues périodes, c'est-à-dire pendant plus de 20 minutes? Rappelez-vous qu'il s'agit de problèmes qui devraient durer 6 mois ou plus.
13. Lorsque vous êtes debout, éprouvez-vous des difficultés à vous pencher et à ramasser un objet sur le plancher (par ex., un soulier)?

14. Éprouvez-vous des difficultés à vous habiller et vous déshabiller?
15. Éprouvez-vous des difficultés à vous mettre au lit et à sortir du lit?
16. Éprouvez-vous des difficultés à vous couper les ongles d'orteils?
17. Éprouvez-vous des difficultés à vous servir de vos doigts pour saisir ou manier un objet?
18. Éprouvez-vous des difficultés à étendre le bras dans n'importe quelle direction pour prendre quelque chose (par ex., au dessus de votre tête)?
19. Éprouvez-vous des difficultés à couper vos aliments?
20. À cause d'une affection ou un problème de santé chronique qui devrait durer 6 mois ou plus, êtes-vous limité(e) dans le genre ou la quantité d'activités que vous pouvez faire . . .  
(i) à la maison? (ii) à l'école ou au travail? (iii) dans vos autres occupations comme les déplacements, les sports ou les loisirs?
21. Un professionnel de l'enseignement ou de la santé vous a-t-il déjà dit que vous aviez des difficultés d'apprentissage?
22. De temps à autre, chacun éprouve des difficultés à se souvenir du nom d'une personne familière ou à apprendre quelque chose de nouveau ou il nous arrive d'être confus pendant quelques instants. Toutefois, avez-vous en permanence des problèmes de mémoire ou d'apprentissage?
23. À cause d'une affection ou d'un problème chronique d'ordre émotif, psychologique, nerveux ou mental, êtes-vous limité(e) dans le genre ou la quantité d'activités que vous pouvez faire . . .  
(i) à la maison? (ii) à l'école ou au travail? (iii) dans vos autres occupations comme les déplacements, les sports ou les loisirs?

## BIBLIOGRAPHIE

- BEBBINGTON, A.C. (1975). A simple method of drawing a sample without replacement. *Applied Statistics*, 24, 136.
- CARTER, R.G., GILES, P.D., et SHERIDAN, M.J. (1982). Description and rationale for the screen tests for the January 1983 Disability Survey. Projet d'établissement d'une base de données sur les invalides, Division de la santé, Statistique Canada.
- CHAMBRE DES COMMUNES (1981). *Obstacles*, Rapport du Comité spécial concernant les invalides et les handicapés. Ottawa.
- COCHRAN, W.G. (1977). *Sampling Techniques*, (3<sup>e</sup> éd.). New York: John Wiley.
- DOLSON, D., GILES, P., et MORIN, J.-P. (1984). Méthode d'enquête sur les personnes souffrant d'une incapacité à l'aide de questions supplémentaires de l'Enquête sur la population active. *Techniques d'enquête*, 10, 203-214.
- DOLSON, D., McCLEAN, K., MORIN, J.-P., et THÉBERGE, A. (1986). Rapport de méthodologie de l'ESLA. Document de travail, Statistique Canada.
- GRABOWIECKI, F. (1982). Discussion of the target population for the Disability Survey. Projet d'établissement d'une base de données sur les invalides, Division de la santé, Statistique Canada.
- GRABOWIECKI, F. (1983). Contenu des données de l'enquête de Statistique Canada auprès des handicapés. Rapport technique, Division de la santé, Statistique Canada.
- LAZARUS, G., et NESICH, R. (1985). A report on the methodology of the Canadian Health and Disability Survey. Document de travail, Statistique Canada.



- McDOWELL, I. (1981). Un examen des questions proposées par l'OECD dans le cadre de l'enquête canadienne. *Revue d'épidémiologie et de santé publique*, 29, 412-429.
- MORIN, J.-P., et DOWLER, L. (1986). Proposition d'une méthodologie pour l'ESLA-institutions. Document de travail, Statistique Canada.
- MORIN, J.-P. (1986). Comparaison initiale de l'ESIC et de l'ESG. Document de travail, Statistique Canada.
- ORGANISATION MONDIALE DE LA SANTÉ, (1980). Classification internationale des déficiences, incapacités et handicaps. Genève, Suisse.
- RAJ, D. (1968). *Sampling Theory*. New York: McGraw-Hill.
- WILSON, R.W., et McNEIL, J.M. (1981). Analyse préliminaire de l'incapacité au sein de l'OCDE, d'après le test initial de l'enquête post-censitaire. *Revue d'épidémiologie et de santé publique*, 29, 469-475.