

Table des matières

1	INTRODUCTION	1
2	CONTEXTE	1
3	OBJECTIFS	3
4	CONCEPTS ET DÉFINITIONS	4
4.1	Concepts et définitions de l'Enquête sur la population active	4
4.2	Définitions et concepts de l'Enquête sur le tabagisme chez les jeunes	6
5	MÉTHODOLOGIE DE L'ENQUÊTE	8
5.1	Composante EPA (supplément)	8
5.1.1	Couverture de la population	8
5.1.2	Plan d'échantillonnage de l'EPA	8
5.1.3	Taille de l'échantillon de l'EPA	12
5.1.4	Renouvellement de l'échantillon de l'EPA	12
5.1.5	Modifications au plan de l'EPA aux fins de l'enquête supplémentaire	13
5.1.6	Taille de l'échantillon, par province, pour l'enquête supplémentaire	14
5.2	Composante École	15
5.2.1	Base de sondage	15
5.2.2	Couverture de la population	15
5.2.3	Stratification	16
5.2.4	Répartition de l'échantillon	16
5.2.5	Sélection des écoles	17
5.2.6	Sélection des classes	17
5.2.7	Traitement des cas spéciaux	17
5.2.8	Remplacement pour la non-réponse	18
6	COLLECTE DES DONNÉES	19
6.1	Composante EPA (supplément)	19
6.2	Composante École	21
7	TRAITEMENT DES DONNÉES	25
7.1	Saisie des données	25
7.2	Contrôle	25
7.2.1	Contrôle du supplément ETJ-EPA (groupe des 15-19 ans)	25
7.2.2	Contrôle de la composante École de l'ETJ (groupe des 10-14 ans)	26
7.3	Codage des questions à réponse libre	26
7.3.1	Marque de cigarettes	27
7.3.2	Problèmes de santé	27
7.3.3	Messages concernant les risques pour la santé	28
7.3.4	Commandite d'événements sportifs et culturels	29

7.4	Création de variables calculées	29
7.5	Pondération	30
7.6	Suppression des renseignements confidentiels	30
8	QUALITÉ DES DONNÉES	32
8.1	Taux de réponse	32
8.1.1	Composante EPA	32
8.1.2	Composante École	32
8.2	Erreurs associées à l'enquête	35
8.2.1	Non-réponse totale	36
8.2.2	Non-réponse partielle	36
8.2.3	Couverture	36
8.2.4	Critères de mesure de l'erreur d'échantillonnage	37
9	LIGNES DIRECTRICES CONCERNANT LA TOTALISATION, L'ANALYSE ET DIFFUSION	38
9.1	Lignes directrices concernant l'arrondissement	38
9.2	Lignes directrices concernant la pondération de l'échantillon aux fins de la totalisation	39
9.2.1	Définition des types d'estimations : catégoriques et quantitatives	39
9.2.2	Totalisation des estimations catégoriques	40
9.2.3	Totalisation des estimations quantitatives	41
9.3	Lignes directrices concernant l'analyse statistique	41
9.4	Lignes directrices concernant la diffusion des C.V.	42
10	TABLEAUX SUR LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE	44
10.1	Comment utiliser les tableaux C.V. aux fins des estimations catégoriques	48
10.2	Exemples de l'utilisation des tableaux sur les c.v. aux fins des estimations catégoriques	50
10.3	Comment utiliser les tables de c.v. pour obtenir les limites de confiance	53
10.4	Exemple d'utilisation des tableaux sur les c.v. pour obtenir les limites de confiance	54
10.5	Comment utiliser les tableaux sur les c.v. pour effectuer un test t	55
10.6	Exemple d'utilisation des tableaux sur les c.v. pour effectuer un test t	55
10.7	Coefficients de variation pour les estimations quantitatives	56
10.8	Seuils de diffusion à l'égard de l'Enquête sur le tabagisme chez les jeunes	56
10.9	Tableaux sur les c.v.	60
11	PONDÉRATION	61
11.1	Pondération pour la composante EPA	61
11.1.1	Procédures de pondération pour l'EPA	61
11.1.2	Procédures de pondération pour l'Enquête sur le tabagisme chez les jeunes	63
11.2	Pondération pour la composante Écoles	65

12	PLAN DE L'ENREGISTREMENT	67
13	QUESTIONNAIRES ET FEUILLES DE CODES	68
13.1	Composante EPA (supplément) de l'ETJ (Formule 08)	68
13.2	Composante École de l'ETJ (Formule 08S)	91
13.3	Questionnaire à l'intention des parents de la composante École (Formule 03S)	111
13.4	Dossier du ménage (Formule 03) et Feuille de codes	115
13.5	Questionnaire de l'Enquête sur la population active (Formule 05) et Feuille de codes	118

1 INTRODUCTION

L'Enquête sur le tabagisme chez les jeunes (ETJ) a été menée par Statistique Canada à l'automne de 1994, pour le compte de Santé Canada. Elle visait principalement à recueillir de l'information sur l'usage du tabac chez les jeunes dans les dix provinces du Canada. C'est la première fois qu'une enquête nationale sur le tabagisme chez les jeunes est menée au Canada.

Cette enquête porte sur les jeunes de 10 à 19 ans. Les données sur les personnes de 10 à 14 ans ont été recueillies au moyen d'un échantillon d'élèves choisis dans des salles de classe de tout le Canada. Les jeunes de 15 à 19 ans ont été interviewés au moyen d'un supplément à l'Enquête sur la population active. Le contenu du questionnaire était pratiquement le même pour les deux composantes; les méthodes de collecte des données, cependant, différaient sensiblement.

Le présent manuel a pour but de faciliter l'utilisation du fichier de microdonnées sur les résultats de l'enquête. Toute question concernant le fichier ou son utilisation devrait être adressée à :

Statistique Canada

Joan Coulter
Division des enquêtes spéciales, Statistique Canada
5^e étage, Immeuble Jean-Talon
Parc Tunney
Ottawa (Ontario) K1A 0T6
Téléphone : (613) 951-3261
Télécopieur : (613) 951-0562

Santé Canada

Margaret Morin
Bureau de contrôle du tabac
Santé Canada
5^e étage, Tour A
11, av. Holland
Localisateur postal 3005B
Ottawa (Ontario) K1A 0K9
Téléphone : (613) 954-0152
Télécopieur : (613) 941-1551

IL EST IMPORTANT QUE LES UTILISATEURS SE FAMILIARISENT AVEC LE CONTENU DU PRÉSENT DOCUMENT AVANT DE PROCÉDER À LA MISE EN TABLEAU, À L'ANALYSE, À LA PUBLICATION OU À LA DIFFUSION DE TOUTE ESTIMATION CALCULÉE À PARTIR DU FICHIER DE MICRODONNÉES DE L'ENQUÊTE SUR LE TABAGISME CHEZ LES JEUNES.

2 CONTEXTE

Depuis 1964, Santé Canada recueille de l'information concernant l'usage du tabac chez les Canadiens au moyen d'un supplément à l'Enquête sur la population active (EPA) et, plus récemment, au moyen de l'Enquête sur le tabagisme au Canada. La plupart des enquêtes sur l'usage du tabac chez les Canadiens n'ont porté que sur les adultes de 15 ans et plus. On accusait donc un manque de données sur les jeunes Canadiens. L'enquête vise à satisfaire le besoin de connaître les tendances d'utilisation et les attitudes des jeunes à l'égard des produits du tabac.

ETJ et politiques et programmes de Santé Canada

Dans le cadre du récent Plan d'action national du gouvernement canadien pour lutter contre la contrebande du tabac, les taxes sur le tabac ont été réduites en février 1994. Puisque des prix moins élevés peuvent inciter les jeunes à fumer davantage, les modifications législatives et réglementaires suivantes ont été promulguées ou le seront prochainement :

- abolir la fabrication des «mini-paquets» qui visent les jeunes consommateurs;
- hausser l'âge légal pour l'achat de cigarettes;
- augmenter les amendes imposées pour la vente de cigarettes à des mineurs;
- limiter l'installation de distributeurs automatiques aux bars, aux tavernes et aux autres débits de boisson semblables; et
- renforcer l'impact des avertissements concernant les risques pour la santé.

On a également lancé une campagne d'éducation du public sur l'usage du tabac. Parmi les composantes de la campagne, citons :

- une campagne médiatique nationale pour sensibiliser les jeunes aux effets nocifs du tabagisme;
- les nouveaux efforts déployés pour rejoindre les familles, les nouveaux parents et toute autre personne dont le comportement sert de modèle aux enfants;
- l'appui aux programmes d'éducation scolaire; et
- les efforts accrus pour rejoindre les jeunes femmes qui commencent à fumer.

L'Enquête sur le tabagisme chez les jeunes (ETJ) constitue un élément clé de la collecte de données de référence que Santé Canada utilisera pour évaluer l'usage du tabac et planifier sa ligne de conduite.

3 OBJECTIFS

L'Enquête sur le tabagisme chez les jeunes (ETJ) est une enquête nationale destinée à recueillir des données de référence aux échelles nationale (Yukon et Territoires du Nord-Ouest exclus) et provinciale relativement aux connaissances, aux attitudes et aux comportements des jeunes Canadiens à l'égard d'une vaste gamme de questions concernant l'usage du tabac. Dans le but d'appuyer les efforts nationaux et provinciaux pour prévenir et réduire l'usage du tabac chez les jeunes Canadiens, il est nécessaire de recueillir des données d'enquête qui permettent d'établir des estimations sur la prévalence du tabagisme dans chaque région et d'effectuer des analyses sur les facteurs sociaux associés au tabagisme.

L'ETJ a été conçue précisément dans le but de recueillir de l'information sur les sujets suivants :

- la prévalence du tabagisme chez les jeunes de 10 à 19 ans;
- les types d'usages du tabac chez les jeunes (p. ex. usage du tabac pour expérimenter, usage occasionnel, usage régulier);
- les facteurs sociaux et démographiques associés au tabagisme (p. ex. ce qui motive les jeunes à fumer, l'influence de la famille et des pairs);
- où et comment les jeunes obtiennent des cigarettes;
- les attitudes et les croyances à l'égard du tabagisme;
- la sensibilisation aux risques pour la santé;
- l'incidence des politiques anti-tabac dans les écoles et le milieu de travail;
- la perception du parrainage de programmes culturels et d'événements sportifs par les sociétés productrices de tabac;
- la reconnaissance d'une marque chez les jeunes et leur réaction face à l'emballage neutre; et
- des données démographiques de base sur les ménages, notamment la composition du ménage, la participation à la population active, l'industrie et la profession des parents.

Les données recueillies à partir de l'Enquête sur le tabagisme chez les jeunes aideront Santé Canada à évaluer et à élaborer des programmes d'éducation du public visant à informer les jeunes Canadiens des risques pour la santé qui sont associés au tabagisme.

Cette enquête fournira également aux décideurs et aux chercheurs des secteurs de la promotion et de la protection de la santé des résultats sommaires précis, à jour et clairs.

4 CONCEPTS ET DÉFINITIONS

Dans le présent chapitre, on énonce des concepts et des définitions qui présentent un intérêt pour les utilisateurs des données relatives à l'Enquête sur le tabagisme chez les jeunes. La section 4.1 expose les concepts et définitions utilisés dans l'Enquête sur la population active, tandis que la section 4.2 porte sur les concepts et définitions propres à l'Enquête sur le tabagisme chez les jeunes. Vous trouverez, au chapitre 12 du présent document, un exemplaire des questionnaires qui ont été utilisés durant l'enquête.

4.1 Concepts et définitions de l'Enquête sur la population active

Situation au regard de l'activité

Situation du répondant sur le marché du travail : un membre de la population hors institutions de 15 ans et plus désigné comme **personne occupée, personne sans emploi ou personne ne participant pas à la population active.**

Personne occupée

Les personnes occupées sont celles qui, durant la semaine de référence :

- a) ont fait un travail¹ quelconque
- b) avaient un travail, mais étaient absents en raison :
 - d'une maladie ou d'une incapacité
 - de responsabilités personnelles ou familiales
 - du mauvais temps
 - d'un conflit de travail
 - de vacances
 - d'autres raisons non précisées ci-dessus (sauf les personnes mises en disponibilité ou celles dont le travail commence à une date ultérieure précise).

Chômeurs

Les chômeurs sont celles qui, durant la période de référence :

¹ Le travail comprend tout travail rémunéré ou en vue d'un bénéfice, c'est-à-dire le travail rémunéré dans le cadre d'une relation employeur-employé ou d'un travail autonome. Il comprend également le travail familial non rémunéré qui se définit comme étant un travail non rémunéré qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'une pratique professionnelle appartenant à un membre apparenté du ménage ou exploitée par lui. De telles activités peuvent comprendre la tenue de livres, la vente de produits, le service aux tables, etc. Les tâches comme les travaux et l'entretien ménagers ne sont pas considérées comme un travail familial non rémunéré.

- a) ne travaillaient pas, avaient recherché activement du travail au cours des quatre dernières semaines (se terminant avec la semaine de référence) et étaient prêtes à travailler²;
- b) n'avaient pas cherché activement du travail au cours des quatre dernières semaines, mais avaient été mises à pied³ et étaient prêtes à travailler;
- c) n'avaient pas cherché activement du travail au cours des quatre dernières semaines, mais devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence, et étaient prêtes à travailler.

Inactifs

Il s'agit de la partie de la population civile hors institutions âgée de 15 ans et plus qui était ni occupée ni en chômage pendant la semaine de référence.

Hors du champ de l'enquête

Les personnes faisant partie de la population résidente non civile ou non permanente durant la période de référence.

Branche d'activité et profession

L'Enquête sur la population active fournit des renseignements sur la profession et la branche d'activité des personnes occupées, des chômeurs et inactifs qui ont occupé un emploi au cours des cinq dernières années. Depuis 1984, ces statistiques sont fondées sur la Classification type des professions de 1980 et sur la Classification type des industries de 1980. Avant 1984, les données étaient groupées selon la Classification des professions du recensement de 1971 et la Classification des activités économiques de 1970.

Semaine de référence

-
- ² Les personnes de ce groupe répondant aux critères suivants sont réputées prêtes à travailler :
 - (i) les étudiants à plein temps qui cherchaient un emploi à temps partiel et qui satisfaisaient également la condition (ii) ci-dessous. (Les étudiants à plein temps qui cherchent un travail à plein temps étaient réputés non prêts à travailler durant la période de référence.)
 - (ii) celles qui ont déclaré n'avoir aucune raison de ne pas prendre un emploi durant la période de référence et celles qui ne le pouvaient pas en raison d'une «maladie ou incapacité», de «responsabilités personnelles ou familiales» ou «parce qu'elles avaient déjà un travail».
 - ³ Les personnes sont classées comme étant mises à pied seulement lorsqu'elles prévoient reprendre le travail qu'elles ont dû quitter pour mise en disponibilité.

Semaine civile entière couverte chaque mois par l'Enquête sur la population active. Cette semaine comprend habituellement le quinzième jour du mois. Les interviews sont menées durant la semaine suivante, appelée Semaine de l'enquête, et la situation de la population active qui est établie est celle de la semaine de référence.

4.2 Définitions et concepts de l'Enquête sur le tabagisme chez les jeunes

Les définitions des catégories de tabagisme sont fondées sur les catégories proposées dans le document de Mills, Stephens et Wilkins intitulé *Rapport sommaire de l'atelier sur les données relatives au contrôle du tabagisme*, lequel a paru dans *Maladies chroniques au Canada* à l'été de 1994. On a apporté certaines modifications mineures aux catégories proposées dans le document afin de les adapter à la population des 10 à 19 ans.

Fume présentement chaque jour

A fumé au moins 100 cigarettes dans sa vie et a fumé au moins une cigarette par jour au cours des 30 jours précédant l'enquête.

Fume présentement, mais non quotidiennement

A fumé au moins 100 cigarettes dans sa vie et a fumé au moins une cigarette au cours des 30 derniers jours, mais n'a pas fumé chaque jour.

Fume présentement - fréquence inconnue

A fumé au moins 100 cigarettes dans sa vie et a fumé au moins une cigarette au cours des 30 derniers jours, mais la fréquence est inconnue.

Fumait autrefois quotidiennement

A fumé 100 cigarettes ou plus dans sa vie, mais n'a pas fumé du tout au cours des 30 derniers jours, et a fumé à un moment donné tous les jours durant sept jours consécutifs.

Fumait autrefois, mais non quotidiennement

A fumé 100 cigarettes ou plus dans sa vie, mais n'a pas fumé du tout au cours des 30 derniers jours, et n'a jamais fumé tous les jours durant sept jours consécutifs.

Fumait autrefois - fréquence inconnue

A fumé 100 cigarettes ou plus dans sa vie, mais n'a pas fumé du tout au cours des 30 derniers jours, et la fréquence autrefois est inconnue.

Commence à fumer

A fumé entre 1 et 99 cigarettes dans sa vie, et fumé au cours des 30 derniers jours.

A déjà fait l'expérience

A fumé entre 1 et 99 cigarettes dans sa vie, mais n'a pas fumé au cours des 30 derniers jours.

N'a jamais fumé

A fumé moins d'une cigarette entière durant sa vie.

