ENQUÊTE DE SURVEILLANCE DE L'USAGE DU TABAC AU CANADA - 2000

FICHIERS DU CYCLE 1

FÉVRIER - JUIN 2000

Table des matières

1.0 Introduction 1

2.0 Contexte 3

3.0 Objectifs 5

4.0 Concepts et définitions 7

5.0 Méthodologie de l'enquête 9
5.1 Champ de l'enquête 9
5.2 Stratification 9
5.3 Plan de sondage et répartition de l'échantillon 10
5.4 Tirage de l'échantillon 10

6.0 Collecte des données 13
6.1 Conception du questionnaire 13
6.2 Collecte et vérification des données 13

7.0 Traitement des données 15
7.1 Saisie des données 15
7.2 Vérification 15
7.3 Création de variables calculées 15
7.4 Pondération 16
7.5 Suppression des renseignements confidentiels 16

8.0 Qualité des données 17
8.1 Taux de réponse des ménages - février à juin 2000 18
8.2 Taux de réponse des personnes - février à juin 2000 19
8.3 Erreurs dans l'enquête 21
8.4 Non-réponse totale 22
8.5 Non-réponse partielle 22
8.6 Couverture 22
8.7 Mesure de l'erreur d'échantillonnage 23

9.0 Lignes directrices pour la publication et la diffusion 25
9.1 Lignes directrices pour l'arrondissement 25
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation 26
9.2.1 Définitions des genres d'estimations: de type nominal par opposition à quantitatives 27
9.2.2 Totalisation d'estimations de type nominal 28
9.2.3 Totalisation d'estimations quantitatives 28
9.3 Lignes directrices pour l'analyse statistique 29
9.4 Lignes directrices pour la diffusion en fonction de la variabilité d'échantillonnage 30

10.0 Tables de variabilité d'échantillonnage approximative 33
10.1 Comment utiliser les tables de c.v. pour les estimations de type nominal 36
10.1.1 Exemples d'utilisation des tables de c.v. pour des estimations de type nominal 38
10.2 Comment utiliser les tables de c.v. pour calculer des limites de confiance 41
10.2.1 Exemple d'utilisation des tables de c.v. pour obtenir des limites de confiance 42
10.3 Comment utiliser les tables de c.v. pour effectuer un test t 43
10.3.1 Exemple d'utilisation des tables de c.v. pour effectuer un test t 43
10.4 Coefficients de variation d'estimations quantitatives 44
10.5 Seuils pour la diffusion des estimations relatives à l'ESUTC - fichier ménages . . 44
10.6 Seuils pour la diffusion des estimations relatives à l'ESUTC - fichier personnes . 45
10.7 Tables de c.v. - fichier ménage 47
10.8 Tables de c.v. - fichier Personnes 47

11.0 Pondération 49
11.1 Pondération pour les fichiers Ménage et Personne 49
11.2 Pondération pour le fichier Ménage 51
11.3 Pondération pour le fichier Personne 51

12.0 Questionnaire 55

13.0 Cliché d'article et variables uniques 57
13.1 Cliché d'article et variables uniques - Ménages 57
13.2 Cliché d'article et variables uniques - Personne 57

1.0 Introduction

L'Enquête de surveillance de l'usage du tabac au Canada a été menée par Statistique Canada entre février et juin 2000, avec la collaboration et l'appui de Santé Canada. Ce manuel a été rédigé en vue de faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Veuillez adresser toutes les demandes concernant l'ensemble des données ou leur utilisation à :

Statistique Canada

Eddy Ross
Division des enquêtes spéciales, Statistique Canada
Section D7
5e étage, immeuble Jean-Talon
Tunney's Pasture
Ottawa (Ontario) K1A 0T6
Téléphone : (613) 951-3240
Télécopieur : (613) 951-0562
Courrier électronique : rossedd@statcan.ca

Santé Canada

Murray Kaiserman
Programme de la lutte au tabagisme
123 Slater, 7ième étage,
Ottawa, Ontario K1A OK9
Téléphone: (613) 954-5851
Télécopieur: (613) 941-1551
Couriel: Murray_Kaiserman@hc-sc.gc.ca

Anne Zaborski
Programme de la lutte au tabagisme
123 Slater, 7ième étage,
Ottawa, Ontario K1A OK9
Téléphone: (613) 954-0152
Télécopieur: (613) 941-1551
Couriel: Anne_Zaborski@hc-sc.gc.ca

IL EST IMPORTANT POUR LES UTILISATEURS DE SE FAMILIARISER AVEC LE CONTENU DE CE DOCUMENT AVANT DE PUBLIER OU DE DIFFUSER TOUTE ESTIMATION PRODUITE À PARTIR DU FICHIER DE MICRODONNÉES DE L'ENQUÊTE SUR LE TABAGISME AU CANADA.

2.0 Contexte

Depuis les années 60, Statistique Canada mène des enquêtes spéciales sur le tabagisme pour Santé Canada sous forme de suppléments à l'Enquête sur la population active du Canada et d'enquêtes téléphoniques à composition aléatoire. Les premières enquêtes portaient uniquement sur le tabagisme, mais depuis quelques années, le tabagisme est devenu un sujet d'enquêtes plus élargies sur la santé. Ces enquêtes ont été menées plus ou moins fréquemment pendant les 10 dernières années.