5 MÉTHODOLOGIE DE L'ENQUÊTE

L'Enquête sur le tabagisme chez les jeunes a été menée en deux étapes. Les données sur les personnes de 10 à 14 ans ont été recueillies au moyen d'un échantillon d'étudiants établi dans les salles de classe de tout le Canada, tandis que les jeunes de 15 à 19 ans ont été interviewés par téléphone dans le cadre d'une enquête supplémentaire de l'Enquête sur la population active.

5.1 Composante EPA (supplément)

L'Enquête sur le tabagisme chez les jeunes a été menée en septembre 1994 auprès d'un échantillon élargi de ménages tirés de l'échantillon de l'Enquête sur la population active (EPA); son plan d'échantillonnage est donc étroitement lié à celui de l'EPA. Le plan de l'EPA est décrit brièvement dans les sections 5.1.1 à 5.1.4⁴. Les sections 5.1.5 et 5.1.6 décrivent comment la composante EPA (supplément) de l'Enquête sur le tabagisme chez les jeunes s'écarte du plan de base de l'EPA.

5.1.1 Couverture de la population

La population cible de l'enquête supplémentaire EPA-ETJ englobait tous les jeunes de 15 à 19 ans. Étaient précisément exclus de la couverture de l'enquête les membres à plein temps des Forces armées canadiennes, les détenus dans les établissements pénitentiaires ainsi que les résidents du Yukon et des Territoires du Nord-Ouest et les personnes vivant sur une réserve indienne. Ces personnes représentent dans l'ensemble une exclusion de 2 % environ de la population de 15 ans et plus. Cette exclusion s'explique par le fait qu'on a utilisé l'EPA comme cadre d'échantillonnage pour l'ETJ. L'EPA consiste en une enquête mensuelle auprès des ménages dont l'échantillon de personnes est représentatif de la population civile hors institutions de 15 ans et plus dans les dix provinces du Canada.

5.1.2 Plan d'échantillonnage de l'EPA

L'échantillon de l'EPA se fonde sur un plan stratifié à plusieurs degrés utilisant une méthode d'échantillonnage probabiliste à tous les degrés du plan. Les principes à l'égard du plan sont les mêmes pour chaque province. Un schéma résumant les degrés du plan est présenté à la fin de la présente section.

Stratification primaire

Les provinces sont d'abord stratifiées en régions économiques -- régions géographiques dont la structure économique est plus ou moins homogène et qui ont été formées en vertu

⁴ On peut se procurer une description détaillée du plan de l'EPA auprès de Statistique Canada, dans une publication intitulée **Méthodologie de l'Enquête sur la population active au Canada, 1984-1990** (publication n° 71-526 au catalogue).

d'ententes fédérales-provinciales. Les régions économiques sont relativement stables dans le temps.

Ces régions économiques constituent des strates primaires, et on effectue une stratification additionnelle à l'intérieur de celles-ci (voir section 5.2.3).

Types de secteurs

Les régions économiques sont à leur tour soumises à une désagrégation en trois catégories: les unités autoreprésentatives (UAR), les unités non autoreprésentatives (UNAR) et les unités spéciales. Les UAR comprennent généralement les régions urbaines dont la population au Recensement de 1981 dépassait 15 000 personnes et dont les caractéristiques particulières de la population active justifiaient la création d'une UAR. Les limites de l'UAR coïncident généralement avec les délimitations établies aux fins du Recensement.

Toutes les UAR de chaque région économique sont comprises dans l'enquête et, comme leur nom l'indique, elles sont représentées par leur propre échantillon.

Les UNAR sont les régions situées à l'extérieur des UAR et comprennent principalement des régions rurales et de petits centres urbains. Chaque région économique contient une UNAR qui est représentée par son propre échantillon.

Une faible proportion (1 % environ) de la population de l'EPA se trouve en établissement (par exemple, le personnel résidant d'hôpitaux ou d'écoles, ou les résidents permanents d'hôtels ou de motels), sur les bases militaires (personnel civil seulement) ou dans des régions éloignées des provinces qui ne sont pas facilement accessibles pour les intervieweurs de l'EPA. Pour des raisons administratives, cette portion de la population est échantillonnée séparément au moyen de la catégorie des unités spéciales. Cette partie de l'échantillon est choisie dans toute la province, sans référence à la stratification utilisée pour les secteurs des UAR et des UNAR.

Stratification secondaire

Les UAR sont ensuite délimitées individuellement en strates qui regroupent des secteurs ayant une situation socio-économique semblable d'après les données du Recensement de 1981. La portée de la stratification (nombre de strates) dépend de la taille de l'UAR.

Dans les régions économiques où la population d'une UNAR constitue une proportion importante de la population de la région économique, l'UNAR est ensuite délimitée en strates urbaines et rurales distinctes. À l'intérieur de chacune des strates, une stratification additionnelle est effectuée pour tenir compte des différences entre un certain nombre de caractéristiques de la population active.

Dans les unités spéciales, des strates sont formées au niveau provincial. Les strates reflètent les principaux types de groupes spéciaux de la population qui requièrent des mesures d'échantillonnage administratives particulières. Ce sont les établissements

militaires, les institutions et les régions éloignées.

Délimitation et sélection de grappes

Dans chacune des strates secondaires qu'on trouve dans les secteurs d'une UAR, un certain nombre de groupes contigus géographiquement de logements ou de grappes sont formés à partir d'une combinaison des chiffres du Recensement de 1981 et du dénombrement sur le terrain. Les grappes correspondent généralement aux îlots des villes et aux côtés d'îlots. La sélection d'un échantillon de grappes (généralement de six ou douze grappes) dans chacune des strates secondaires représente la première étape de l'échantillonnage dans les secteurs des UAR.

Dans chacune des strates secondaires des régions des UNAR, un certain nombre de grandes régions géographiques sont délimitées de façon à ce que chacune reflète la composition de la strate dans laquelle elle est située à l'égard d'un certain nombre de caractéristiques socio-économiques. Deux ou quatre de ces secteurs, appelés unités primaires d'échantillonnage (UPÉ) sont choisis dans l'échantillon de chaque strate secondaire. Dans chaque UPÉ choisie, un certain nombre de plus petits groupes de logements ou de grappes contigus géographiquement sont ensuite formés en utilisant des caractéristiques physiques bien définies que l'on peut reconnaître sur les cartes et sur le terrain.

Dans les secteurs spéciaux, les secteurs de dénombrement du recensement (régions géographiques couvertes par des enquêteurs du Recensement) constituent la première étape de la sélection. Parmi les groupes sélectionnés, s'il y a lieu, des groupes de logements contigus ou des grappes sont formés, et la sélection d'un échantillon de ceux-ci constitue la seconde étape de l'échantillonnage.

Sélection des logements

Dans les trois types de secteurs (UAR, UNAR et secteurs spéciaux), les grappes choisies sont d'abord visitées par les enquêteurs sur le terrain qui préparent une liste de tous les logements privés situés dans la grappe. On choisit ensuite un échantillon de six logements (en moyenne) à partir de la liste établie. Cela représente l'étape finale de l'échantillonnage.

Dans les 17 plus grandes UAR, un échantillon d'appartements situés dans de grands immeubles d'habitation est sélectionné à partir d'un registre distinct fondé sur des renseignements fournis par la SCHL. Cette mesure a pour but d'assurer une meilleure représentation des personnes qui habitent en appartement et de minimiser l'incidence de la croissance des grappes en raison de la construction de nouveaux immeubles d'habitation.

Sélection des personnes

On obtient des données démographiques pour toutes les personnes dont le logement choisi constitue leur lieu de résidence habituel. Les données de l'EPA sont recueillies auprès de tous les membres civils de ménages qui ont 15 ans ou plus.

EPA - PLAN DE L'ÉCHANTILLON

À chaque étape du plan de l'échantillon, on fait appel à des méthodes d'échantillonnage aléatoires afin de s'assurer que l'échantillon est aléatoire tout en étant représentatif de la population visée.

Le plan de l'échantillon est le même pour chaque province.

Chaque province est constituée d'un certain nombre de régions économiques, soit de secteurs de structure économique semblable constitués sur la base des ententes fédérale-provinciales.

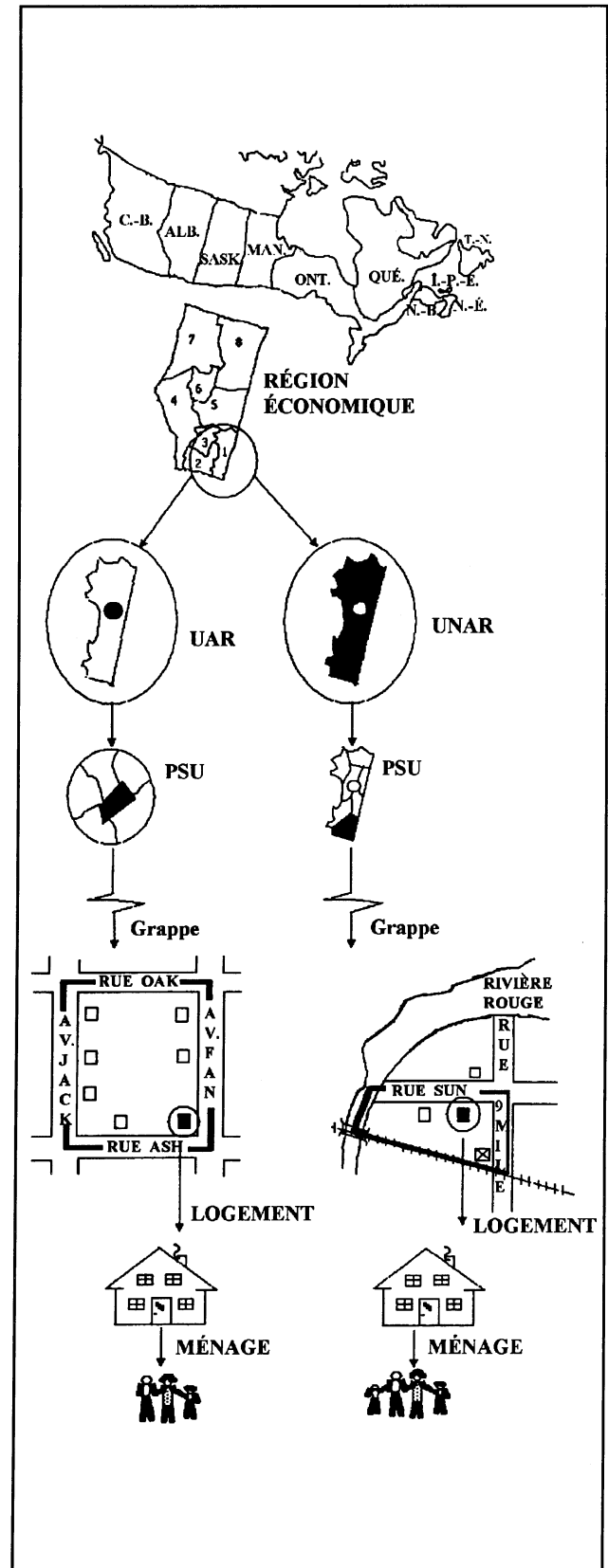
Chaque région économique est divisée en unités autoreprésentatives (UAR), en unités non autoreprésentatives (UNAR) et en secteurs spéciaux. Les UAR sont des villes de plus de 15,000 habitants ou des villes dont les caractéristiques justifient qu'elles soient considérées comme des unités autoreprésentatives. Les UNAR sont les régions qui se situent à l'extérieur des UAR. Les secteurs spéciaux sont constitués par les établissements militaires, les hôpitaux et autres institutions, et les régions éloignées.

Les UAR et les UNAR sont délimitées en unités primaires d'échantillonnage (UPÉ), qui sont des secteurs qu'un interviewer peut visiter facilement. Un échantillon d'UPÉ est ensuite sélectionné.

Les UPÉ sélectionnées sont alors délimitées en grappes de logements qui correspondent à des îlots ou à des côtés d'îlots (dans les secteurs urbains) et qui correspondent à des frontières physiques identifiables (dans les secteurs ruraux). On sélectionne ensuite un échantillon des grappes et les agents recenseurs sur le terrain dressent la liste de tous les logements privés se trouvant dans les grappes sélectionnées.

À l'intérieur de chaque grappe sélectionnée, un échantillon des logements est sélectionné à partir de la liste des logements.

À l'intérieur de chaque logement sélectionné, l'EPA permet de recueillir les renseignements correspondant à chaque membre du ménage de 15 ans ou plus.



à chaque membre du ménage de 15 ans ou plus.

5.1.3 Taille de l'échantillon de l'EPA

Dans l'EPA, la taille de l'échantillon de personnes admissibles est déterminée de façon à répondre aux critères de précision statistique établis pour diverses caractéristiques de la population active aux échelles provinciale et infraprovinciale, à satisfaire les exigences des administrations fédérale, provinciales et municipales et à répondre aux besoins d'une multitude d'autres utilisateurs des données.

L'échantillon mensuel de l'EPA compte 68 000 logements environ. Après avoir exclu les logements vides, démolis ou convertis à des usages non résidentiels, ainsi que les logements ne comportant que des personnes inadmissibles, en construction et saisonniers, il reste quelque 58 000 logements occupés par une ou plusieurs personnes admissibles. À partir de ces logements, on obtient des renseignements sur quelque 112 000 citoyens âgés de plus de 15 ans.

5.1.4 Renouvellement de l'échantillon de l'EPA

L'EPA utilise un plan d'échantillons permanents selon lequel l'échantillon mensuel complet de logements est composé de six échantillons permanents, ou groupes de renouvellement, dont la taille est à peu près égale. Chacun de ces échantillons permanents est réputé, en lui-même, être représentatif de la population entière de l'EPA. Tous les logements faisant partie d'un groupe de renouvellement demeurent dans l'échantillon de l'EPA pendant six mois consécutifs, après quoi ils sont remplacés (retirés de l'échantillon) par un nouvel échantillon permanent de logements choisi à partir des mêmes grappes ou de grappes semblables. Les échantillons permanents sont échelonnés, et un nouvel échantillon permanent est utilisé chaque mois; un échantillon permanent est retiré de l'échantillon chaque mois et remplacé par un nouveau.

Ce modèle de renouvellement a été adopté pour garantir que l'échantillon de logements reflète de façon constante les changements qui touchent l'ensemble d'habitations courant et pour minimiser tout problème de non-réponse ou le fardeau qui serait imposé au répondant si les ménages devaient demeurer dans l'échantillon pendant plus de six mois. Cette méthode présente également l'avantage statistique de fournir une base d'échantillonnage commune pour les comparaisons de mois en mois et à court terme des caractéristiques de l'EPA.

Grâce à la méthode du renouvellement de l'échantillon, on peut facilement mener des enquêtes supplémentaires en utilisant le plan de l'EPA pour avoir un échantillon dont la taille est plus ou moins complète.

5.1.5 Modifications au plan de l'EPA aux fins de l'enquête supplémentaire

L'Enquête sur le tabagisme chez les jeunes a nécessité cinq groupes de renouvellement tirés de l'échantillon de l'EPA de septembre 1994, sauf en Ontario, où on a utilisé trois groupes de renouvellement. En vue de produire des estimations à l'échelle provinciale, l'échantillon de l'ETJ a été augmenté dans toutes les provinces, sauf en Ontario et au Québec, par l'addition de groupes de renouvellement qui avaient déjà été retirés de l'échantillon de l'EPA (voir le tableau à la page suivante). Bien que l'EPA recueille de l'information auprès de tous les membres admissibles de ménages ayant 15 ans et plus, l'Enquête sur le tabagisme chez les jeunes ne visait que les membres du ménage ayant entre 15 et 19 ans au 1^{er} septembre 1994. Les réponses par procuration n'étaient pas permises.

On a aussi modifié le plan d'échantillonnage de l'ETJ de façon à minimiser le chevauchement avec d'autres enquêtes supplémentaires qui avaient lieu en même temps. Dans l'un des cinq groupes de renouvellement de l'EPA de septembre 1994, on a interviewé les jeunes de 15 à 19 ans seulement si aucun des membres du ménage n'avait moins de deux ans. Dans trois des quatre groupes de renouvellement restants de l'EPA de septembre 1994, on a interviewé les jeunes de 15 à 19 ans seulement si aucun des membres du ménage n'avait moins de 12 ans. Pour le dernier groupe de renouvellement choisi à partir de l'EPA de septembre 1994, aucune restriction ne s'appliquait. De même, aucune restriction n'a été appliquée sur les groupes de renouvellement supplémentaires utilisés pour augmenter l'échantillon de l'ETJ et qui avaient déjà été retirés de l'EPA.

Province	Nombre de groupes de renouvellement tirés de l'EPA de septembre 1994	Nombre de groupes de renouvellement tirés d'EPA précédentes	Nombre total de groupes de renouvellement échantillonnés
Terre-Neuve	5	6	11
Île-du-Prince-Édouard	5	6	11
Nouvelle-Écosse	5	6	11
Nouveau-Brunswick	5	6	11
Québec	5	0	5
Ontario	3	0	3
Manitoba	5	6	11
Saskatchewan	5	6	11
Alberta	5	4	9
Colombie-Britannique	5	5	10

5.1.6 Taille de l'échantillon, par province, pour l'enquête supplémentaire

Le tableau suivant indique le nombre de jeunes de 15 à 19 ans qui sont inclus dans les échantillons de renouvellement de l'EPA et qui étaient admissibles à l'enquête supplémentaire, soit l'Enquête sur le tabagisme chez les jeunes.