En février 1994, une modification de la législation qui autorisait la diminution des taxes sur les cigarettes a été adoptée. Comme il n'existait aucune donnée d'enquête datant d'immédiatement avant l'adoption de cette modification, mesurer exactement l'impact de ce changement était difficile pour Santé Canada et pour les analystes intéressés .

Santé Canada souhaite être en mesure de mieux réagir à de futurs changements, l'Enquête de surveillance de l'usage du tabac a été développée en vue de fournir à Santé Canada et ses partenaires des données continues et prévisibles sur l'usage du tabac et des sujets connexes.

Cette publication est la troisième pour l'Enquête de surveillance de l'usage du tabac au Canada. La première publication faite en janvier 2000 couvrait la période de févier à juin 1999 et en octobre 2000 les données pour la période juillet à décembre 1999 de même qu'un sommaire pour l'année 1999 ont été publiés. Cette publication comprend les données pour la période de février à juin 2000.

3.0 Objectifs

Le principal objectif de l'enquête est de fournir en permanence une source de données sur la prévalence de l'usage du tabac permettant de suivre l'évolution de la prévalence. Cet objectif est distinct de celui de l'Enquête nationale sur la santé de la population (ENSP), qui recueille des données sur l'usage du tabac à partir d'un échantillon longitudinal afin de déterminer quelles sont les personnes qui changent leurs habitudes de fumer, les facteurs possibles qui contribuent à ce changement ainsi que les facteurs de risque possibles pour ceux qui commencent à fumer et qui continuent à le faire plus ou moins longtemps. Comme l'ENSP recueille des données tous les deux ans et que celles-ci sont diffusées environ un an après la fin du cycle de collecte, elle ne comble pas les besoins de Santé Canada qui exige une couverture temporelle continue et une diffusion rapide. De plus, les données de l'ENSP ne sont pas suffisamment détaillées dans le cas de la population la plus à risque, soit les 15 à 24 ans.

L'Enquête de surveillance de l'usage du tabac au Canada aidera Santé Canada à observer la prévalence de l'usage du tabac par province et par groupe d'âge-sexe (15 à 19 ans, 20 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 ans ou plus) deux fois par an. Il sera possible de déceler annuellement un changement de 3 % ou plus dans la prévalence de l'usage du tabac parmi les groupes d'âge à l'échelle nationale. Les données comprises dans ces fichiers représentent les données recueillies pour les six premiers mois de 2000. Les données de 1999 ont été diffusées en octobre 2000. Les données continueront à être recueillies de façon constante, selon la disponibilité des fonds.

4.0
Concepts et définitions

Étant donné que l'Enquête de surveillance de l'usage du tabac au Canada a été réalisée par téléphone, on a fait appel à une terminologie facile à comprendre pour éviter les longues explications. L'analyse et l'interprétation des données exigent néanmoins le recours à des concepts et définitions standards. Les questions de l'enquête ont été conçues à la lumière de ces définitions.

Situation actuelle

1. Fumeur quotidien :	Personne qui fume actuellement des cigarettes tous les jours.
2. Fumeur occasionnel :	Personne qui fume des cigarettes à l'heure actuelle, mais non tous les jours.
3. Non-fumeur :	Personne qui ne fume pas la cigarette à l'heure actuelle.
4. Fumeur :	Personne qui fume actuellement des cigarettes quotidiennement ou à l'occasion.

Antécédents

1. Ancien fumeur :	Personne qui a fumé au moins 100 cigarettes au cours de sa vie, mais qui ne fume pas à l'heure actuelle.
2. Fumeur expérimental :	Personne qui a fumé au moins une cigarette, mais moins de 100 cigarettes, et qui ne fume pas la cigarette à l'heure actuelle.
3. Abstinent à perpétuité :	Personne n'a jamais fumé de cigarette de sa vie.
4. Personne ayant déjà fumé :	Personne qui est un fumeur ou un ancien fumeur.
5. Personne n'ayant jamais fumé :	Fumeur expérimental ou abstinent à perpétuité.

Prévalence de l'usage de la cigarette

Proportion de la population qui fume des cigarettes à l'heure actuelle.

5.0 Méthodologie de l'enquête

Le premier cycle de l'Enquête de surveillance de l'usage du tabac au Canada a été réalisé entre le 1ier février et le 30 juin 2000 par enquête téléphonique à composition aléatoire (CA), une technique qui consiste à générer des numéros de téléphone au hasard par ordinateur. L'interview a été menée au téléphone.

5.1 Champ de l'enquête

Toute personne âgée de 15 ans ou plus et vivant au Canada faisait partie de la population cible de l'Enquête de surveillance de l'usage du tabac au Canada, exception faite des personnes suivantes:

1. les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut;
2. les pensionnaires à temps plein d'un établissement institutionnel.

Comme l'enquête se faisait à partir d'un échantillon de numéros de téléphone, les ménages (et donc tous ceux qui en font partie) ne possédant pas de téléphone ont été exclus de l'échantillon. Les personnes n'ayant pas de téléphone représentent moins de 3 % de la population cible. Toutefois, les estimations de l'enquête ont été pondérées afin d'inclure les personnes n'ayant pas de téléphone.

5.2 Stratification

Pour s'assurer que des personnes de partout au Canada soient représentées dans l'échantillon, on a divisé chacune des dix provinces en strates ou aires géographiques. En général, à l'intérieur de chaque province, une strate de régions métropolitaines de recensement (RMR) et une autre de régions autres que des RMR étaient délimitées. L'Île-du-Prince-Édouard ne possédait qu'une seule strate. En Ontario et au Québec, Toronto et Montréal possédaient chacune une troisième strate. Les RMR, régions définies aux fins du recensement, correspondent approximativement à des villes d'au moins 100 000 habitants.

5.3 Plan de sondage et répartition de l'échantillon

Le plan de sondage est un échantillon spécial aléatoire stratifié à deux phases de numéros de téléphone. On utilise un plan de sondage à deux phases afin d'accroître la représentation des personnes appartenant aux groupes d'âge de 15 à 19 ans ou de 20 à 24 ans dans l'échantillon. Durant la première phase, on sélectionne les ménages par CA et durant la seconde, une ou deux personnes (ou aucune) sont sélectionnées d'après la composition du ménage.

Comme l'objectif principal de l'enquête est de produire des estimations fiables dans les dix provinces, on vise ***un nombre égal de répondants dans chaque province***. Durant les six premiers mois de collecte, on a cherché à obtenir les réponses de 5 000 personnes âgées entre 15 et 24 ans et de 5 000 personnes âgées de 25 ans ou plus à l'échelle du Canada, c'est-à-dire de 500 personnes de chaque groupe d'âge par province. La taille initiale de l'échantillon de numéros de téléphone dépendait du taux de réponse prévu et du taux de succès prévu de la CA (proportion des numéros de téléphones échantillonnés appartenant à des ménages). Pour obtenir les tailles d'échantillon voulues, il a fallu apporter deux rajustements à la méthodologie standard de la CA. Premièrement, les probabilités de sélection au sein du ménage étaient inégales. Deuxièmement, les ménages comptant uniquement des personnes de 25 ans ou plus ont été sous-échantillonnés. On a estimé qu'au total, environ 148 000 numéros de téléphones par année seraient nécessaires pour obtenir les 20 000 répondants par année requis. Cela suppose un taux de réponse de 75 % et qu'environ 20 % des ménages contiennent des personnes âgées entre 15 et 24 ans. Le taux de succès varie considérablement selon la province, avec une moyenne générale prévue d'environ 43 %.

5.4 Tirage de l'échantillon

L'échantillon de l'Enquête de surveillance de l'usage du tabac au Canada a été créé à l'aide d'une méthode améliorée d'échantillonnage par CA, appelée la méthode d'élimination des banques non valides (EBNV). Pour chaque combinaison province-strate, une liste de banques valides (indicatif régional suivi de cinq chiffres) a été dressée à partir des fichiers administratifs des compagnies de téléphone. Une banque valide se définit, pour les fins des enquêtes sociales, comme une banque qui contient au moins un numéro de téléphone résidentiel valide. Ainsi, toutes les banques ayant seulement des numéros de téléphone non assignés, non valides, ou d'affaires ont été exclues de la base de sondage.

Ensuite, un échantillon systématique des banques (avec banques de remplacement) a été sélectionné à l'intérieur de chaque strate. Pour chaque banque sélectionnée, un numéro à deux chiffres (de 00 à 99) a été produit au hasard. Ce numéro aléatoire a été ajouté à la banque pour former un numéro de téléphone complet. Grâce à cette méthode, il était possible d'incorporer dans l'échantillon des numéros de téléphone résidentiels publiés et non publiés, de même que des numéros de téléphone d'affaires ou non valides (c.-à-d. qui ne sont pas ou qui n'ont jamais été en service).

Chaque numéro de téléphone de l'échantillon a été composé pour vérifier si il correspondait bien à un ménage. Si c'était le cas, on demandait à l'interlocuteur de fournir des informations sur les membres individuels du ménage. On utilisait l'âge des membres pour déterminer qui dans le ménage serait sélectionné pour l'interview sur l'usage du tabac. Les interviews par procuration n'étaient pas acceptées.

Pour s'assurer de joindre un nombre suffisant de personnes appartenant aux jeunes groupes d'âge, la sélection aléatoire a été structurée pour qu'au moins une personne âgée entre 15 et 19 ans ou entre 20 et 24 ans soit sélectionnée au sein du ménage, s'il en existait une. En effet, environ 76 % de tous les ménages au Canada comprennent seulement des personnes de plus de 25 ans, 20 % regroupent des personnes de plus de 25 ans qui vivent avec des personnes de 15 à 19 ans ou de 20 à 24 ans, et seulement 4 % des ménages ne contiennent personne âgé de plus de 25 ans. Si tous les âges étaient sélectionnés et retenus avec la même probabilité, le groupe des 25 ans et plus serait surreprésenté compte tenu des objectifs de l'enquête. De ce fait, pour économiser le coût d'interviews additionnels, une partie des personnes sélectionnées dans le groupe d'âge des 25 ans et plus ont été rejetées et n'ont pas été interviewées sur l'usage du tabac. Lorsque plus d'un des groupes d'âge de 15 à 19 ans, de 20 à 24 ans, ou de 25 ans et plus étaient représentés dans le ménage, deux personnes étaient sélectionnées. Les deux personnes sélectionnées dans un même ménage appartenaient toujours à des groupes d'âge différents. Cela garantissait qu'aucun impact négatif sur la précision des estimations par groupe d'âge ne serait causé par une corrélation à l'intérieur du ménage. Il y avait un léger impact sur la précision de l'estimation totale pour tous les âges, mais la taille de l'échantillon était suffisante pour que les conséquences restent minimales.