PROVINCE	TAILLE DE L'ÉCHANTILLON
Terre-Neuve	1 155
Île-du-Prince-Édouard	447
Nouvelle-Écosse	1 186
Nouveau-Brunswick	1 124
Québec	1 444
Ontario	1 138
Manitoba	1 242
Saskatchewan	1 352
Alberta	1 288
Colombie-Britannique	1 327
CANADA	11 703

5.2 Composante École

Le plan d'échantillonnage du volet des écoles de l'Enquête sur le tabagisme chez les jeunes consiste en un plan par grappes stratifié à deux degrés où les écoles constituent les unités primaires d'échantillonnage et les classes les unités secondaires. Tous les élèves d'une classe sélectionnée se retrouvent dans l'échantillon.

5.2.1 Base de sondage

On a utilisé une base-liste d'écoles couvrant l'ensemble des écoles publiques et privées du Canada. Des renseignements assez complets sur les effectifs des écoles selon le niveau scolaire et l'âge des élèves étaient disponibles sur cette base de données. Ces renseignements dataient de l'année scolaire 1991-92. On n'avait cependant pas d'information géographique à propos des écoles, à part la province où celles-ci étaient situées. On n'avait pas non plus de renseignements à propos des classes dans les écoles.

5.2.2 Couverture de la population

La population-cible consiste en l'ensemble des jeunes de 10 à 14 ans demeurant au Canada, à l'exception de ceux demeurant au Yukon et dans les Territoires du Nord-Ouest, et de

ceux vivant en institution ou dans les réserves indiennes. La population effectivement enquêtée diffère quelque peu de la population-cible. Ces différences peuvent être regroupées dans les catégories suivantes:

- A) Pour des raisons opérationnelles, on a exclu les jeunes fréquentant des écoles spéciales (e.g. pour aveugles, sourd-muets) ainsi que les écoles situées sur les bases militaires.
- B) Seuls les jeunes fréquentant les niveaux 5 à 9 ont été considérés pour l'enquête. Le faible nombre de jeunes âgés de 10 ans en 4^{ième} année, ou de 14 ans en 10^{ième} année (au moment de la collecte des données) a motivé cette décision.
- C) Les jeunes fréquentant des classes de petite taille, c'est-à-dire comptant moins de 10 élèves au total, ont été exclus.

On estime que l'ensemble des exclusions représente au total environ 8% de la population-cible.

5.2.3 Stratification

Le plan d'échantillonnage comporte deux niveaux de stratification. Tout d'abord, chaque province représente une strate principale. Ensuite, on a utilisé une stratification implicite par niveau scolaire (de 5 à 9 inclusivement, donc 5 strates secondaires par province) en procédant à la sélection de l'échantillon dans chaque niveau de manière indépendante.

5.2.4 Répartition de l'échantillon

L'échantillon d'écoles a été réparti également entre les provinces. Les exigences de précision des résultats étaient représentées par une proportion minimale estimable (0.10) assortie d'un coefficient de variation maximal (16.5%), au niveau province et sexe, donc pour le groupe d'âge complet des 10-14 ans.

Pour déterminer premièrement la taille d'échantillon nécessaire dans chaque province, on a dû faire certaines hypothèses à propos des taux de réponse à plusieurs niveaux (commission scolaire, écoles, parents, enfants), ainsi qu'à propos de la taille moyenne d'une classe et de la perte éventuelle des élèves non-éligibles (i.e. n'ayant pas l'âge requis). Les taux de réponse espérés ont été déduits des résultats de l'enquête-pilote qui avait eu lieu l'année précédente. On a supposé que la taille moyenne d'une classe était de 30 élèves. Pour chaque province, on a calculé qu'un échantillon de 80 écoles était suffisant pour satisfaire aux exigences de précision. L'échantillon total d'écoles était donc de 800 écoles.

Dans chaque province, on a réparti l'échantillon d'écoles également entre les niveaux scolaire d'intérêt. On a donc sélectionné exactement 16 écoles de chaque niveau scolaire.

5.2.5 Sélection des écoles

L'échantillon d'écoles a été sélectionné de manière systématique avec probabilité proportionnelle à la taille des écoles, c'est-à-dire au nombre total d'étudiants pour chaque niveau. Afin de garantir une bonne représentativité par taille des commissions scolaires, ainsi que pour les écoles elles-mêmes, le fichier des écoles a été trié par taille de commission scolaire, premièrement, et par taille d'école à l'intérieur de chaque commission scolaire, deuxièmement. Ensuite, on a calculé le pas de sondage approprié et procédé à une sélection systématique des écoles.

Pour l'Ile-du-Prince-Edouard, certaines écoles étaient tellement importantes en termes d'effectifs qu'on les a sélectionnées avec certitude dans l'échantillon. Techniquement, ceci correspond en fait à la création d'une strate additionnelle pour chaque niveau.

Etant donné que la sélection était faite indépendamment pour chaque niveau, une école de grande taille pouvait être sélectionnée plus d'une fois, pour des niveaux différents. Ceci s'est produit surtout dans les petites provinces. Le nombre maximum de fois qu'une école a été ainsi sélectionnée a été quatre fois à l'Ile-du-Prince-Edouard.

5.2.6 Sélection des classes

La sélection des unités secondaires d'échantillonnage, c'est-à-dire les classes, s'est effectuée sur le terrain. L'intervieweur avait comme instruction de dresser une liste de toutes les classes du niveau voulu et de procéder à la sélection d'une de ces classes en utilisant une grille de sélection aléatoire. Tous les élèves de la classe sélectionnée étaient retenus dans l'échantillon final. Pour plus de renseignements sur les procédures de collecte, prière de consulter la section 6.2.

5.2.7 Traitement des cas spéciaux

On définit ici les cas spéciaux comme étant tous les cas où il y a eu changement touchant une école sélectionnée et affectant la collecte des données. Ceci inclut tous les cas de déménagement, fermetures et fusion d'écoles. Dans tous ces cas, il était important d'agir de manière à respecter la nature probabiliste de l'échantillon.

La règle générale à suivre était la suivante. L'intervieweur devait tout d'abord tenter de déterminer à quel école les élèves du niveau d'intérêt avaient été relocalisés. Ensuite, si les deux conditions suivantes étaient respectées: a) les élèves avaient été relocalisés à une nouvelle école (c'est-à-dire n'apparaissant pas sur la base-liste d'écoles), et b) aucun élève d'aucune autre école n'avait été relocalisé à cette nouvelle école, alors la nouvelle école remplaçait l'ancienne dans l'échantillon. Dans le cas contraire, il n'y avait pas de remplacement.

En pratique, l'information à propos de la relocalisation des élèves n'était pas toujours disponible ou complète. La décision de remplacer ou non l'école devenait alors plus complexe. Au total, on a observé 26 cas spéciaux (3.3% de l'échantillon total) pour lesquels 10 remplacements ont été effectués.

5.2.8 Remplacement pour la non-réponse

Dans le but de conserver la taille d'échantillon finale voulue, on a adopté une stratégie de remplacement des écoles pour lesquelles on n'avait pu obtenir la participation à l'enquête. Ce refus de participer pouvait provenir de la commission scolaire ou du principal de l'école. Dans le premier cas, il pouvait y avoir plusieurs écoles d'impliquées par cette décision.

La stratégie consistait, dans le cas du refus de la commission scolaire, à remplacer celle-ci par la commission scolaire dont les effectifs pour les niveaux d'intérêt étaient les plus semblables. La commission scolaire remplaçante devait être située dans la même province et évidemment consentir à participer à l'enquête. Dans un deuxième temps, on devait trouver des écoles dans cette commission scolaire remplaçante pour remplacer celles appartenant à la commission scolaire originale. Encore une fois, la similarité des effectifs a été le critère choisi.

Dans le cas des écoles pour lesquelles un refus était obtenu, on a procédé comme décrit plus haut mais en s'assurant de sélectionner l'école remplaçante dans la même commission scolaire.

Au total, 52 écoles, soit 6.5% de l'échantillon, ont été remplacées de la manière décrite.

6 COLLECTE DES DONNÉES

La collecte des données pour l'Enquête sur le tabagisme chez les jeunes (ETJ) a été divisée en deux composantes : la composante Enquête sur la population active (EPA) pour les jeunes de 15 à 19 ans; et la composante École pour les personnes de 10 à 14 ans. Les interviews ont été menées en vertu des dispositions de la *Loi sur la statistique* concernant la participation volontaire. Les réponses par procuration n'étaient pas permises.

6.1 Composante EPA (supplément)

La collecte des données pour la composante EPA a été menée durant la Semaine d'enquête de l'EPA et durant la Semaine post-enquête de septembre 1994 (du 19 septembre au 1^{er} octobre 1994). Il a fallu consacrer deux semaines à cet exercice pour deux raisons principales. D'abord, l'échantillon comptait à la fois des ménages faisant partie des groupes de renouvellement courants et des groupes qui avaient déjà été retirés de l'EPA. Cela a permis aux intervieweurs de se concentrer sur la collecte des données auprès des ménages faisant partie des groupes de renouvellement courants durant la Semaine de l'enquête et sur la collecte auprès des groupes retirés de l'enquête durant la semaine suivante. Deuxièmement, la période de collecte de deux semaines donnait aux intervieweurs une autre semaine pour communiquer avec les jeunes qui étaient difficiles à rejoindre. (De plus amples renseignements sur le nombre de groupes de renouvellement choisis dans chaque province sont donnés à la section 5.1.5.)

Tous les intervieweurs de l'EPA ont participé à la collecte des données. Ils étaient 870 environ, répartis dans toutes les régions des dix provinces du Canada.

Les intervieweurs avaient une heure pour se préparer à mener l'enquête pendant laquelle ils ont lu le Manuel de l'intervieweur et effectué l'exercice de révision, simulé une interview avec un parent ou un ami et abordé toute question avec leur surveillant avant le début de l'enquête.

La plupart des interviews ont été menées par téléphone puisque le premier groupe de renouvellement (le groupe de renouvellement qui ne fait que commencer, c'est-à-dire le premier des six mois dans l'échantillon de l'EPA) de l'EPA a été exclu de l'échantillon. Le premier groupe de renouvellement a été exclu de l'échantillon de l'ETJ puisque, durant le premier mois, l'EPA est menée au moyen d'interviews en personne. Les interviews avec les ménages qui n'ont pas consenti à des interviews téléphoniques pour l'EPA ont été réalisées au moyen de visites en personne. Une seule visite était permise pour ces ménages. Des méthodes de collecte spéciale pour l'EPA ont été établies pour certaines régions éloignées du pays. Les jeunes de ces régions éloignées n'ont été interviewés que si l'EPA était menée durant le mois de septembre. Les jeunes des colonies huttérites ont été exclus.

On a remis aux intervieweurs une Liste de contrôle de la tâche qui leur permettait

d'identifier chaque ménage et chaque jeune ayant été choisi pour l'interview. Les intervieweurs ont aussi reçu un questionnaire étiqueté de l'ETJ pour chaque jeune choisi. (Remarque : tous les jeunes des groupes de renouvellement choisis n'étaient pas admissibles à l'interview en raison d'un chevauchement avec un certain nombre d'enquêtes supplémentaires et longitudinales. Plus de renseignements à ce sujet sont donnés dans la section 5.1.5).

Dans le cas des groupes de renouvellement courants, les intervieweurs de l'EPA ont dû adopter deux modes de collecte : les interviews assistées par ordinateur pour l'EPA et les interviews papier et crayon pour la composante École. Avant la Semaine d'enquête de l'EPA, les intervieweurs ont identifié chaque ménage faisant partie de la tâche et qui comptait un jeune admissible ou plus; ils ont ensuite inscrit l'acronyme de l'enquête et le nombre de jeunes sélectionnés sous «Note temporaire» sur leur écran «Voir et choisir les cas».

Dans les groupes de renouvellement courants, les intervieweurs ont présenté l'ETJ une fois l'entrevue de l'EPA terminée. L'intervieweur posait alors quelques brèves questions de filtrage au répondant du ménage de l'EPA pour confirmer que le jeune choisi était admissible à l'interview de l'ETJ (âgé entre 15 et 19 ans au 1^{er} septembre 1994) et demandait ensuite à parler au jeune choisi. La plupart des jeunes ont été interviewés au moment de l'interview de l'EPA. Si le jeune n'était pas disponible à ce moment-là, les intervieweurs ne devaient pas faire plus de cinq rappels pour rejoindre le jeune. Si le jeune était temporairement absent du logement et que le répondant du ménage ne pouvait donner un numéro de téléphone pour le joindre, on devait essayer de retracer le jeune. Ces tentatives étaient faites uniquement par téléphone.

Les méthodes de collecte utilisées pour les groupes retirés de l'enquête étaient très semblables. Les intervieweurs appelaient les ménages, donnaient leur nom et présentaient l'ETJ, posaient quelques brèves questions de filtrage à un membre responsable du ménage et demandaient de parler au jeune admissible. Ces ménages n'ont pas participé aux interviews de l'EPA. Les renseignements de l'EPA concernant les parents et les membres du ménage ont été tirés des données recueillies au moment de leur dernier mois de participation. Ici encore, des rappels et des tentatives de localisation ont été effectués pour retracer les jeunes admissibles.

La méthode de collecte de données par téléphone adoptée aux fins de la composante EPA a soulevé trois préoccupations :

1. Les parents permettraient-ils aux intervieweurs de parler à leurs enfants au sujet d'une enquête sur le tabagisme?
2. Le jeune sentirait-il que l'entrevue téléphonique est privée et qu'il peut répondre librement aux questions?

3. Serait-il difficile de joindre des jeunes de ce groupe d'âge par téléphone et en découlerait-il un nombre élevé de non-réponses?

Ces préoccupations n'ont pas causé de réels problèmes durant la collecte des données. Le taux de réponse pour la composante EPA a été légèrement plus élevé que prévu. On visait un taux de réponse de 80 % à l'échelle nationale, et le taux de réponse réel a été de 81,1 % (voir la section 8.1.1 pour plus de détails sur les taux de réponse à l'ETJ). Bien que les commentaires consignés sur la liste de contrôle de la tâche ont révélé que certains parents ont écouté les interviews, c'était rarement le cas.

Au chapitre de la collecte, on se préoccupait également de savoir si les intervieweurs comprenaient et appliqueraient de façon uniforme les diverses règles d'admissibilité pour jusqu'à 11 groupes de renouvellement choisis. Il ne semble pas y avoir eu de problèmes à cet égard.

Un problème est cependant survenu et a été corrigé par la suite. Les jeunes âgés de 14 ans au moment de l'interview de l'EPA (qui devaient avoir 15 ans sous peu et seraient admissibles aux interviews) ont été, dans la plupart des cas, exclus par inadvertance des mesures de sélection. Des étiquettes ont rapidement été produites pour ces jeunes, et la collecte des données a été menée à partir des bureaux régionaux en décembre 1994. On n'a connu aucun problème relativement à la collecte des données auprès des jeunes qui avaient été omis. On a changé la date déterminant l'âge du jeune (15-19 ans), de sorte que l'admissibilité du jeune a été changée du 1^{er} septembre 1994 au 1^{er} décembre pour ce groupe.

6.2 Composante École

La collecte pour la composante École a été menée sur une période de trois mois, de septembre à novembre 1994. La collecte comprenait de nombreuses activités, dont envoyer par la poste une lettre de présentation aux écoles sélectionnées, administrer les méthodes de présélection des classes, mener des interviews avec les enfants dans les salles de classe et faire remplir les questionnaires à l'intention des parents. Ces activités de collecte ont été précédées d'un long processus d'approbation par les commissions scolaires qui a commencé en mai 1994.

En raison du chevauchement avec l'Enquête nationale sur la santé de la population et l'Enquête longitudinale nationale sur les enfants, on a demandé aux bureaux régionaux d'embaucher de nouveaux intervieweurs pour l'ETJ. Toutes les régions ont été en mesure de recruter des intervieweurs expérimentés de l'EPA, sauf la région du Pacifique où la majorité des intervieweurs étaient nouveaux.

Le recrutement était fondé sur une tâche moyenne de trois salles de classe par intervieweur. Si plus d'une salle de classe, et jusqu'à un maximum de quatre, était

sélectionnée dans une école, toutes les salles de classe étaient assignées au même intervieweur. Cette mesure permettait de s'assurer que chaque école ne traitait qu'avec un seul intervieweur et qu'un nombre minimal de visites étaient effectuées. Le recrutement était également fondé sur la distribution géographique des écoles par rapport au lieu de résidence des intervieweurs.

Les intervieweurs ont suivi une autoformation de deux heures au cours de laquelle ils ont lu le Manuel de l'intervieweur et effectué l'exercice de révision, simulé une interview avec un parent ou un ami et abordé toute question avec leur surveillant avant le début de l'enquête. Les nouveaux intervieweurs de la région du Pacifique ont également reçu une formation initiale à l'intention des nouveaux intervieweurs.

Voici un résumé du processus de collecte des données :

Premier contact avec l'école

Peu de temps après que les bureaux régionaux ont envoyé par la poste les lettres de présentation aux écoles choisies, des intervieweurs ont téléphoné à chaque école pour :

- présenter l'ETJ au directeur de l'école;
- obtenir la collaboration du directeur relativement à l'enquête;
- fixer un rendez-vous pour une première visite à l'école; et
- vérifier l'adresse de l'école et obtenir les directions, au besoin.