La sélection des personnes s'est faite selon la logique détaillée suivante :

1. Si toutes les personnes du ménage ont entre 15 et 19 ans, une personne est sélectionnée au hasard.
2. Si toutes les personnes du ménage ont entre 20 et 24 ans, une personne est sélectionnée au hasard.
3. Si toutes les personnes du ménage ont 25 ans ou plus, une personne est sélectionné au hasard. Toutefois, cette personne n'est retenue que dans une proportion des cas.
4. Si des membres du ménage ont entre 15 et 19 ans et que les autres ont entre 20 et 24 ans, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge.
5. Si des membres du ménage ont entre 15 et 19 ans et que les autres ont 25 ans ou plus, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge. Toutefois, la

personne sélectionnée n'est retenue que dans une proportion des cas.

6. Si des membres du ménage ont entre 20 et 24 ans et que les autres ont 25 ans ou plus, deux personnes sont sélectionnées au hasard, une de chaque groupe d'âge. Toutefois, la personne sélectionnée n'est retenue que dans une proportion des cas.

7. Si les trois groupes d'âge sont représentés dans le ménage, deux des groupes d'âge sont sélectionnés au hasard et la règle 4, 5 ou 6 s'applique.

6.0 Collecte des données

6.1 Conception du questionnaire

La conception du questionnaire de cette enquête se base fortement sur l'Enquête sur le tabagisme au Canada de 1994. Quelques questions ont été ajoutées par souci de cohérence avec les enquêtes internationales qui utilisent le concept de l'habitude de consommation de tabac « durant les 30 derniers jours ». Le questionnaire contient les mêmes questions que le cycle 1 de l'ESUTC de 2000.

Des spécifications définissant les limites valides et garantissant la cohérence d'une question à l'autre ont été intégrées dans l'application ITAO dans la mesure du possible. Des contrôles de cohérences additionnels ont été réalisés durant la phase de traitement des données.

6.2 Collecte et vérification des données

Le test final de l'application ITAO a eu lieu en janvier 2000. Dès le début de février 2000, la collecte des données s'est effectuée chaque mois.

Les données ont été recueillies à l'aide des techniques de l'interview téléphonique assistée par ordinateur (ITAO). Notre système ITAO contient plusieurs modules génériques qui s'adaptent rapidement à la plupart des types d'enquête. Un module frontal contient un ensemble de codes de réponse standards, qui s'appliquent à toutes les issues possible des appels, ainsi que les scénarios correspondants qui sont lus par les intervieweurs. Une approche normalisée a été utilisée pour présenter notre organisme, le nom et le but de l'enquête, les clients de l'enquête, l'utilisation qui sera faite des résultats et la durée de l'interview. Nous avons expliqué aux répondants comment ils avaient été sélectionnés pour l'enquête, que leur participation à l'enquête était volontaire et que les renseignements fournis resteraient strictement confidentiels. Les intervieweurs avaient accès à des écrans d'aide grâce auxquels ils pouvaient répondre aux questions les plus fréquemment posées par les répondants.

L'application ITAO garantissait l'entrée des seules réponses valides et le bon enchaînement des questions. Des contrôles intégrés à l'application garantissaient la cohérence des réponses, repéraient et corrigeaient les

valeurs aberrantes et déterminaient à qui étaient posées les différentes questions. Ainsi, à la fin du processus de collecte, les données étaient déjà passablement « épurées ».

Les intervieweurs ont reçu une formation sur l'utilisation de l'application ITAO et sur le contenu de l'enquête. En plus de la formation en classe, les intervieweurs complétaient une série d'interviews simulées afin de se familiariser avec l'enquête, ses concepts et ses définitions. Tout a été mis en œuvre pour conserver le même groupe d'intervieweurs chaque mois. Cela réduisait le temps de formation et produisait des données de meilleure qualité et plus cohérentes.

Les cas ont été distribués dans quatre bureaux régionaux de Statistique Canada. La charge de travail et les intervieweurs de chaque bureau étaient supervisés par un chargé de projet. L'ordonnanceur automatique utilisé dans le système ITAO garantissait que les cas étaient assignés au hasard aux intervieweurs et que les appels se faisaient à différents moments de la journée pendant des jours différents de la semaine pour maximiser la probabilité de contact. Un maximum de 17 appels par cas ont été tentés. Une fois le maximum atteint, un intervieweur principal révisait le cas et déterminait si d'autres appels seraient tentés.