Un certain nombre de situations se sont produites durant le contact initial qui ont permis de déterminer si la classe (ou même l'école) faisait partie du champ de l'enquête. Par exemple, l'école pouvait ne plus exister. La classe pouvait ne plus exister et les étudiants pouvaient avoir été déplacés dans une autre école. Le directeur de l'école pouvait refuser de participer. On a mis au point une stratégie de remplacement, et chaque cas a été traité individuellement. (Pour plus de renseignements, se reporter à la section 5.2.8).

Durant la collecte, bien des directeurs d'école n'étaient pas au courant des détails de l'enquête parce que l'information ne leur avait pas été transmise par la commission scolaire. Les directeurs d'école ont d'ailleurs déploré le fait qu'ils n'aient pas reçu la documentation envoyée aux commissions scolaires.

Première visite à l'école

À son arrivée à l'école, l'intervieweur devait se présenter au directeur et expliquer brièvement les activités de collecte. Une formule de sélection de classe étiquetée a été utilisée pour vérifier les activités de collecte. Sur la formule était indiqué le niveau scolaire choisi. Si l'école comptait plus d'une classe du niveau choisi, l'intervieweur devait utiliser la grille de sélection sur l'étiquette pour choisir au hasard l'une des classes.

L'intervieweur devait inscrire le nom, la date de naissance et le numéro de téléphone de chacun des élèves présents dans la classe. Pour chaque élève âgé entre 10 et 14 ans au 1^{er} septembre 1994, l'intervieweur avait préparé une trousse contenant une lettre de présentation et une formule de consentement parental que l'élève devait apporter à la maison. Le directeur ou le professeur de la classe devait distribuer les formules de consentement parental et s'assurer qu'elles lui étaient remises dûment remplies. L'intervieweur devait expliquer qu'il reviendrait à l'école une semaine plus tard pour prendre les formules remplies. Dans certains cas, le directeur ne voulait pas divulguer la date de naissance et le numéro de téléphone des élèves. On a mis au point des méthodes pour contourner la difficulté.

Deuxième visite à l'école

Durant la deuxième visite, l'intervieweur devait :

- cueillir les formules de consentement parental dûment remplies; et
- fixer un rendez-vous pour retourner à l'école une semaine plus tard en vue de mener la séance en classe.

Interviews avec les parents

L'intervieweur devait revoir les formules de consentement parental pour déterminer quels enfants étaient admissibles à participer à l'enquête. (L'admissibilité dépendait de l'âge du jeune et du consentement parental). Si une formule de consentement parental n'était pas retournée, l'intervieweur sollicitait un consentement par téléphone. Une entrevue était menée par téléphone avec le parent ou le gardien de chaque enfant admissible. La plupart de ces entrevues ont été menées avant la séance en classe.

Séance en classe (troisième visite à l'école)

En prévision de la séance en classe, l'intervieweur devait préparer un questionnaire pour chaque enfant admissible en fonction de la langue d'entrevue privilégiée que le parent avait notée sur la formule de consentement parental. Pour conserver l'anonymat, le nom de l'enfant n'était pas inscrit sur la formule. Il ne figurait en fait que sur une enveloppe pour garantir que le bon questionnaire avait été remis à l'élève.

En classe, l'intervieweur a suivi la démarche présentée ci-dessous :

- Il se présente aux élèves.
- Il explique le but de l'enquête.
- Il demande au professeur de distribuer les enveloppes aux élèves.
- Il distribue, face en dessous sur chaque bureau, le document concernant la reconnaissance des marques de cigarettes et demande aux élèves de ne pas

- le retourner.
- Il lit à haute voix l'introduction du questionnaire.
 - Il explique que le document ne doit être retourné que pour les quelques dernières questions du questionnaire.
 - Il fait les sept premières questions avec les élèves pour leur montrer comment effectuer les différents types d'entrées.
 - Il explique comment remplir la roue à la question 19.
 - Il dit aux élèves de ne pas placer les questionnaires remplis dans l'enveloppe, mais de les laisser sur leur bureau face en dessous et séparément de l'enveloppe.
 - Il dit aux élèves de ne pas hésiter à lever la main pour poser des questions.
 - Il demande aux élèves de remplir le questionnaire.
 - Il répond à toutes les questions.
 - Il remercie les élèves et l'enseignant de leur collaboration et de leur appui.
 - Il ramasse les questionnaires remplis, la documentation et les enveloppes.

Les séances en classe ont duré en moyenne de 30 à 40 minutes. On demandait aux professeurs de rester dans la classe pour des raisons de discipline, mais de ne pas circuler dans la classe de façon à respecter la confidentialité.

7 TRAITEMENT DES DONNÉES

Le principal résultat de l'Enquête sur le tabagisme chez les jeunes consiste en un fichier de microdonnées «épuré». La présente section résume les étapes de traitement suivies pour produire ce fichier.

7.1 Saisie des données

Les données des questionnaires suivants ont été saisies dans six bureaux régionaux à l'aide du système DC2, installation de saisie et de collecte de données sur UNIX qui a été mise au point par Statistique Canada.

- F08 (Questionnaire ETJ pour les jeunes de 15 à 19 ans - - composante EPA)
- F08S (Questionnaire ETJ pour les jeunes de 10 à 14 ans -- composante École)
- F04S (Formule de sélection des classes pour la composante École)
- F03S (Questionnaire à l'intention des parents pour la composante École)

Avant la saisie, les documents remplis ont été regroupés en lots de 20 questionnaires. Un fichier de contrôle était disponible pour la composante EPA seulement du fait que les données ont été extrapolées à partir des fichiers de l'EPA.

On a procédé à une vérification par échantillon comme mesure de contrôle de la qualité pour les formules F08 et F08S. Aux fins de la saisie des données, un échantillon de lots a été choisi et les champs clés ont été vérifiés, c'est-à-dire qu'ils ont été entrés de nouveau et vérifiés. Les résultats de la vérification ont révélé que les préposés à la saisie des données sont demeurés bien en deçà du seuil d'erreur visé de 2 à 3 %. On a estimé en fait que le taux d'erreur sur le terrain était de 0,075 % pour la composante EPA et de 0,260 % pour la composante École.

Après la saisie des données, les documents ont été envoyés aux Centres de documents fédéraux des Archives nationales.

7.2 Contrôle

7.2.1 Contrôle du supplément ETJ-EPA (groupe des 15-19 ans)

Le contrôle du supplément ETJ-EPA a été appliqué «de haut en bas». On a examiné chaque question pour vérifier la présence d'un code ou de codes valides si plus d'un étaient acceptables. Si aucun n'était présent, un code de «non-réponse» de «9», «99», «999» ou «9999» (non déclaré) a été entré. Quant aux questions pour lesquelles l'enchaînement était valide, le chiffre «6», «96», «996» ou «9996» a été attribué. Lorsqu'une question «passer à»

était laissée en blanc, on vérifiait les deux enchaînements de questions possibles pour établir la direction. Si cela était possible, on attribuait la bonne valeur à la question «passer à». Si le cheminement ne pouvait être déterminé, on attribuait à la question «passer à» et à toutes les questions subséquentes jusqu'à la prochaine question «passer à» le chiffre «9» ou «99», signifiant «non déclaré». Pour les questions comportant un élément de réponse «Ne sait pas», la valeur a été normalisée de manière à ce que «7», «97» ou «997» apparaissent toujours.

7.2.2 Contrôle de la composante École de l'ETJ (groupe des 10-14 ans)

Le questionnaire École a été conçu sans directive «passer à» puisqu'on pensait que les répondants de 10 à 14 ans ne les suivraient pas correctement et que cela risquait de compromettre la qualité des données. Ainsi, tous les répondants ont dû répondre à toutes les questions du questionnaire École et ce, dans un ordre différent du questionnaire EPA. Même si les deux questionnaires contenaient à peu près la même série de questions, le programme de contrôle utilisé pour la composante EPA ne pouvait être répété de façon exacte.

L'équipe a décidé de contrôler la composante École en utilisant la même logique descendante que celle utilisée pour contrôler la série de données de l'EPA. Dans cette optique, les cheminements comparables devaient être déterminés avant que le programme de contrôle puisse être écrit. Ce programme a ensuite été écrit de manière à être aussi comparable que possible à la composante EPA. Les mêmes valeurs imputées ont été ajoutées lorsque cela était pertinent.

7.3 Codage des questions à réponse libre

En prévision de la saisie des données, les commis des bureaux régionaux ont attribué des codes à cinq postes du questionnaire, soit :

	Composante <u>EPA</u>	Composante <u>École</u>
Marque de cigarettes que le jeune fume habituellement	Q14	Q21
Marque de cigarettes que le jeune préfère	Q21	Q24
Problèmes de santé	Q51	Q46
Messages d'avertissement concernant les risques pour la santé	Q54	Q48
Commandite d'événements sportifs et culturels	Q59	Q52b

Les questions de base inscrites sur les questionnaires pour la composante EPA (supplément) et sur ceux de la composante École étaient légèrement différentes.

Un manuel de codage comportant des procédures de codage, des exemples de réponse, un exercice de révision et une liste de codes a été élaboré et a servi principalement à la formation. La formation relative au codage a été donnée dans une salle de classe.

Le surveillant du codage a simulé une vérification des postes codés pour s'assurer que les codes avaient été attribués correctement.

7.3.1 Marque de cigarettes

La liste de codes pour les marques de cigarettes a été fournie par Santé Canada. Elle est composée de 450 codes, dont 153 ont été retranscrits sur une liste distincte des dix marques de cigarettes les plus en demande pour faciliter le codage des réponses du questionnaire. Les codes à trois chiffres n'ont pas été groupés selon un ordre séquentiel et variaient en fonction des différentes marques et descriptions.

Pour la composante EPA, toutes les réponses ont été codées manuellement. Les codeurs devaient examiner la réponse concernant la marque et toutes les réponses relatives aux descriptions (c'est-à-dire, régulière, avec filtre, menthol, etc.) afin d'attribuer un code.

Pour la composante École, on n'a obtenu que la marque, sans description. Ainsi, si une marque énumérée dans le questionnaire avait été cochée, elle était automatiquement codée, et seules les réponses écrites «Autre, précisez» devaient être codées manuellement.

7.3.2 Problèmes de santé

On a demandé aux jeunes de décrire les problèmes de santé que peuvent avoir des gens qui fument pendant de nombreuses années. Pour la composante EPA, les intervieweurs ont enregistré chaque réponse en cochant la réponse appropriée sur la liste des problèmes de santé courants incluse dans le questionnaire, ou ont inscrit les problèmes non mentionnés dans la liste. Les réponses précisées par écrit, jusqu'à un maximum de cinq réponses, ont exigé un codage manuel. Pour la composante École, les élèves ont inscrit leurs réponses dans l'espace prévu sur le questionnaire. Un maximum de huit réponses a été codé manuellement.

La liste des codes concernant les problèmes de santé comptait 64 codes à deux chiffres qui étaient groupés en catégories principales. Il ne fallait pas tenir compte des erreurs d'épellation, mais certaines réponses dans la composante École étaient difficiles à coder en raison d'une mauvaise épellation ou d'une calligraphie illisible.

Aux fins du fichier de microdonnées, la question comparable de la composante École a été réorganisée à partir de variables à code seulement, pour correspondre au format de la composante EPA (une liste et des cases de code). Les codes correspondants à la liste ont été convertis en une série de variables à cocher, et les codes restants ont été assignés aux variables des cases de code. Comme la composante École comptait trois cases de code supplémentaires, même après réorganisation de la question, il a fallu plus de cinq cases de code. Le cliché d'article prévoit une ou deux cases de code additionnelles qui ne s'appliqueraient qu'à la composante École puisqu'il n'a pas été possible d'autoriser les répondants à la composante EPA à donner plus de cinq réponses.

Toutes les réponses dans les variables «case de code» sont demeurées justifiées et classées par ordre numérique ascendant. Aucun dédoublement de code n'a été permis, à l'exception du code «95» pour «Toute autre réponse».

7.3.3 Messages concernant les risques pour la santé

Aux jeunes qui ont déclaré avoir vu des messages concernant les risques pour la santé sur les paquets de cigarettes, on a demandé de citer autant de messages qu'ils le pouvaient. Pour la composante EPA, les intervieweurs enregistraient chaque réponse en cochant le message approprié sur la liste. Pour la composante École, on a demandé aux élèves d'écrire, dans leurs propres mots, n'importe quel message concernant les risques pour la santé qu'ils se rappelaient avoir vu sur des paquets de cigarettes. Ces réponses, huit au plus, ont ensuite été codées manuellement.

Il était peu probable que les jeunes des deux composantes de l'enquête puissent citer intégralement les messages concernant les risques pour la santé qu'on peut lire sur les paquets de cigarettes. Certains mots clés figurant sur le questionnaire (pour la composante EPA) et sur la liste des codes (pour la composante École) étaient imprimés en caractères gras. Si la réponse comprenait les mots imprimés en caractères gras, elle était codée.

Quelques-uns des messages concernant les risques pour la santé comportaient des mots clés semblables. Par exemple, le mot «poumon» paraît dans les messages sur «le cancer des poumons» et «les maladies pulmonaires». Dans ces cas, les codeurs ont attribué le code qui se rapprochait le plus de la réponse de l'élève. On n'a pas tenu compte des erreurs d'épellation et, comme pour la catégorie des problèmes de santé, certaines réponses de la composante École ont été difficiles à coder parce qu'elles étaient mal écrites ou illisibles.

Aux fins du fichier de microdonnées, la question comparable de la composante École a été réorganisée de façon à correspondre au format de la composante EPA. Les codes qui correspondent à la liste de réponses incluse dans le questionnaire de l'EPA ont été convertis en une série de variables «à cocher», qui indiquent si un message particulier concernant les risques pour la santé a été nommé par le répondant. La liste présentée dans le cliché d'article a été quelque peu modifiée à partir du questionnaire de la composante

EPA du fait que la composante École comportait un message additionnel dans sa structure de codage. Le dernier message concernant les risques pour la santé «La fumée du tabac peut nuire à votre enfant» ne s'applique qu'à la composante École.

7.3.4 Commandite d'événements sportifs et culturels

On a demandé aux jeunes s'ils connaissaient des événements sportifs ou culturels qui sont commandités par des sociétés productrices de tabac et de nommer tous les commanditaires et événements qu'ils connaissaient. Chacune de ces réponses, cinq au plus, a été codée manuellement. Pour la composante EPA, on a demandé aux intervieweurs d'inscrire la réponse verbale des jeunes et, pour la composante École, les élèves devaient écrire leurs réponses dans l'espace prévu sur le questionnaire.

Les listes de commanditaires et d'événements ont été fournies par Santé Canada. Une structure de codage à trois chiffres a été mise au point pour coder les réponses en deux groupes principaux :

1. Bonne correspondance entre le commanditaire et l'événement.
2. Seul l'événement est donné, ou un événement avec un mauvais commanditaire.

Il y avait 23 codes de correspondance pour le commanditaire et l'événement. Le nom officiel de l'événement n'était pas nécessaire, et comme il était peu probable que les répondants donnent le nom exact de l'événement, les mots clés étaient inscrits en caractères gras sur la liste de codes. Si une réponse écrite comprenait les mots imprimés en caractères gras, elle était alors codée.

Il y avait 18 codes pour un événement seulement ou pour un événement avec un mauvais commanditaire. Les événements inscrits par les jeunes qui ne figuraient pas sur la liste de codes étaient codés dans «Tout autre événement» et considérés comme une réponse incorrecte. On n'a pas tenu compte des erreurs d'épellation, mais certaines réponses ont été difficiles à coder parce qu'elles étaient mal écrites ou illisibles.

7.4 Création de variables calculées

En vue de faciliter l'analyse des données, un certain nombre d'éléments d'information (variables) figurant au fichier de microdonnées ont été calculés en combinant les postes du questionnaire, par une addition ou par un autre calcul. D'autres ont été calculés en assignant un pseudonyme afin d'éviter de révéler une marque de cigarettes ou une société productrice de tabac en particulier.

Les variables calculées se trouvent dans le cliché d'article immédiatement après les questions de l'ETJ. Les variables calculées sont identifiées par le préfixe «DV».

7.5 Pondération

Toute estimation dans un échantillon probabiliste, comme celui de l'EPA, part du principe suivant lequel chaque personne incluse dans l'échantillon «représente», en plus d'elle-même, plusieurs autres personnes non comprises dans l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 personnes de la population.

À l'étape de la pondération, on calcule le nombre pour chaque enregistrement. Ce poids figure au fichier de microdonnées et doit être utilisé pour calculer des estimations significatives à partir de l'enquête. Par exemple, si l'on doit calculer le nombre de personnes qui n'ont jamais fumé une cigarette entière, on choisit les enregistrements se rapportant aux personnes qui, dans l'échantillon, possèdent ces caractéristiques, et on additionne les poids entrés dans ces enregistrements.

Le chapitre 11 donne des détails sur la méthode permettant de calculer ces poids.