7.0 Traitement des données

Le principal produit de l'Enquête de surveillance de l'usage du tabac au Canada sont deux fichiers de microdonnées « épurés », le premier sur les ménages, le second sur les personnes. Ce chapitre présente un bref aperçu des étapes du traitement associées à la réalisation de ces fichiers.

7.1 Saisie des données

Les données ayant été recueillies à l'aide de l'ITAO, un système de collecte de données séparé n'était pas nécessaire puisque les informations étaient entrées directement dans les systèmes des bureaux régionaux par les intervieweurs durant l'interview.

7.2 Vérification

La première étape du traitement de l'enquête a été de combiner tous les fichiers mensuels en un seul fichier. Toutes les valeurs « hors-limite » des fichiers de données ont été remplacées par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Le premier type d'erreurs traitées étaient les erreurs dans l'enchaînement des questions, qui se traduisaient par des réponses à des questions qui ne s'adressaient pas au répondant (et donc auxquelles il n'aurait pas dû répondre). Dans ce cas, une vérification par ordinateur a automatiquement éliminé les données superflues d'après l'enchaînement dicté par les réponses aux questions précédentes, ou occasionnellement aux suivantes.

Le second type de contrôles décelait l'information manquante dans des questions auxquelles il aurait fallu répondre. Par exemple, lorsqu'un répondant refusait de répondre à une question à laquelle il y aurait normalement eu une réponse, cette question était codée « non déclaré ».

7.3 Création de variables calculées

Un certain nombre de données élémentaires du fichier de microdonnées ont été calculées en regroupant des postes du questionnaire pour faciliter l'analyse des données. Le nombre moyen de cigarettes fumées quotidiennement, la suffisance du revenu, etc.., sont des variables calculées.

7.4 Pondération

Le principe sur lequel repose une estimation dans un échantillon probabiliste veut que chaque personne en représente plusieurs autres qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne représente 50 membres de la population.

La pondération est l'étape où l'on calcule le nombre (poids) de personnes représentées par chaque répondant de l'enquête. Il faut utiliser ce poids dans le calcul des estimations faites à partir du fichier de microdonnées. Par exemple, si l'on veut estimer le nombre de personnes au Canada qui fument tous les jours, on sélectionne les enregistrements des personnes qui possèdent cette caractéristique (Q010 = '1') et l'on additionne les poids de ces enregistrements . La pondération des ménages et des personnes se fait séparément tous les six mois.

7.5 Suppression des renseignements confidentiels

Il convient de mentionner que le fichier de microdonnées « à grande diffusion » décrit ci-dessus diffère à certains égards importants du fichier principal de l'enquête conservé par Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants à l'enquête. Les utilisateurs qui désirent accéder à des renseignements exclus du fichier de microdonnées peuvent acheter des totalisations personnalisées. Les estimations produites seront communiquées à l'utilisateur, à la condition qu'elles soient conformes aux lignes directrices pour l'analyse et la diffusion décrites au chapitre 9 du présent document.

Identificateurs géographiques : Le fichier principal de l'enquête comprends des identificateurs géographiques explicites pour la province et la strate (RMR, régions autres que des RMR, Toronto, Montréal). Il renferme aussi le code postal des répondants. Le fichier à grande diffusion de l'enquête comprends les identificateurs géographiques pour la province et la strate, mais non les identificateurs géographiques inférieurs à la strate (c.-à-d. Le code postal). Lorsque cela est possible, le code postal peut être lié à d'autres identificateurs géographiques et l'on peut totaliser les données à partir de ce nouveau niveau géographique; la liaison et la totalisation sont exécutées contre recouvrement des coûts.

8.0
Qualité des données

Pour l'Enquête de surveillance de l'usage du tabac au Canada, les taux de réponse calculés tiennent compte :

- du taux de numéros de téléphone résolus, un numéro de téléphone résolu étant un numéro résidentiel, d'affaires ou non admissible confirmé;
- du taux de succès, un succès étant un numéro résidentiel confirmé appartenant à un ménage ou des **données valides sur le ménage**;

Le taux de téléphone résolus et le taux de succès s'appliquent à la fois au fichier Ménage et au fichier Personne.

- du taux de réponse des ménages pour le fichier des ménages, une réponse étant un ménage pour lequel l'âge de toutes les personnes sur la liste a été fourni et des **données valides sur le ménage**;

Le taux de réponse des ménages ne s'applique qu'au fichier Ménage.

- du taux de réponse des ménages pour le fichier des personnes, une réponse étant un ménage résidentiel confirmé avec un liste remplie ou des **données valides sur la personne**;
- du taux d'achèvement des listes, une réponse étant un ménage pour lequel l'âge de toutes les personnes sur la liste a été fourni;
- du taux de réponse combiné des ménages pour le fichier des personnes, une réponse étant un ménage avec une liste remplie valide;
- du taux de réponse des personnes, une réponse étant un enregistrement contenant l'âge de toutes les personnes sur la liste et des données valides sur la personne.

Le taux de réponse des ménages pour le fichier Personne, le taux de résolution des numéros, le taux de réponse combiné des ménages et ne s'appliquent que pour le fichier Personne.

Taux de numéros de téléphone résolus et taux de succès par Province

Provinces	Total des nos de téléphone produits	Total des nos résolus	Taux de nos résolus	Total des ménages	Taux de succès
Terre-Neuve	6 120	6 076	99,3%	2 155	35,5%
Île-du-Prince-Édouard	5 921	5 773	97,5%	2 507	43,4%
Nouvelle-Écosse	6 025	5 919	98,2%	2 657	44,9%
Nouveau-Brunswick	7 104	7 010	98,7%	2 178	31,1%

Québec	5 275	5 195	98,5%	2 628	50,6%
Ontario	5 610	5 446	97,1%	2 521	46,3%
Manitoba	5 999	5 929	98,8%	2 592	43,7%
Saskatchewan	5 643	5 601	99,3%	2 300	41,1%
Alberta	5 209	5 157	99,0%	2 581	50,0%
Colombie-Britannique	5 490	5 367	97,8%	2 834	52,8%
Total	58 396	57 473	98,4%	24 953	43,4%

8.1 Taux de réponse des ménages - février à juin 2000

Le **répondant d'un ménage** doit remplir la liste sans aucun refus à propos de l'âge des personnes, et des données valides sur le ménage doivent exister. Il y a eu 2,518 (10.1%) ménages non répondants.

Taux de réponse des ménages répondants par province.

Province	Nombre total de ménages	Ménages répondants	Taux de réponse des ménages
Terre-Neuve	2 155	2 050	95,1%
Île-du-Prince-Édouard	2 507	2 303	91,9%
Nouvelle-Écosse	2 657	2 436	91,7%
Nouveau-Brunswick	2 178	1 972	90,5%
Québec	2 628	2 315	88,1%
Ontario	2 521	2 099	83,3%
Manitoba	2 592	2 327	89,8%
Saskatchewan	2 300	2 100	91,3%
Alberta	2 581	2 307	89,4%
Colombie-Britannique	2 834	2 526	89,1%
Canada	24 953	22 435	89,9%

Taux de réponse des ménages répondants par mois.

Mois d'enquête	Nombre total de ménages	Ménages répondants	Taux de réponse
Février	4 976	4 529	91,0%
Mars	4 943	4 530	91,6%
Avril	4 975	4 394	88,3%
Mai	5 086	4 532	89,1%
Juin	4 973	4 450	89,5%
Canada	24 953	22 435	89,9%

8.2 Taux de réponse des personnes - février à juin 2000

Une **personne répondante** possède les caractéristiques suivantes :

- Le numéro de téléphone correspondant à la personne sélectionnée appartient à un ménage.
- La liste a été remplie sans aucun refus à propos de l'âge des personnes.
- La personne sélectionnée doit être âgée de 15 ans ou plus au moment de l'enquête. L'âge inscrit dans la liste a été vérifié à l'aide de la date de naissance donnée par la personne sélectionnée.
- La personne sélectionnée doit avoir répondu au moins aux questions clés concernant les habitudes de fumer.

Dans 12 571 ménages, des données sur le ménage ont été recueillies mais personne n'a été sélectionné pour continuer l'Enquête sur l'usage du tabac. (Pour plus d'informations, voir « Tirage de l'échantillon » au chapitre 5.4). Dans les ménages restants, on a sélectionné une personne dans 8 484 ménages et deux personnes dans 1 472 ménages. 155 (1.4%) personnes ont refusé de remplir le questionnaire et 780 (6,8%) personnes étaient non-répondantes pour d'autres raisons.

Taux de réponse des ménages, taux d'achèvement des listes et taux combiné de réponse des ménages par province

Province	Total des ménages	Total des ménages avec liste remplie	Taux de réponse des ménages	Ménages avec données de liste valides	Taux d'achèvement des listes	Taux de réponse des ménages combiné
Terre-Neuve	2 155	2 106	97,7%	2 055	97,6%	95,4%
Île-du-Prince-Édouard	2 507	2 366	94,4%	2 306	97,5%	92,0%
Nouvelle-Écosse	2 657	2 509	94,4%	2 443	97,4%	91,9%
Nouveau-Brunswick	2 178	2 047	94,0%	1 978	96,6%	90,8%
Québec	2 628	2 382	90,6%	2 324	97,6%	88,4%
Ontario	2 521	2 207	87,5%	2 109	95,6%	83,7%
Manitoba	2 592	2 451	94,6%	2 342	95,6%	90,4%
Saskatchewan	2 300	2 187	95,1%	2 115	96,7%	92,0%
Alberta	2 581	2 415	93,6%	2 316	95,9%	89,7%
Colombie-Britannique	2 834	2 622	92,5%	2 539	96,8%	89,6%
Canada	24 953	23 292	93,3%	22 527	96,7%	90,3%

Taux de réponse des personnes répondantes par province.

Province	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
Terre-Neuve	1 206	1,096	90,7%
Île-du-Prince-Édouard	1 253	1,159	92,0%
Nouvelle-Écosse	1 223	1,134	92,4%
Nouveau-Brunswick	1 075	975	90,3%
Québec	1 172	1,090	93,3%
Ontario	1 054	947	91,7%
Manitoba	1 057	978	92,1%
Saskatchewan	1 021	963	93,4%
Alberta	1 146	1,073	92,6%
Colombie-Britannique	1 221	1,078	87,6%
Canada	11 428	10,493	91,6%

Taux de réponse des personnes répondantes par mois.