7.6 Suppression des renseignements confidentiels

Il faut signaler que le fichier de microdonnées «à grande diffusion» diffère à bien des égards des fichiers «principaux» de l'enquête détenus par Statistique Canada. Ces différences résultent de mesures qui ont été prises pour préserver l'anonymat des répondants à l'enquête et des sociétés productrices de tabac. Les usagers désirant avoir accès à des renseignements exclus des fichiers de microdonnées peuvent acheter des tableaux personnalisés. Les estimations générées seront offertes aux usagers, sous réserve que soient respectées les directives concernant l'analyse et la diffusion qui sont énoncées à la section 9 du présent document. Les données qui identifient des marques de cigarettes précises ou des sociétés productrices de tabac ne sauraient, en aucun cas, être diffusées au public.

L'Enquête sur le tabagisme chez les jeunes contenait deux types de renseignements différents qui devaient être protégés pour répondre aux exigences de la *Loi sur la statistique* au chapitre de la confidentialité :

- 1) les données qui permettraient d'identifier un répondant particulier ou le ménage auquel il appartient;
- 2) l'identification de marques de cigarettes ou de sociétés productrices de tabac précises.

La plupart des variables dans la partie de l'enquête destinée aux répondants portaient sur des caractéristiques qui ne permettraient pas d'identifier un répondant particulier (par exemple, s'il a déjà essayé de fumer; pourquoi, selon lui, les jeunes de son âge commencent à fumer; quels problèmes de santé les gens peuvent connaître en fumant durant de nombreuses années). Certaines variables, cependant, ont dû être regroupées sous

forme de catégorie agrégée en raison du faible taux de réponse dans les catégories détaillées. Citons, par exemple, des variables comme l'état civil, la branche d'activité et la profession.

Les renseignements concernant le ménage du répondant comportaient également des variables pour lesquelles des mesures ont été prises pour prévenir l'identification des répondants. Dans les variables applicables aux parents du répondant, les catégories détaillées ont été groupées en catégories agrégées. Par exemple, les codes détaillés de la branche d'activité et de la profession ont été groupés sous forme de variables à deux chiffres. Pour les autres membres du ménage, certains renseignements sur des personnes ont été supprimés. Lorsque c'était possible, on a créé des variables calculées pour résumer les caractéristiques du ménage. Par exemple, l'âge des autres membres du ménage est donné sous forme de résumé, indiquant qu'il y a dans le ménage des membres de 0 à 9 ans, de 10 à 19 ans ou de 20 ans et plus.

Quelques variables identifiaient des marques de cigarette et des sociétés productrices de tabac en particulier. Des mesures ont été prises pour s'assurer que de tels renseignements ne soient pas divulgués; on a supprimé, notamment, les variables qui identifiaient directement les noms des marques et des sociétés productrices. D'autres variables telles que la teneur en goudron, en nicotine et en monoxyde de carbone des cigarettes ont été supprimées afin d'éviter la possibilité d'identifier une marque particulière par une combinaison de variables uniques. Autant que possible, les variables calculées ont été créées pour résumer l'information de façon à ce qu'une analyse soit quand même possible. De plus amples détails sur les regroupements et la suppression de variables sont donnés dans le cliché d'article (voir le chapitre 12).

8 QUALITÉ DES DONNÉES

8.1 Taux de réponse

8.1.1 Composante EPA

Le tableau présenté ci-dessous résume les taux de réponse à la composante EPA de l'Enquête sur le tabagisme chez les jeunes. Le taux de réponse correspond au nombre de personnes ayant répondu à l'ETJ en tant que pourcentage du nombre de personnes ayant répondu à l'EPA dans l'échantillon sélectionné.

	Nombre de répondants à l'Enquête sur le tabagisme chez les jeunes	Taux de réponse des personnes à l'Enquête sur le tabagisme chez les jeunes
Terre-Neuve	990	85.7 %
Île-du-Prince-Édouard	375	83.9 %
Nouvelle-Écosse	944	79.6 %
Nouveau-Brunswick	866	77.1 %
Québec	1 303	90.2 %
Ontario	920	80.8 %
Manitoba	941	75.8 %
Saskatchewan	1 099	81.3 %
Alberta	1 030	80.0 %
Colombie-Britannique	1 023	77.1 %
CANADA	9 491	81.1 %

8.1.2 Composante École

On a identifié plusieurs niveaux de non-réponse pour cette composante de l'enquête. Dans les paragraphes qui suivent, on discute de chacune des composantes de la non-réponse et on présente deux tableaux descriptifs.

Premièrement, une certaine non-réponse au niveau des commissions scolaires et des écoles

a été observée. Grâce à la stratégie de remplacement des refus décrite à la Section 5.2.8, on a pu remplacer la totalité des 52 cas de refus observés, soit 6.5% du total des classes. En ce qui a trait aux cas spéciaux tels que décrits à la Section 5.2.7, ceux-ci étaient au nombre de 26 (3.3% du total) et on a remplacé 10 de ces écoles par de nouvelles écoles.

On a également noté cinq cas-problèmes. Tout d'abord, les questionnaires des parents n'ont pu être appariés aux questionnaires correspondant des jeunes dans deux écoles des provinces Atlantiques, soit une à l'Île-du-Prince-Édouard et une au Nouveau-Brunswick. Dans un autre cas, dans une école de la Nouvelle-Écosse, il y a eu erreur dans le niveau scolaire sélectionné: une classe de 7^{ième} année a été sélectionnée au lieu d'une classe de 9^{ième} année. Finalement, deux écoles du Manitoba qui avaient fermé ont été remplacées dans l'échantillon par deux nouvelles écoles alors qu'elles n'auraient pas dû l'être. Il est à noter que dans ces trois derniers cas, on a pu conserver dans l'échantillon les écoles où la collecte avait eu lieu, en supposant qu'elles avaient été sélectionnées dès le départ et en ajustant les poids d'enquête en conséquence.

Le tableau suivant montre la distribution finale des classes répondantes, c'est-à-dire où la collecte a eu lieu, selon la province et le niveau scolaire.

Nombre de classes répondantes selon la province et le niveau scolaire

Province	Niveau scolaire					Total
	5	6	7	8	9	
Terre-Neuve	16	15	16	16	15	78
Île-du-Prince-Édouard	16	15	16	16	16	79
Nouvelle-Écosse	16	14	16	16	15	77
Nouveau-Brunswick	15	16	15	15	16	77
Québec	16	16	16	15	16	79
Ontario	16	16	16	16	15	79
Manitoba	16	16	16	16	15	79
Saskatchewan	15	16	16	15	15	77
Alberta	16	16	16	16	16	80
Colombie-Britannique	16	16	15	16	16	79
Total	158	156	158	157	155	784

La deuxième composante de la non-réponse a trait au refus des parents de laisser leur enfant participer à l'enquête. Environ 91.0% des parents, soit 16,262 sur un total de 17,877 jeunes éligibles, ont consenti à ce que leur enfant participe à l'enquête. De plus, on estime que 70.9% des parents ont donné un consentement écrit. Le reste des consentements a été obtenu par téléphone selon la procédure décrite à la Section 6.2.

Une fois le consentement parental obtenu, il était toujours possible que le jeune soit absent de la classe au moment de la collecte ou refuse de participer à l'enquête. De plus, un certain nombre de questionnaires des jeunes ont dû être rejetés car ils ne répondaient pas aux critères de qualité minimale. Tout ceci étant considéré, 93.6% des jeunes éligibles pour l'enquête et pour lesquels un consentement avait été obtenu, soit 15,217 sur 16,262, ont fourni un questionnaire acceptable.

En ce qui a trait aux parents, ceux-ci pouvaient également refuser de répondre à leur questionnaire ou être absents de la maison lorsqu'on les appelait. Pour ce qui est de la qualité du questionnaire, le seul critère considéré a été que les parents consentent à partager leurs réponses avec Santé Canada. Au total, on a obtenu un questionnaire utilisable dans 93.2% des cas, soit dans 15,161 cas sur 16,262. Le nombre de cas où un questionnaire de parent était rempli mais un consentement au partage n'était pas obtenu était assez faible (moins de 2%) pour qu'on décide tout simplement de rejeter ces questionnaires.

Finalement, la dernière étape consistait à appairer les questionnaires des enfants et ceux des parents à l'aide d'un identificateur unique. Le nombre total de questionnaires appariés avec succès a été de 14,270. Ceci représentait 93.8% des questionnaires valides des jeunes et 94.1% des questionnaires valides des parents.

Le tableau suivant donne un sommaire des taux de réponse globaux obtenus. Ceux-ci représentent le ratio du nombre final de questionnaires valides appariés sur le nombre total de jeunes éligibles à participer à l'enquête dans les classes échantillonnées. Ces taux incorporent donc tous les niveaux de non-réponse décrits précédemment (à l'exception de celui des écoles). Au niveau du Canada, le taux de réponse global s'éleve à 79.8%.

Taux de réponse globaux selon la province

Province	Jeunes éligibles	Questionnaires	Taux de réponse (%)
Terre-Neuve	1815	1476	81.3
Ile-du-Prince-Edouard	1739	1430	82.2
Nouvelle-Ecosse	1759	1431	81.4
Nouveau-Brunswick	1792	1430	79.8
Québec	2023	1556	76.9
Ontario	1785	1260	70.6
Manitoba	1622	1370	84.5
Saskatchewan	1636	1360	83.1
Alberta	1863	1516	81.4
Colombie-Britannique	1843	1441	78.2
Total	17787	14270	79.8

8.2 Erreurs associées à l'enquête

L'enquête produit des estimations à partir des données recueillies auprès et au sujet d'un échantillon de personnes donné. On aurait obtenu des estimations légèrement différentes si un recensement complet avait été mené avec les mêmes questionnaires, intervieweurs, surveillants, méthodes de traitement, etc., que ceux qui ont été utilisés dans l'enquête. La différence entre les estimations obtenues de l'échantillon et celles résultant d'un dénombrement complet effectué dans des conditions semblables est appelée erreur d'échantillonnage de l'estimation.

D'autres erreurs, qui ne sont pas liées à l'échantillonnage, peuvent se produire à pratiquement n'importe quelle étape d'une enquête. Les intervieweurs peuvent mal interpréter les instructions, les répondants peuvent faire des erreurs en répondant aux questions, les réponses peuvent être inscrites incorrectement sur le questionnaire et des erreurs peuvent survenir au moment de traiter et de totaliser les données. Ce sont tous des exemples d'erreurs non liées à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires qui se produisent auront peu d'incidence sur les estimations calculées à partir de l'enquête. Toutefois, des erreurs qui se produisent systématiquement entraîneront des biais dans les estimations de l'enquête.

Beaucoup de temps et d'efforts ont été consacrés à la réduction du nombre d'erreurs non liées à l'échantillonnage dans l'enquête. Des mesures de contrôle de la qualité ont été implantées à chaque étape du cycle de collecte et de traitement des données afin de vérifier la qualité des données. Au nombre de ces mesures, citons l'utilisation d'intervieweurs très compétents, une formation approfondie des intervieweurs concernant les procédures et le questionnaire de l'enquête, l'observation des intervieweurs pour déceler des problèmes au niveau de la conception du questionnaire ou de l'interprétation des instructions, des procédures visant à garantir un minimum d'erreurs pendant la saisie des données et la vérification du codage et de la qualité du contrôle pour évaluer la logique du traitement.

8.2.1 Non-réponse totale

La non-réponse totale peut être une source majeure d'erreurs non liées à l'échantillonnage dans plusieurs enquêtes, selon la mesure dans laquelle les répondants et les non-répondants diffèrent en fait de caractéristiques d'intérêt. Il y a non-réponse totale lorsque l'intervieweur est incapable d'entrer en communication avec le répondant ou que le répondant a refusé de participer à l'enquête. Il faut alors ajuster le poids des ménages qui ont répondu à l'enquête pour compenser ceux qui n'ont pas répondu.

8.2.2 Non-réponse partielle

Dans la plupart des cas, il y a non-réponse partielle à l'enquête lorsque le répondant n'a pas compris la question ou l'a mal interprétée, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée. Une non-réponse partielle est signalée par des codes sur le fichier de microdonnées (p. ex. «non déclaré»).

8.2.3 Couverture

A) Couverture pour la composante EPA

Comme on l'a mentionné à la section 5.1.1 (couverture de la population), la population cible du supplément ETJ-EPA était entièrement composée de jeunes âgés de 15 à 19 ans. Étaient cependant expressément exclus de la couverture de cette enquête les membres à plein temps des Forces armées canadiennes, les détenus des établissements pénitentiaires ainsi que les résidents du Yukon et des Territoires du Nord-Ouest et les personnes vivant sur une réserve indienne. Ces groupes représentent une exclusion de 2 % environ de la population âgée de 15 ans et plus. Les membres de ces groupes exclus peuvent posséder des caractéristiques uniques qui ne seront pas reflétées dans les estimations de l'enquête. Les utilisateurs devraient agir avec circonspection lorsqu'ils analysent des sous-groupes de population dont les caractéristiques peuvent être mises en corrélation avec celles des groupes de population qui ont été exclus du supplément ETJ-EPA.

B) Couverture pour la composante des écoles

Comme on l'a expliqué dans la Section 5.2.2, la population d'enquête diffère quelque peu de la population-cible. Il est possible que les jeunes appartenant à l'une ou l'autre des catégories d'exclusion possèdent des caractéristiques différentes des autres jeunes par rapport au tabagisme. L'utilisateur devrait particulièrement faire preuve de prudence si son analyse porte sur un sous-groupe dont les caractéristiques sont corrélées avec celles des groupes exclus, par exemple le sous-groupe des jeunes de 14 ans.

8.2.4 Critères de mesure de l'erreur d'échantillonnage

Comme il est inévitable que les estimations faites à partir d'une enquête-échantillon soient sujettes à des erreurs d'échantillonnage, il est statistiquement judicieux pour les chercheurs de donner aux usagers une idée de l'ampleur de cette erreur. La présente section énonce les critères de mesure de l'erreur d'échantillonnage que Statistique Canada utilise couramment; de même, SC incite les utilisateurs qui produisent des estimations à partir des fichiers de microdonnées à les utiliser également.

Pour mesurer la taille potentielle des erreurs d'échantillonnage, on se fonde sur l'erreur type des estimations calculées à partir des résultats de l'enquête.

Cependant, en raison de la grande variété des estimations qui peuvent être produites à partir d'une enquête, l'erreur type d'une estimation est habituellement exprimée par rapport à l'estimation à laquelle elle se rapporte. La mesure qui en résulte, qu'on appelle «coefficient de variation» (C.V.) d'une estimation, est obtenue en divisant l'erreur type de l'estimation par l'estimation elle-même et est exprimée sous forme de pourcentage de l'estimation.

Par exemple, supposons que, d'après les résultats de l'enquête, on estime que 23,4 % des jeunes de 15 à 19 ans au Canada sont présentement des fumeurs et que l'erreur type de l'estimation est établie à 0,0043. Le coefficient de variation de l'estimation se calcule alors de la façon suivante :

$$\frac{.0043}{.234} \times 100\% = 1.8\%$$

9 LIGNES DIRECTRICES CONCERNANT LA TOTALISATION, L'ANALYSE ET DIFFUSION

La présente section expose les lignes directrices auxquelles doivent se conformer les utilisateurs qui veulent totaliser, analyser, publier ou autrement diffuser toute donnée calculée à partir des bandes de microdonnées de l'enquête. En suivant ces lignes directrices, les utilisateurs de microdonnées devraient être en mesure de produire les mêmes chiffres que ceux produits par Statistique Canada et, en même temps, d'élaborer des chiffres inédits jusqu'à maintenant d'une manière qui soit conforme aux lignes directrices établies.

9.1 Lignes directrices concernant l'arrondissement

De façon à ce que les estimations qui sont destinées à être publiées ou autrement diffusées et qui ont été calculées à partir des bandes de microdonnées correspondent à celles produites par Statistique Canada, on incite les utilisateurs à suivre les lignes directrices suivantes concernant l'arrondissement de ces estimations :

- a) Les estimations contenues dans le corps d'un tableau statistique doivent être arrondies à la centaine d'unités la plus proche en utilisant la technique d'arrondissement habituelle. Dans un arrondissement habituel, si le premier ou seul chiffre à éliminer se situe entre 0 et 4, le dernier chiffre qui doit être retenu ne change pas. Si le premier ou seul chiffre à éliminer se situe entre 5 et 9, le dernier chiffre à être retenu est augmenté de 1. Par exemple, dans l'arrondissement normal à la centaine la plus proche, si les deux derniers chiffres se situent entre 00 et 49, ils sont changés à 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont changés à 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux et les totaux partiels marginaux qu'on trouve dans les tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ils sont eux-mêmes arrondis à la centaine d'unités la plus proche en utilisant la technique habituelle.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir des composantes non arrondies (c.-à-d. les numérateurs et (ou) les dénominateurs) et être ensuite arrondis eux-mêmes à une décimale près à l'aide de la technique habituelle. Dans un arrondissement normal à un chiffre, si le chiffre final ou le seul chiffre à éliminer se situe entre 0 et 4, le dernier à être retenu est inchangé. Si le premier ou le seul chiffre à éliminer se situe entre 5 et 9, le dernier chiffre à être retenu est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou ratios) doivent être calculées à partir

de leurs composantes non arrondies correspondantes et sont ensuite arrondies elles-mêmes à la centaine la plus proche (ou à la décimale la plus proche) en utilisant la technique habituelle.