Mois	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
Février	2 363	2,198	92,6%
Mars	2 350	2,180	92,1%
Avril	2 225	2,011	89,9%
Mai	2 264	2,098	92,5%
Juin	2 226	2,006	90,6%
Canada	11 428	10,493	91,6%

Nombre cible de répondants et taux de réponse des personnes par groupe d'âge

Groupe d'âge	Nombre cible de répondants	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes
15-19	2500	3 052	2 881	93,3%
20-24	2 500	2 589	2 317	90,4%
25 +	5 000	5 787	5	91,2%
Canada	10 000	11 428	10,493	91,6%

8.3 Erreurs dans l'enquête

L'enquête permet de produire des estimations fondées sur les données recueillies auprès des personnes faisant partie de l'échantillon. On aurait pu obtenir des estimations quelque peu différentes si l'on avait procédé à un recensement complet en faisant appel au même questionnaire et aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que pour cette enquête. La différence entre les estimations obtenues à partir de l'échantillon et celles tirées d'un dénombrement complet effectué dans des conditions semblables est appelée l'erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les phases de l'enquête. Il se peut que les intervieweurs comprennent mal les instructions, que les répondants se trompent en répondant aux questions, que des réponses soient mal entrées dans

l'ordinateur et que des erreurs se produisent au moment du traitement ou de la totalisation des données. Ces erreurs sont autant d'exemples d' erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations tirées de l'enquête. Toutefois, les erreurs qui se produisent systématiquement contribueront à biaiser les estimations de l'enquête. On a consacré beaucoup de temps et d'efforts à réduire le nombre d'erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment donné aux intervieweurs une formation poussée pour ce qui est des procédures d'enquête et de l'application ITAO, observé les intervieweurs pour repérer les problèmes liés à la conception du questionnaire ou à la mauvaise interprétation des instructions, et testé l'application ITAO pour s'assurer que les contrôles des limites, les vérifications et le déroulement des questions étaient tous programmés correctement.

8.4 Non-réponse totale

Dans bien des enquêtes, la non-réponse totale peut être une source importante d'erreurs non dues à l'échantillonnage, selon la mesure dans laquelle les répondants et les non-répondants diffèrent quant aux caractéristiques présentées. Il y a non-réponse totale lorsque la personne sélectionnée n'a pas pu être contactée ou lorsqu'elle a refusé de participer à l'enquête. On a traité la non-réponse totale en ajustant le poids des ménages ou des personnes qui ont répondu à l'enquête afin de compenser celles qui n'y ont pas répondu.

8.5 Non-réponse partielle

Il y a non-réponse partielle lorsque le répondant a refusé de répondre à une question ou lorsqu'il ne pouvait pas se souvenir de l'information demandée. Des codes dans le fichier de microdonnées indiquent les non-réponses partielles.

8.6 Couverture

Tel que mentionné dans le chapitre 5.1 (Champ de l'enquête), moins 3 % des ménages au Canada n'ont pas le téléphone. Les personnes qui vivent dans ces ménages ont peut-être des caractéristiques uniques qui ne seront pas reflétées dans les estimations de l'enquête. Les utilisateurs devraient faire preuve de prudence lorsqu'ils analysent des sous-groupes de la population

dont les caractéristiques peuvent être corrélées au fait de ne pas avoir le téléphone.

8.7 Mesure de l'erreur d'échantillonnage

Comme il est inévitable que les estimations d'une enquête par sondage comportent une erreur d'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur d'échantillonnage. La mesure de l'importance éventuelle des erreurs d'échantillonnage est fondée sur l'erreur-type des estimations découlant des résultats de l'enquête. Cependant, en raison de la grande diversité des estimations que l'on peut tirer d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure qui en résulte, appelée coefficient de variation (c.v.) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons qu'en se basant sur les résultats de l'enquête, on estime que 31 % des Canadiens fument actuellement la cigarette, et qu'on constate que cette estimation a une erreur type de 0,0056. Le coefficient de variation de l'estimation serait calculé ainsi :

$$\left(\frac{.0056}{.31}\right) x\ 100\% \ '\ 1{,}8\%$$

9.0 Lignes directrices pour la publication et la diffusion

Ce chapitre décrit les lignes directrices que doivent suivre les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données tirées du fichier de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs d'arriver aux mêmes chiffres que Statistique Canada, tout en étant en mesure d'obtenir, conformément à ces lignes directrices, des chiffres qui n'ont pas été publiés.

9.1 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées de ces fichiers de microdonnées correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices ci-après en ce qui concerne l'arrondissement de ces estimations :

a) Les estimations dans le corps d'un tableau statistique doivent être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Dans cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 à 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.

b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.

c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et (ou) les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on cherche à arrondir à un seul chiffre par cette

technique, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.

d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.

e) Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique d'arrondissement que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

f) Des estimations non arrondies ne doivent en aucun cas être publiées ou diffusées par les utilisateurs. Des estimations qui ne sont pas arrondies laissent supposer qu'elles sont beaucoup plus précises qu'elles ne le sont en réalité.