- e) Dans les cas où, en raison de limites techniques ou autres, il faut utiliser une autre technique d'arrondissement, on publiera ou on diffusera des estimations qui diffèrent des estimations correspondantes publiées par Statistique Canada. On incite donc les utilisateurs à noter la raison de telles différences dans les documents publiés ou diffusés.
- f) Des estimations non arrondies ne sauraient, en aucun cas, être publiées ou autrement diffusées par les utilisateurs. Les estimations non arrondies supposent une précision de loin supérieure à celle qu'on a obtenue.

9.2 Lignes directrices concernant la pondération de l'échantillon aux fins de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête sur le tabagisme chez les jeunes n'était pas autopondéré. Au moment de produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

S'ils n'utilisent pas les bons poids, les estimations calculées à partir des bandes de microdonnées ne peuvent être jugées représentatives de la population de l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également savoir que certains progiciels ne permettent pas de produire des estimations qui correspondent exactement à celles offertes par Statistique Canada du fait que les poids ne font pas l'objet d'un traitement similaire.

9.2.1 Définition des types d'estimations : catégoriques et quantitatives

Avant de voir comment on peut totaliser et analyser les données de l'Enquête sur le tabagisme chez les jeunes, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être générées à partir du fichier de microdonnées concernant l'Enquête sur le tabagisme chez les jeunes.

Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre de répondants ou du pourcentage de la population visée par l'enquête qui possèdent certaines caractéristiques ou qui entrent dans une quelconque catégorie définie. On peut citer comme exemples le nombre de jeunes qui n'ont jamais fumé une cigarette entière ou la proportion du nombre

de jours d'usage du tabac au cours des 30 derniers jours. Une estimation du nombre de personnes possédant une certaine caractéristique peut également être considérée comme une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Avez-vous déjà fumé une cigarette entière?

R : Oui/Non

Q : Au cours des 30 derniers jours, durant combien de jours avez-vous fumé une ou plusieurs cigarettes?

R : Aucun / 1-5 jours / 6-10 jours / 11-20 jours / 21-29 jours / 30 jours

Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et d'autres mesures de tendances centrales de quantités qui sont fondées sur certains ou sur tous les membres de la population visée par l'enquête. En outre, elles comprennent précisément les estimations de la forme X/Y , où X est une estimation du total de la population visée par l'enquête et Y , une estimation du nombre de personnes de la population visée par l'enquête contribuant à ce total.

Par exemple, le nombre moyen de fois qu'un jeune a tenté de cesser de fumer. Le numérateur est une estimation du nombre total de tentatives de cesser de fumer, et le dénominateur est le nombre de jeunes qui ont déclaré avoir essayé d'arrêter de fumer.

Exemples de questions quantitatives

Q : Combien de fois avez-vous essayé d'arrêter de fumer?

R : |_|_| Nombre de fois

Q : Quel âge aviez-vous lorsque vous avez essayé de cesser de fumer pour la première fois?

R : |_|_| Ans

9.2.2 Totalisation des estimations catégoriques

Les estimations du nombre de personnes possédant une certaine caractéristique peuvent être obtenues à partir du fichier de microdonnées en additionnant les poids finaux de tous les enregistrements possédant la ou les caractéristiques d'intérêt. Les proportions et les ratios de la forme X/Y sont obtenus en :

- a) additionnant les poids finaux des enregistrements ayant la caractéristique d'intérêt pour le numérateur (X);

- b) additionnant les poids finaux des enregistrements ayant la caractéristique d'intérêt pour le dénominateur (Y);
- c) divisant l'estimation du numérateur par l'estimation du dénominateur.

9.2.3 Totalisation des estimations quantitatives

Les estimations quantitatives peuvent être obtenues à partir du fichier de microdonnées en multipliant la valeur de la variable d'intérêt par le poids final pour chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements d'intérêt. Par exemple, pour obtenir une estimation du nombre total d'heures hebdomadaires travaillées par des jeunes ayant un emploi rémunéré, on multiplie la valeur déclarée à la Q73_63 (heures hebdomadaires travaillées) par le poids final pour l'enregistrement, puis on additionne cette valeur pour tous les enregistrements avec la Q72_62=1 (Oui, le répondant a un emploi pour lequel il est rémunéré).

Pour obtenir une moyenne pondérée de la forme X/Y, le numérateur (X) est calculé comme pour une estimation quantitative, et le dénominateur (Y) est calculé comme pour une estimation catégorique. Par exemple, pour estimer le nombre moyen d'heures hebdomadaires travaillées par les jeunes ayant un emploi rémunéré, on doit :

- a) estimer le nombre total d'heures hebdomadaires travaillées de la façon décrite ci-dessus;
- b) estimer le nombre de personnes faisant partie de cette catégorie en additionnant les poids finaux de tous les enregistrements avec la Q72_62=1;
- c) diviser l'estimation a) par l'estimation b).

9.3 Lignes directrices concernant l'analyse statistique

L'Enquête sur le tabagisme chez les jeunes repose sur un plan complexe, comportant une stratification et de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données d'enquêtes aussi complexes pose des problèmes aux analystes du fait que le plan d'enquête et les probabilités de sélection ont une incidence sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées.

Bien qu'un grand nombre des procédures d'analyse qu'on trouve dans l'ensemble des programmes statistiques permettent l'utilisation de poids, la signification ou la définition du poids dans ces procédures diffère de celle qui est jugée appropriée dans un cadre d'enquête-échantillon. Résultat : même si, dans de nombreux cas, les estimations produites par les programmes sont exactes, les variances qui sont calculées sont presque non significatives.

Pour un grand nombre de techniques d'analyse (par exemple, la régression linéaire, la

régression logistique, l'analyse de la variance), il existe une méthode qui peut rendre l'application des programmes normalisés plus significative. Si les poids figurant sur les enregistrements sont rééchelonnés de façon à ce que le poids moyen soit (1), les résultats produits par les programmes normalisés seront alors plus raisonnables; ils ne tiendront toujours pas compte de la stratification ni du groupement du plan de l'échantillon, mais ils tiendront compte des probabilités inégales de sélection. On peut effectuer le rééchelonnement en divisant chaque poids par le poids global moyen avant d'effectuer l'analyse.

Dans le but de fournir une méthode permettant d'évaluer la qualité des estimations calculées, Statistique Canada a produit une série de tableaux sur la variabilité d'échantillonnage approximative (communément appelés «tableaux C.V.») aux fins de l'Enquête sur le tabagisme chez les jeunes. Ces tableaux permettent d'obtenir des coefficients de variation approximatifs pour les proportions et les estimations de type catégorique. Voir le chapitre 10 pour de plus amples détails.

9.4 Lignes directrices concernant la diffusion des C.V.

Avant de diffuser et (ou) de publier toute estimation faite à partir de ces bandes de microdonnées, les utilisateurs devraient d'abord déterminer le nombre de répondants inclus dans le calcul de l'estimation. Si le chiffre est inférieur à 30, l'estimation pondérée ne devrait pas être diffusée, peu importe la valeur du coefficient de variation pour cette estimation. Pour les estimations pondérées fondées sur des échantillons de 30 ou plus, les utilisateurs devraient établir le coefficient de variation de l'estimation **arrondie** et suivre les lignes directrices énoncées ci-dessous.

Lignes directrices concernant la variabilité d'échantillonnage

Type d'estimation	c.v. (%)	Lignes directrices
1. Sans restriction	0,0 - 16,5	Les estimations peuvent être utilisées pour publication générale sans restriction. Elles ne nécessitent aucune note spéciale.
2. Avec restriction	16,6 - 25,0	Les estimations peuvent être utilisées pour diffusion générale sans restriction, mais doivent être accompagnées d'une mise en garde à l'intention des utilisateurs subséquents relativement à la grande variabilité d'échantillonnage associée aux estimations. De telles estimations devraient être marquées de la lettre Q (ou de toute autre façon semblable).
3. Confidentiel	25,1 - 33,3	Les estimations peuvent être utilisées pour diffusion générale sans restriction seulement lorsque les variabilités d'échantillonnage sont obtenues en utilisant une méthode de calcul de la variance exacte. À moins d'obtenir des variances exactes, de telles estimations devraient être éliminées et remplacées par des (---) dans les tableaux statistiques.
4. Non diffusable	33,4 ou plus	Les estimations ne peuvent être diffusées sous aucune forme NI en aucun cas. Dans les tableaux statistiques, de telles estimations devraient être éliminées et remplacées par des (---).

10 TABLEAUX SUR LA VARIABILITÉ D'ÉCHANTILLONNAGE APPROXIMATIVE

En vue de fournir des coefficients de variation qui seraient applicables à une grande variété d'estimations catégoriques produites à partir de ce fichier de microdonnées et qui pourraient être utilisées d'office par l'utilisateur, une série de tableaux sur la variabilité d'échantillonnage approximative a été produite. Ces tableaux de consultation permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation qui a été calculée à partir des données de l'enquête.

Les coefficients de variation (C.V.) sont calculés en utilisant la formule de variance pour l'échantillonnage aléatoire simple et en intégrant un facteur qui tient compte de la nature pluridimensionnelle et regroupée du plan d'échantillonnage. Ce facteur, qu'on appelle effet du plan, a été établi en calculant d'abord les effets du plan sur une vaste gamme de caractéristiques, puis en choisissant parmi celles-ci une valeur conservatrice pouvant être utilisée dans les tableaux de consultation qui s'appliqueraient alors à l'ensemble des caractéristiques.

Le tableau ci-dessous présente les effets du plan, les tailles d'échantillon et les chiffres de population, par province, qui ont été utilisés pour produire les tableaux sur la variabilité d'échantillonnage approximative.

Composante EPA

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
Terre-Neuve	2,02	990	48 385
Île-du-Prince-Édouard	1,56	375	10 045
Nouvelle-Écosse	2,16	944	63 819
Nouveau-Brunswick	1,41	866	54 934
Québec	2,59	1 303	487 428
Ontario	2,95	920	703 632
Manitoba	2,50	941	74 768
Saskatchewan	1,75	1 099	71 267
Alberta	2,15	1 030	185 245
Colombie-Britannique	2,45	1 023	232 411
Provinces de l'Atlantique	1,95	3 175	177 183
Prairies	2,70	3 070	331 280
Canada	5,47	9 491	1 931 934

Composante École

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
Terre-Neuve	1,91	1 476	44 997
Île-du-Prince-Édouard	1,93	1 430	9 725
Nouvelle-Écosse	1,85	1 431	61 877
Nouveau-Brunswick	2,40	1 430	51 935
Québec	2,66	1 556	478 242
Ontario	2,05	1 260	711 502
Manitoba	2,22	1 370	74 755
Saskatchewan	1,94	1 360	76 704
Alberta	2,16	1 516	201 801
Colombie-Britannique	2,18	1 441	237 721
Provinces de l'Atlantique	2,58	5 767	168 534
Prairies	2,53	4 246	353 260
Canada	4,96	14 270	1 949 259

Les deux composantes

PROVINCE	EFFET DU PLAN	TAILLE DE L'ÉCHANTILLON	POPULATION
Terre-Neuve	2,33	2 466	93 382
Île-du-Prince-Édouard	2,96	1 805	19 770
Nouvelle-Écosse	2,43	2 375	125 696
Nouveau-Brunswick	1,91	2 296	106 869
Québec	2,67	2 859	965 670
Ontario	3,20	2 180	1 415 134
Manitoba	2,78	2 311	149 523
Saskatchewan	1,91	2 459	147 971
Alberta	2,42	2 546	387 046
Colombie-Britannique	2,56	2 464	470 132
Provinces de l'Atlantique	2,87	8 942	345 717
Prairies	2,73	7 316	684 540
Canada	5,61	23 761	3 881 193

Tous les coefficients de variation dans les tableaux sur la variabilité d'échantillonnage approximative sont approximatifs et, de fait, officieux. Les estimations de la variance réelle pour des variables particulières peuvent être obtenues auprès de Statistique Canada contre remboursement des frais. Les estimations de la variance réelle permettraient aux utilisateurs de diffuser des estimations autrement non diffusables, c.-à-d. des estimations dont les coefficients de variation se situent dans la catégorie «confidentiel».

Rappel : Si le nombre d'observations sur lesquelles une estimation est fondée est inférieur à 30, l'estimation pondérée ne devrait pas être diffusée, peu importe la valeur du coefficient de variation pour cette estimation. Cette restriction est attribuable au fait que les formules utilisées pour estimer la variance ne sont pas exactes pour les échantillons de petite taille.

10.1 Comment utiliser les tableaux C.V. aux fins des estimations catégoriques

Les règles énoncées ci-dessous devraient aider l'utilisateur à déterminer les coefficients de variation approximatifs inclus dans les tableaux de variabilité d'échantillonnage en vue d'effectuer des estimations du nombre de personnes, de la proportion ou du pourcentage de la population visée par l'étude qui possède une certaine caractéristique, ainsi que des ratios et des différences entre les estimations obtenues.

Règle n° 1 : Estimation du nombre de personnes possédant une caractéristique (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Sur le tableau de variabilité d'échantillonnage, repérez, pour la région géographique appropriée, le nombre estimé dans la colonne à l'extrême gauche du tableau (intitulé «Numérateur de pourcentage») et suivez les astérisques (s'il y a lieu) jusqu'au premier chiffre. Ce chiffre représente le coefficient de variation approximatif.

Règle n° 2 : Estimation des proportions ou des pourcentages de population possédant une caractéristique

Le coefficient de variation d'une proportion ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage est fondé. Les proportions ou pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, où la proportion ou le pourcentage est fondé sur un sous-groupe de la population. Par exemple, la proportion de «jeunes de 15 à 19 ans qui fument présentement» est plus fiable que le nombre estimé de «jeunes de 15 à 19 qui fument présentement». (Veuillez noter que, dans les tableaux, la valeur du c.v. diminue de gauche à droite).

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique couverte par le tableau, le c.v. de la proportion ou du pourcentage est le même que le c.v. du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut utiliser la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (p. ex., les personnes d'un groupe d'âge ou d'un sexe particulier), on doit faire référence à la proportion ou au pourcentage (première ligne du tableau) et au numérateur de la proportion ou du pourcentage (au bas de la colonne de gauche du tableau). L'intersection de la ligne et de la colonne appropriées donne le coefficient de variation.

Règle n° 3 : Estimation des différences entre les agrégats ou les pourcentages

L'erreur type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur type prise séparément.

Ainsi, l'erreur type d'une différence ($\hat{d} = \hat{x}_1 - \hat{x}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{x}_1 \alpha_1)^2 + (\hat{x}_2 \alpha_2)^2}$$

où \hat{x}_1 est l'estimation 1, \hat{x}_2 l'estimation 2, et α_1 et α_2 représentent les coefficients de variation de \hat{x}_1 et de \hat{x}_2 , respectivement. Le coefficient de variation de \hat{d} est donné par $\sqrt{\hat{d}} / \hat{d}$. Cette formule est exacte pour la différence entre les caractéristiques distinctes et non mises en corrélation, mais n'est autrement que purement approximative.

Règle n° 4 : Estimation des ratios

Lorsque le numérateur est un sous-ensemble du dénominateur, le ratio devrait être converti en pourcentage, et la règle n° 2 devrait s'appliquer. Ceci s'appliquerait, par exemple, au cas où le dénominateur est le nombre de jeunes de 15 à 19 ans et le numérateur, le nombre de «jeunes de 15 à 19 ans qui fument présentement».

Lorsque le numérateur n'est pas un sous-ensemble du dénominateur, comme le ratio du nombre de «jeunes de 15 à 19 ans qui fument présentement» comparé au nombre de «jeunes de 15 à 19 ans qui ont déjà fumé mais ne fument plus», l'écart type du ratio des estimations est à peu près égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément et multiplié par R.

C'est-à-dire, l'erreur type d'un ratio ($\hat{R} = \hat{x}_1 / \hat{x}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{x}_1 et \hat{x}_2 , respectivement.

Le coefficient de variation de \hat{R} est donné par $\sqrt{\hat{R}} / \hat{R}$. La formule aura tendance à

surévaluer l'erreur, si \hat{x}_1 et \hat{x}_2 sont en corrélation directe et à sous-évaluer l'erreur si \hat{x}_1 et \hat{x}_2 sont en corrélation inverse.

Règle n° 5 : Estimation des différences de ratios

Dans ce cas, les règles 3 et 4 sont combinées. Les c.v. pour les deux ratios sont d'abord établis avec la règle n° 4, puis le c.v. de leur différence est établi en appliquant la règle n° 3.

10.2 Exemples de l'utilisation des tableaux sur les c.v. aux fins des estimations catégoriques

Les exemples concrets suivants visent à aider les utilisateurs à appliquer les règles exposées ci-dessus.

Exemple 1 : Estimation du nombre de personnes possédant une caractéristique (agrégats)

Supposons qu'un utilisateur estime que 451 858 jeunes de 15 à 19 ans étaient des fumeurs durant la période de référence. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Retournez au tableau des c.v. pour le CANADA.
- 2) L'agrégat estimé (451 858) n'apparaît pas dans la colonne de gauche («numérateur de pourcentage»); il est donc nécessaire d'utiliser le nombre le plus près de cet agrégat, soit 450 000.
- 3) On détermine le coefficient de variation pour un agrégat estimé en se reportant à la première entrée sans astérisque se trouvant sur cette ligne, soit 4,3 %.
- 4) Ainsi, le coefficient de variation approximatif de l'estimation est 4,3 %.
La conclusion suivant laquelle 451 858 jeunes de 15 à 19 ans étaient des fumeurs au moment de la période de référence peut être diffusée, sans restriction.

Exemple n° 2 : Estimation des proportions ou des pourcentages de population possédant une caractéristique

Supposons que l'utilisateur estime que $341\,236/451\,858 = 75,5\%$ des jeunes de 15 à 19 ans qui étaient des fumeurs durant la période de référence ont déclaré qu'ils fumaient tous les jours. Comment l'utilisateur détermine-t-il le coefficient de variation de cette

estimation?

- 1) Retournez au tableau pour le CANADA.
- 2) Comme l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c.-à-d., les jeunes de 15 à 19 ans qui étaient des fumeurs durant la période de référence), il faut utiliser à la fois le pourcentage (75,5 %) et la portion du numérateur du pourcentage (341 236) pour déterminer le coefficient de variation.
- 3) Le numérateur, 341 236, n'apparaît pas dans la colonne de gauche («Numérateur de pourcentage»); il faut donc utiliser le chiffre le plus près, soit 350 000. De la même façon, le pourcentage estimé n'apparaît dans aucune des colonnes; il faut donc utiliser le chiffre le plus près, soit 70,0 %.
- 4) Le chiffre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 3,1 %, représente le coefficient de variation à utiliser.
- 5) Ainsi, le coefficient de variation approximatif de l'estimation est 3,1 %. La conclusion suivant laquelle 75,5 % des jeunes de 15 à 19 ans qui fumaient durant la période de référence ont fumé tous les jours peut être diffusée, sans restriction.

Exemple n° 3 : Estimation des différences entre les agrégats ou les pourcentages

Supposons qu'un utilisateur estime que $177\,052/224\,852 = 78,7\%$ des filles âgées de 15 à 19 ans qui fumaient au moment de la période de référence ont déclaré fumer tous les jours, tandis que $164\,184/227\,006 = 72,3\%$ des garçons âgés de 15 à 19 ans qui fumaient durant la période de référence ont déclaré fumer tous les jours. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) En utilisant le tableau des c.v. pour le CANADA de la même manière que celle décrite à l'exemple 2, on obtient un c.v. de 4,1 % de l'estimation pour les filles, et un c.v. de 4,7 % de l'estimation pour les garçons.
- 2) En utilisant la règle n° 3, l'erreur type d'une différence $\hat{d} = \hat{x}_1 - \hat{x}_2$ est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{x}_1 \alpha_1)^2 + (\hat{x}_2 \alpha_2)^2}$$

où \hat{x}_1 est l'estimation 1, \hat{x}_2 est l'estimation 2 et α_1 et α_2 représentent les coefficients de variation de \hat{x}_1 et \hat{x}_2 , respectivement.

Ainsi, l'erreur type de la différence $\hat{d} = (0,787-0,723) = 0,065$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(.787)(.041)]^2 + [(.723)(.047)]^2} \\ &= \sqrt{(.001041) + (.001155)} \\ &= .047\end{aligned}$$

- 3) Le coefficient de la variation de \hat{d} est donné par $\sqrt{\hat{d}} / \hat{d} = 0,047/0,065 = 0,72$.
- 4) Ainsi, le coefficient de variation approximatif de la différence entre les estimations est 72 %. Cette estimation ne saurait en aucun cas être diffusée et devrait être éliminée et remplacée par des tirets.

Exemple n° 4 : Estimation de ratios

Supposons que, d'après les estimations d'un utilisateur, 177 052 jeunes filles de 15 à 19 ans fumaient tous les jours durant la période de référence, et 164 184 garçons de 15 à 19 ans fumaient tous les jours durant la période de référence. L'utilisateur est intéressé à comparer l'estimation des femmes par rapport à celle des hommes sous forme de ratio. Comment doit-il déterminer le coefficient de variation de cette estimation?

- 1) D'abord, cette estimation est une estimation par quotient, où le numérateur de l'estimation ($= \hat{x}_1$) est le nombre de jeunes filles âgées de 15 à 19 ans qui fumaient tous les jours durant la période de référence. Le dénominateur de l'estimation ($= \hat{x}_2$) est le nombre de garçons de 15 à 19 ans qui fumaient tous les jours durant la période de référence.
- 2) Retournez au tableau des c.v. pour le CANADA.
- 3) Le numérateur de cette estimation par quotient est 177 052. Le chiffre le plus près est 200 000. Le coefficient de variation pour cette estimation est établi en se reportant à la première entrée sans astérisque se trouvant sur cette ligne, soit 6,9 %.
- 4) Le dénominateur de cette estimation par quotient est 164 184. Le chiffre le plus près de cette estimation est 150 000. Le coefficient de variation pour cette estimation est établi en se reportant à la première entrée sans astérisque se trouvant sur cette ligne,

soit 8,2 %

- 5) Ainsi, on obtient le coefficient de variation approximatif de l'estimation par quotient en utilisant la règle n° 4, soit :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 , respectivement.

Soit :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(.069)^2 + (.082)^2} \\ &= 0.107\end{aligned}$$

Ainsi, le rapport obtenu entre les filles et les garçons de 15 à 19 ans qui fumaient tous les jours durant la période de l'enquête est 177 052/164 184, soit 1,08:1. Le coefficient de variation de cette estimation est de 10,7 %, résultat pouvant être diffusé, sans restriction.

10.3 Comment utiliser les tables de c.v. pour obtenir les limites de confiance

Bien qu'on utilise largement les coefficients de variation, l'intervalle de confiance d'une estimation constitue une mesure intuitivement plus significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue un énoncé à l'égard du niveau de confiance suivant lequel la valeur réelle pour la population se situe dans une échelle de valeurs précise. Par exemple, un intervalle de confiance de 95 % peut être décrit de la façon suivante :

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, alors, dans 95 % des échantillons, l'intervalle couvrira la valeur réelle de la population.

En utilisant l'erreur type d'une estimation, on peut obtenir des intervalles de confiance pour les estimations en supposant que, avec un échantillonnage répété de la population, les diverses estimations qu'on obtiendrait pour une caractéristique de la population sont normalement distribuées pour la valeur réelle de la population. Selon cette hypothèse, les probabilités sont de 68 sur 100 environ que la différence entre une estimation d'échantillon et la valeur réelle de la population soit moins d'une erreur type, de 95 sur 100 environ que la différence soit moins de deux erreurs types et de 99 sur 100 environ que la différence soit moins de trois erreurs types. Ces différents coefficients de confiance sont appelés «niveaux de confiance».

Les intervalles de confiance pour une estimation, \hat{x} , sont généralement exprimés par deux nombres, soit un en deçà de l'estimation et l'autre au-dessus de l'estimation, tel que $(\hat{x} - k, \hat{x} + k)$, où k est déterminé selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Les intervalles de confiance pour une estimation peuvent être calculés directement à partir des tableaux sur la variabilité d'échantillonnage approximative en déterminant d'abord à partir du tableau approprié le coefficient de variation de l'estimation \hat{x} et en utilisant ensuite la formule suivante pour le convertir en intervalle de confiance CI :

$$CI_x = [\hat{X} - t \hat{X} \alpha_{\hat{x}}, \hat{X} + t \hat{X} \alpha_{\hat{x}}]$$

où α_x est le coefficient de variation déterminé de \hat{x} , et

- t = 1 si on désire un intervalle de confiance de 68 %;
- t = 1.6 si on désire un intervalle de confiance de 90 %;
- t = 2 si on désire un intervalle de confiance de 95 %;
- t = 3 si on désire un intervalle de confiance de 99 %.

Remarque : Les lignes directrices concernant la publication qui s'appliquent à l'estimation s'appliquent également à l'intervalle de confiance. Par exemple, si l'estimation ne peut pas être diffusée, l'intervalle de confiance ne peut pas l'être non plus.

10.4 Exemple d'utilisation des tableaux sur les c.v. pour obtenir les limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée de jeunes de 15 à 19 ans qui étaient des fumeurs durant la période de référence et qui ont déclaré fumer tous les jours (tiré de l'exemple 2, section 10.2) serait calculé de la façon suivante :

$$\hat{x} = 75,5 \% \text{ (ou exprimé comme proportion} = 0,755)$$

$$t = 2$$

$\alpha_{\hat{x}} = 3,1\%$ (0,031 exprimé comme proportion) est le coefficient de variation de cette estimation, déterminée à partir des tableaux.

$$CI \hat{x}_1 = \{0,755 - (2) (0,55) (0,031), 0,755 + (2) (0,755) (0,031)\}$$

$$CI \hat{x}_1 = \{0,755 - 0,047, 0,755 + 0,047\}$$

$$CI \hat{x}_1 = \{0,708, 0,802\}$$

Avec un intervalle de confiance de 95 %, on peut affirmer qu'entre 70,8 % et 80,2 % des jeunes de 15 à 19 ans qui étaient de fumeurs durant la période de référence fumaient tous les jours.

10.5 Comment utiliser les tableaux sur les c.v. pour effectuer un test t

On peut aussi utiliser des erreurs types pour effectuer un test d'hypothèse, procédure qui vise à établir une distinction entre les paramètres d'une population à l'aide des estimations d'échantillon. Les estimations d'échantillon peuvent être des nombres, des moyennes, des pourcentages, des ratios, etc. Le test peut être effectué à différents niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes lorsque, en fait, elles sont identiques.

Supposons que X_1 et X_2 sont des estimations d'échantillon pour deux caractéristiques d'intérêt. Supposons que l'erreur type de la différence $\hat{x}_1 - \hat{x}_2$ est σ_d .

Si $t = \frac{\hat{x}_1 - \hat{x}_2}{\sigma_d}$ se situe entre -2 et 2, alors aucune conclusion quant à la différence entre

les caractéristiques n'est justifiée à un niveau de signification de 5 %. Si, toutefois, ce ratio est plus petit que -2 ou plus grand que +2, la différence observée est significative au niveau 0,05. C'est donc dire que les caractéristiques sont significatives.

10.6 Exemple d'utilisation des tableaux sur les c.v. pour effectuer un test t

Supposons que nous désirons tester, à un niveau de signification de 5 %, l'hypothèse suivant laquelle il n'y a pas de différence entre la proportion de filles de 15 à 19 ans qui fumaient tous les jours durant la période de référence et la proportion de garçons de 15 à 19 ans qui fumaient tous les jours durant la période de référence. À partir de l'exemple n° 3, section 10.2, l'erreur type de référence entre les deux estimations a été établie à 0,047. Ainsi,

$$t = \frac{\hat{x}_1 - \hat{x}_2}{\sigma_d} = \frac{.787 - .723}{.047} = \frac{.065}{.047} = 1.38.$$

Puisque $t = 1,38$ est inférieur à 2, on doit conclure qu'il n'y a pas de différence significative entre les deux estimations à un niveau de signification de 0,05.

10.7 Coefficients de variation pour les estimations quantitatives

Pour les estimations quantitatives, il aurait fallu produire des tableaux spéciaux pour déterminer leur erreur d'échantillonnage. Comme la plupart des variables concernant l'Enquête sur le tabagisme chez les jeunes sont principalement catégoriques, cela n'a pas été fait.

Règle générale, cependant, le coefficient de variation d'un total quantitatif sera plus grand que le coefficient de variation d'estimation catégorique correspondante (c.-à-d., l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation catégorique correspondante ne peut pas être diffusée, l'estimation quantitative ne peut pas l'être non plus. Par exemple, le coefficient de variation du «nombre total de semaines travaillées par les jeunes de 15 à 19 ans qui avaient un emploi rémunéré» serait plus grand que le coefficient de variation de la proportion correspondante de «jeunes de 15 à 19 ans qui avaient un emploi rémunéré». Ainsi, si le coefficient de variation de la proportion ne peut pas être diffusé, alors le coefficient de la variation de l'estimation quantitative correspondante ne peut pas l'être non plus.

Les coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise en utilisant une technique appelée pseudo-réplication. Elle consiste à diviser les enregistrements sur les fichiers de microdonnées en sous-groupes (ou répliquations) et à déterminer la variation dans l'estimation d'une réplification à l'autre. Les utilisateurs qui désirent calculer les coefficients de variation pour les estimations quantitatives peuvent communiquer avec Statistique Canada pour obtenir des conseils sur l'attribution des enregistrements aux répliquations appropriées et sur les formules à utiliser pour effectuer ces calculs.

10.8 Seuils de diffusion à l'égard de l'Enquête sur le tabagisme chez les jeunes

La taille minimale de l'estimation aux niveaux provincial, régional et canadien est précisée dans le tableau ci-dessous. Les estimations inférieures à la taille minimale donnée dans la colonne «Non diffusable» ne peuvent être diffusées en aucun cas. Il faut signaler qu'il y a trois tableaux ayant des seuils de diffusion pour l'Enquête sur le tabagisme chez les jeunes, un pour la composante EPA, un autre pour la composante École et un dernier tableau pour les deux composantes. Ces tableaux doivent être utilisés avec les estimations appropriées. Par exemple, si des estimations sont créées pour le groupe des 10-14 ans seulement (en utilisant seulement le fichier de la composante École), le tableau qu'il convient de consulter serait alors le tableau des seuils de diffusion pour la composante École. La même

procédure s'applique si des estimations sont créées pour le groupe des 15-19 ans (en utilisant le fichier de la composante EPA seulement); c.-à-d., il faut consulter le tableau des seuils de diffusion pour la composante EPA. Cependant, lorsque des estimations sont créées pour le groupe des 10-19 ans (en enchaînant les deux fichiers), il faut alors utiliser le troisième tableau de la série, lequel est le tableau des seuils de diffusion pour les deux composantes.

Tableau des seuils de diffusion -- Composante EPA

Province	Sans restriction	Avec restriction	Confidentiel	Non diffusable
Terre-Neuve	3 500 et +	1 500-3 500	1 000-1 500	moins de 1 000
Île-du-Prince-Édouard	1 500 et +	1 000-1 500	500-1 000	moins de 500
Nouvelle-Écosse	5 000 et +	2 500-5 000	1 500-2 500	moins de 1 500
Nouveau-Brunswick	3 000 et +	1 500-3 000	1 000-1 500	moins de 1 000
Québec	33 000 et +	15 000-33 000	8 500-15 000	moins de 8 500
Ontario	74 000 et +	34 500-74 000	20 000-34 500	moins de 20 000
Manitoba	6 500 et +	3 000-6 500	1 500-3 000	moins de 1 500
Saskatchewan	4 000 et +	2 000-4 000	1 000-2 000	moins de 1 000
Alberta	13 000 et +	6 000-13 000	3 500-6 000	moins de 3 500
Colombie-Britannique	19 000 et +	8 500-19 000	5 000-8 500	moins de 5 000
Provinces de l'Atlantique	4 000 et +	1 500-4 000	1 000-1 500	moins de 1 000
Prairies	10 500 et +	4 500-10 500	2 500-4 500	moins de 2 500
CANADA	40 000 et +	17 500-40 000	10 000-17 500	moins de 10 000

Tableau des seuils de diffusion -- Composante École

Province	Sans restriction	Avec restriction	Confidentiel	Non diffusable
Terre-Neuve	2 000 et +	1 000-2 000	500-1 000	moins de 500
Île-du-Prince-Édouard	500 et +	250-500	50-250	moins de 50
Nouvelle-Écosse	3 000 et +	1 500-3 000	500-1 500	moins de 500
Nouveau-Brunswick	3 000 et +	1 500-3 000	1 000-1 500	moins de 1 000
Québec	28 500 et +	12 500-28 500	7 500-12 500	moins de 7 500
Ontario	40 000 et +	18 000-40 000	10 500-18 000	moins de 10 500
Manitoba	4 000 et +	2 000-4 000	1 000-2 000	moins de 1 000
Saskatchewan	4 000 et +	1 500-4 000	1 000-1 500	moins de 1 000
Alberta	10 000 et +	4 500-10 000	2 500-4 500	moins de 2 500
Colombie-Britannique	12 500 et +	5 500-12 500	3 000-5 500	moins de 3 000
Provinces de l'Atlantique	2 500 et +	1 000-2 500	500-1 000	moins de 500
Prairies	7 500 et +	3 500-7 500	2 000-3 500	moins de 2 000
CANADA	24 500 et +	11 000-24 500	6 000-11 000	moins de 6 000

Tableau des seuils de diffusion -- Les deux composantes

Province	Sans restriction	Avec restriction	Confidentiel	Non diffusable
Terre-Neuve	3 000 et +	1 500-3 000	1 000-1 500	moins de 1 000
Île-du-Prince-Édouard	1 000 et +	500-1 000	250-500	moins de 250
Nouvelle-Écosse	4 500 et +	2 000-4 500	1 000-2 000	moins de 1 000
Nouveau-Brunswick	3 000 et +	1 500-3 000	1 000-1 500	moins de 1 000
Québec	32 000 et +	14 000-32 000	8 000-14 000	moins de 8 000
Ontario	72 500 et +	32 500-72 500	18 500-32 500	moins de 18 500
Manitoba	6 500 et +	3 000-6 500	1 500-3 000	moins de 1 500
Saskatchewan	4 000 et +	2 000-4 000	1 000-2 000	moins de 1 000
Alberta	13 000 et +	6 000-13 000	3 500-6 000	moins de 3 500
Colombie-Britannique	17 500 et +	7 500-17 500	4 500-7 500	moins de 4 500
Provinces de l'Atlantique	4 000 et +	2 000-4 000	1 000-2 000	moins de 1 000
Prairies	9 500 et +	4 000-9 500	2 500-4 000	moins de 2 500
CANADA	33 500 et +	14 500-33 500	8 000-14 500	moins de 8 000

10.9 Tableaux sur les c.v.

De façon à garantir la qualité d'une estimation individuelle, les principes énoncés à la section 10.8 doivent être suivis au moment de consulter les tableaux sur les c.v. Cela signifie que, lorsqu'une estimation a été créée pour le groupe des 10-14 ans seulement, on doit se reporter aux tableaux sur les c.v. pour la composante École. Lorsqu'une estimation a été créée pour le groupe des 15-19 ans seulement, on doit se reporter aux tableaux sur les c.v. pour la composante EPA. Finalement, lorsqu'une estimation a été créée pour le groupe des 10-19 ans, on doit consulter la troisième série de tableaux sur les c.v. qui a été créée pour les deux composantes.

11 PONDÉRATION

11.1 Pondération pour la composante EPA

Comme l'Enquête sur le tabagisme chez les jeunes (composante EPA) a utilisé un sous-échantillon de l'échantillon de l'EPA, le calcul des poids pour les enregistrements de l'enquête était clairement associé à la procédure de pondération utilisée pour l'EPA. Celle-ci est décrite brièvement ci-dessous, et une description de la procédure qui a été suivie aux fins de l'ETJ vient ensuite.

11.1.1 Procédures de pondération pour l'EPA

Dans l'EPA, le poids final associé à chaque enregistrement est le produit des facteurs suivants : le poids de base, le sous-poids des grappes, le facteur de compensation pour les non-réponses, le facteur rural-urbain et le facteur d'ajustement du ratio province-âge-sexe. Chacun de ces facteurs est décrit ci-dessous.

Poids de base

Dans un échantillon probabiliste, le plan d'échantillonnage lui-même détermine quels poids doivent être utilisés pour produire des estimations non biaisées de la population. Chaque enregistrement doit être pondéré par l'inverse de la probabilité de choisir la personne à laquelle l'enregistrement fait référence. Dans l'exemple d'un échantillon aléatoire simple de 2 %, cette probabilité serait de 0,02 % pour chaque personne, et les enregistrements doivent être pondérés par $1/0,02 = 50$. Comme toutes les personnes admissibles dans un logement sont interviewées (directement ou par procuration), cette probabilité est fondamentalement la même que la probabilité que le logement soit choisi.

Sous-poids des grappes

La délimitation des grappes est telle que le nombre de logements inclus dans l'échantillon augmente très légèrement, avec une croissance modérée, dans l'ensemble des logements. Une croissance importante peut être tolérée dans une grappe isolée avant que l'échantillon additionnel pose un problème de collecte sur le terrain. Cependant, si une croissance touche plus d'une grappe faisant partie de la tâche d'un intervieweur, l'effet cumulatif de toutes les augmentations peut créer un problème. Dans les grappes, où une croissance substantielle a été enregistrée, on utilise un sous-échantillonnage comme moyen de maintenir la charge de travail de l'intervieweur à un niveau acceptable. Le sous-poids des grappes représente l'inverse de ce ratio de sous-échantillonnage dans les grappes où a eu lieu un sous-échantillonnage.

Non-réponse

Malgré les contrôles stricts de l'EPA, il est inévitable qu'on obtienne des non-réponses et ce,

en dépit de toutes les tentatives faites par les intervieweurs. Le taux de non-réponse de l'EPA est de 5 % environ. Pour certains types de non-réponses (p.ex., ménage temporairement absent, refus), les données recueillies auprès du ménage au cours d'une interview d'un mois précédent, s'il y a lieu, sont rappelées et utilisées comme données du mois courant pour ce ménage.

Dans d'autres cas, on compense les non-réponses en augmentant proportionnellement les poids des ménages répondants. Le poids de chaque enregistrement de réponse est augmenté par le ratio du nombre de ménages qui auraient dû être interviewés, divisé par le nombre qui l'ont réellement été. Cet ajustement est effectué séparément pour les régions géographiques appelées unités de compensation. Il est fondé sur l'hypothèse suivant laquelle les ménages qui ont été interviewés représentent les caractéristiques de ceux qui auraient dû être interviewés. Dans la mesure où cette hypothèse est fautive, les estimations seront quelque peu biaisées.

Facteur rural-urbain

Dans les UNAR où la population rurale et urbaine est insuffisante pour former des strates urbaines et rurales explicites, chaque unité primaire d'échantillonnage (UPÉ) est composée de parties à la fois urbaine et rurale. Les données sur la population totale dans les régions urbaines et rurales peuvent être tirées du Recensement de 1981 pour chaque UPÉ ainsi que pour chaque région économique (RE) où une stratification urbaine/rurale explicite n'est pas effectuée. Une comparaison par RE avec les chiffres ruraux ou urbains réels du Recensement de 1981 montre si les UPÉ sélectionnés sous-représentent ou sur-représentent les régions respectives. Le ratio des chiffres ruraux-urbains réels est divisé par les estimations correspondantes. Ces deux facteurs sont calculés pour chaque RE pertinente au moment de la sélection des UPÉ et sont entrés dans chaque enregistrement d'échantillon, en fonction de la région appropriée (rurale ou urbaine) de l'UNAR. Tout changement à ces facteurs est incorporé au moment du renouvellement des UPÉ.

Sous-poids de l'EPA

Le produit des facteurs de pondération décrits plus haut est appelé «sous-poids de l'EPA». Tous les membres du même logement échantillonné ont le même sous-poids.

Ajustements infraprovinciaux et du groupe province-âge-sexe

On peut utiliser un sous-poids pour obtenir une estimation valide de toute caractéristique pour laquelle des données ont été recueillies par l'EPA. Notamment, des estimations sont produites à l'égard du nombre total de personnes de 15 ans et plus dans les régions économiques provinciales et dans les 24 grandes régions métropolitaines, ainsi qu'à l'égard du nombre de groupes âge-sexe désignés dans chacune des dix provinces.

Des estimations indépendantes sont disponibles mensuellement pour divers groupes d'âge et

de sexe, par province. Ces projections de population sont fondées sur les plus récentes données du Recensement, les enregistrements de naissance et de décès, et les estimations de migration. En dernière étape, ces données auxiliaires servent à transformer le sous-poids en poids final. À cette fin, on utilise un modèle de régression linéaire. La régression est établie de manière à garantir que le total des poids finaux produits correspond aux projections du recensement pour les variables auxiliaires, notamment les divers groupes âge-sexe, les régions économiques et les régions métropolitaines de recensement.

Cette procédure de pondération garantit l'uniformité avec les chiffres du Recensement externes et garantit également que le même poids a été attribué à chaque membre de la famille économique.

11.1.2 Procédures de pondération pour l'Enquête sur le tabagisme chez les jeunes

Les principes qui sous-tendent le calcul des poids pour l'Enquête sur le tabagisme chez les jeunes sont identiques à ceux adoptés aux fins de l'EPA. Cependant, d'autres ajustements ont dû être effectués à l'égard des poids de l'EPA de façon à obtenir un poids final pour les enregistrements individuels sur le fichier de microdonnées de l'Enquête sur le tabagisme chez les jeunes. Ces ajustements sont décrits ci-dessous :

- 1) Un ajustement visant à tenir compte des non-réponses additionnelles à l'enquête supplémentaire, c'est-à-dire, les non-réponses à l'Enquête sur le tabagisme chez les jeunes pour les personnes qui ont répondu à l'EPA ou pour qui les données de l'EPA du mois précédent ont été copiées. Cet ajustement tient également compte du poids attribué aux jeunes de 15 à 19 ans qui ont été exclus de l'enquête en raison de restrictions imposées sur les ménages dans certains groupes de renouvellement (décrits à la section 5.5). Les poids doivent être ajustés pour tenir compte de ces jeunes puisqu'ils sont admissibles du fait de leur âge, mais n'ont pas été interviewés pour l'ETJ en raison de considérations liées au fardeau de réponse. En principe, ces jeunes ne sont pas des non-répondants, bien qu'ils seront inclus dans l'ajustement des non-réponses. De fait, ils ne sont pas inclus dans les taux de réponse.
- 2) Un ajustement pour tenir compte de l'utilisation d'un échantillon élargi, sauf en Ontario et au Québec, où on a utilisé un sous-échantillon plutôt que l'échantillon complet de l'EPA.
- 3) Après que les ajustements précités ont été effectués, on procède à un réajustement dans le but de tenir compte des projections indépendantes selon la province, l'âge et le sexe. Il faut signaler le fait que, comme les âges d'admissibilité sont de 15 à 19 ans seulement, les ajustements dans l'ETJ sont effectués par année simple plutôt que par groupes d'âge, comme cela a été le cas dans l'EPA.

On tient compte de l'ajustement (1) en multipliant le sous-poids de l'EPA pour chaque

enregistrement de réponse de l'Enquête sur le tabagisme chez les jeunes :

$$\frac{\text{somme des sous-poids de l'EPA chez les jeunes admissibles qui ont répondu à l'EPA}}{\text{somme des sous-poids de l'EPA chez les jeunes admissibles qui ont répondu à l'ETJ}}$$

pour obtenir un sous-poids de non-réponse ajusté pour l'Enquête sur le tabagisme chez les jeunes (POIDS 1).

On effectue l'ajustement (2) en multipliant le POIDS 1 pour chaque enregistrement de réponse de l'ETJ par :

$$\frac{\text{nombre de groupes de renouvellement dans l'EPA}}{\text{nombre de groupes de renouvellement choisis pour l'ETJ}}$$

pour obtenir un sous-poids ajusté de l'Enquête sur le tabagisme chez les jeunes (POIDS 2). Signalons le fait que le numérateur dans l'équation ci-dessus sera six (6) dans tous les cas puisqu'il y a toujours six groupes de renouvellement dans l'EPA. Le dénominateur différera selon la province puisqu'on a choisi un nombre différent de groupes de renouvellement pour les diverses provinces (voir le tableau en 5.1.5).

À cette étape, le poids comprend trois composantes : le sous-poids de l'EPA, l'ajustement des non-réponses et l'ajustement de la sélection des groupes de renouvellement. Une quatrième composante, l'ajustement du groupe province-sexe-âge décrite ci-dessous a été ajoutée pour accroître la précision des estimations.

Comme nous l'avons déjà mentionné, des estimations indépendantes sont disponibles mensuellement pour divers groupes d'âge et de sexe, par province. Ces projections de population sont fondées sur les données du dernier recensement, les enregistrements de naissance et de décès et les estimations de migration. En utilisant une estimation de calage, les poids sont calés ou ajustés en utilisant des données auxiliaires pour obtenir un poids final. On effectue un calage pour garantir que le total des poids finaux produits correspond aux projections du recensement pour les variables auxiliaires, notamment les divers groupes âge-sexe pour chaque province. Précisons que, contrairement à l'EPA, l'ETJ a utilisé des années simples plutôt que des groupes d'âge. Ce processus rehausse la fiabilité des estimations qui peuvent être produites à partir de l'Enquête sur le tabagisme chez les jeunes.

L'ajustement (3) est effectué par le programme d'estimation de calage. Il faut simplement multiplier le POIDS 2 pour chaque répondant de l'Enquête sur le tabagisme chez les jeunes:

$$\frac{\text{population totale pour le groupe province-sexe-âge } i}{\text{somme du POIDS 2 pour les répondants de l'enquête dans le groupe province-sexe-âge } i}$$

Le poids résultant (FINWT) est le poids final qui figure sur le fichier de microdonnées de l'Enquête sur le tabagisme chez les jeunes.

Mise en garde : Les poids s'appliquent aux répondants et aux données sur les répondants seulement, et non aux membres du ménage et aux parents du répondant. Le plan d'échantillonnage est tel qu'il n'est pas approprié de créer des totaux pondérés pour les variables qui font référence à des données sur les parents et les membres du ménage, sauf s'ils sont combinés avec ceux du répondant. Par exemple, il est tout à fait acceptable de trouver le nombre total de répondants de 15 à 19 ans qui fument et qui vivent avec un parent seulement. Cependant, il ne serait pas approprié de calculer le nombre total de parents célibataires qui ont des enfants de 15 à 19 ans qui fument.

11.2 Pondération pour la composante Écoles

La détermination du poids associé à chaque enregistrement requiert plusieurs étapes, chacune d'entre elles reflétant soit un degré d'échantillonnage, soit un ajustement particulier.

1. Poids de l'école

On commence par calculer le poids initial de chaque école sélectionnée. Celui-ci est égal à l'inverse de la probabilité de sélection de l'école dans la strate considérée. Comme on l'a vu dans la section 5.2.5, cette probabilité est proportionnelle à l'importance des effectifs de l'école pour le niveau scolaire en question.

2. Ajustement pour la non-réponse au niveau de l'école

Cet ajustement tient compte du fait que certaines écoles sélectionnées ne sont pas sur le fichier final. Puisque le nombre total d'écoles sélectionnées dans chaque strate était toujours égal à 16, on multiplie le poids de l'école par le ratio suivant:

$$\frac{16}{\text{nombre d'écoles répondantes pour cette strate}}$$

3. Ajustement pour la sélection d'une classe

Cet ajustement correspond en fait au deuxième degré d'échantillonnage, lorsqu'une seule classe de l'école est sélectionnée parmi toutes celles du niveau scolaire considéré. On multiplie simplement le poids obtenu à l'étape précédente par le nombre total de classes de l'école pour ce niveau.

4. Ajustement pour la non-réponse des élèves

Cet ajustement vise à représenter les élèves pour lesquels un enregistrement final n'a pas été obtenu, peu importe la raison. On multiplie le poids obtenu à l'étape précédente par le ratio suivant:

$$\frac{\text{nombre d'élèves éligibles dans la classe}}{\text{nombre d'élèves répondants dans cette classe}}$$

5. Ajustement par province-âge-sexe

Ce dernier ajustement vise à faire concorder les totaux de l'enquête aux projections calculées à partir des données du recensement pour la population-cible. Cet ajustement est effectué au niveau province-âge-sexe, donc pour un total de $10*5*2=100$ cellules. Le poids obtenu à l'étape précédente est multiplié par le ratio suivant:

$$\frac{\text{population projetée pour cellule province-âge-sexe}}{\text{à laquelle appartient l'enregistrement}} \\ \frac{\text{somme des poids de tous les enregistrements}}{\text{appartenant à cette cellule}}$$

12 PLAN DE L'ENREGISTREMENT

L'ordre suivant lequel les variables du questionnaire paraissent dans le cliché d'article est le même que celui qu'on trouve dans la composante EPA. Le nom de la variable contient le numéro de la question pour les deux composantes. Par exemple, la variable «Q9_17» indique qu'il s'agit de la question n° 9 dans la composante EPA et de la question 17 dans la composante École. Dans certains cas, comme le champ compte 8 bytes au plus, on a limité cette stratégie de numérotation; par exemple, dans «Q11B1_19», il n'est pas explicite que la question de la composante École ait également une partie B1.

Quelques questions posées dans la composante EPA n'ont pas été posées dans la composante École, et vice-versa. Pour ces cas, on a utilisé XX comme numéro pour la question non existante; par exemple, Q14B1_XX ou QXX_68A.

Les variables créées aux fins du regroupement de données confidentielles commencent par «CL» ou par «GRP», et sont données à la fin du cliché d'article.

La section des commentaires apparaissant sous toutes les variables comprend un énoncé universel qui montre, par numéro de question, quels répondants ont répondu à chaque question.

Dans le cliché d'article, les questions à réponses multiples sont représentées par plusieurs variables, chaque variable ayant une valeur de réponse «oui», «non», «enchaînement valide», «non déclaré». Le nombre de variables dépend du nombre de réponses possibles (une variable pour chaque réponse). Lorsqu'une réponse particulière était indiquée, on cochait «oui». Autrement, on cochait «non», signifiant aucune réponse particulière. On cochait «non déclaré» seulement si aucune réponse n'était donnée pour la question entière et s'il s'agissait d'une question à laquelle le répondant aurait dû répondre.

Toutes les variables s'assortissent d'une série de codes normalisés pour signaler les valeurs manquantes. Ces valeurs sont les suivantes :

6 (ou 96, 996, etc.), si la question a été sautée en raison d'une réponse valide à une question précédente (c'est-à-dire, non applicable);

7 (ou 97, 997, etc.), si la réponse à la question était «Ne sait pas»;

9 (ou 99, 999, etc.), pour indiquer que la réponse était «Non déclaré».

Une courte description (carte D), d'une longueur de 40 bytes, a été écrite pour chaque question et peut être utilisée dans le logiciel de calcul.

13 QUESTIONNAIRES ET FEUILLES DE CODES

- Composante EPA (supplément) de l'ETJ (Formule 08)
- Composante École de l'ETJ (Formule 08S)
- Questionnaire à l'intention des parents de la composante École (Formule 03S)
- Dossier du ménage (Formule 03) et Feuille de codes
- Questionnaire de l'Enquête sur la population active (Formule 05) et Feuille de codes

13.1 Composante EPA (supplément) de l'ETJ (Formule 08)

Le questionnaire de la composante EPA (Formule 08) a servi à recueillir des données sur l'usage du tabac chez les jeunes de 15 à 19 ans. Un intervieweur posait les questions au téléphone.