9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête de surveillance de l’usage du tabac au Canada n'était pas autopondéré. Pour produire des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer le poids approprié.

Sinon, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population observée et ne correspondront pas à celles de Statistique Canada.

Les utilisateurs doivent également noter qu'en raison de la manière dont ils traitent le champ du poids, certains progiciels ne permettent peut être pas d'obtenir des estimations correspondant exactement à celles de Statistique Canada.

9.2.1 Définitions des genres d'estimations: de type nominal par opposition à quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ESUTC, il est utile de décrire les deux principaux genres d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir des fichiers de microdonnées pour l'Enquête de surveillance de l'usage du tabac au Canada.

Estimations de type nominal

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes dans la population visée par l'enquête qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Le nombre de fumeurs actuels de cigarettes et la proportion de fumeurs quotidiens qui ont tenté de cesser de fumer sont des exemples de type nominal. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être appelée estimation d'un agrégat.

Exemples de questions de type nominal

Q Avez-vous fumé des cigarettes au cours des 30 DERNIERS JOURS?
R Oui/Non

Q Qu'est-ce qui vous a incité à cesser de fumer?
R Problèmes de santé / maladie liée à la cigarette d'un ami/ Grossesse / Avis d'un médecin / inquiétude au sujet de sa santé / Maladie / Accident

Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités basées sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi explicitement des estimations de la forme $\hat{X}/\hat{Y}$ où $\hat{X}$ est une estimation de la quantité totale pour la population visée par l'enquête et $\hat{Y}$ est une estimation du nombre de personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de cigarette fumés les fins de semaine. Si l'on écrit cette moyenne sous la forme X/Y, le numérateur (X) est une estimation du nombre total de fumées les samedi et dimanche alors que le dénominateur (Y) est une estimation du nombre de personnes ayant contribué à cette quantité.

Exemples de questions quantitatives

Q Au cours des 7 derniers jours, à compter d'hier, combien de cigarettes avez-vous fumé lundi?
R |_|_| cigarettes

Q À quel âge avez-vous fumé votre première cigarette?
R |_|_| ans

9.2.2 Totalisation d'estimations de type nominal

On peut obtenir des estimations du nombre de personnes possédant une caractéristique particulière à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient les proportions et les rapports en:

(a) additionnant les poids finals des enregistrements possédant la caractéristique qui nous intéresse, pour le numérateur(X),
(b) en additionnant les poids finals des enregistrements possédant la caractéristique qui nous intéresse, pour le dénominateur(Y), puis
(c) en divisant l'estimation du numérateur par celle du dénominateur.

9.2.3 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Par exemple, pour obtenir une estimation du nombre total de cigarettes fumées le samedi, il faut multiplier la valeur déclarée à Q090SAT (nombre de cigarettes fumées le samedi) par le poids final de l'enregistrement, puis additionner cette valeur pour tous les enregistrements portant la mention Q090sat <96 (tous les répondants qui ont donné une réponse à ce champ).

Afin d'obtenir une moyenne pondérée exprimée sous la forme X/Y, le numérateur (X) est calculé comme une estimation quantitative et le dénominateur (Y), comme une estimation de type nominal. Par exemple, pour estimer le nombre moyen de cigarettes fumées le samedi,

(a) estimer le nombre total de cigarettes fumées le samedi, comme on l'a vu ci-dessus;

(b) estimer le nombre de personnes dans cette catégorie par l'addition des poids finals de tous les enregistrements ayant la mention Q090SAT < 96; puis
(c) diviser l'estimation obtenue en «a» par celle calculée en «b».

9.3 Lignes directrices pour l'analyse statistique

L'Enquête de surveillance de l'usage du tabac au Canada est fondée 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 provenant d'enquêtes aussi complexes présente des difficultés aux analystes parce que le plan d'enquête et les probabilités de sélection influent sur les méthodes d'estimation et de calcul de la variance qui doivent être utilisées. Pour que les estimations et les analyses de l'enquête soient sans biais, les poids de l'enquête doivent être utilisés.

Bien que de nombreuses méthodes d'analyse qui font partie des progiciels statistiques permettent d'utiliser des poids, le poids diffère sous l'angle de la définition ou du sens de celui qui convient à une enquête par sondage, de sorte que si les estimations faites au moyen de ces progiciels sont exactes dans bien des cas, les variances calculées n'ont pratiquement aucune signification. Les variances pour des estimations simples comme totaux, proportions et rapport (pour variables qualitatives) sont fournis dans les tableaux de variabilité de l'échantillon.

Dans le cas de nombreuses techniques d'analyse (par exemple, la régression linéaire, la régression logistique, l'analyse de variance), il existe un moyen de rendre l'application des progiciels standard plus significative, en incorporant les probabilités inégales de sélection . La méthode transforme les poids des enregistrements de manière à ce que le poids moyen soit un (1).

Par exemple, si nous voulions faire une analyse pour les répondants masculins. Les étapes pour transformer les poids seraient:

- sélectionner tous les répondants du fichier qui ont répondu SEX=hommme
- Calculer le poids MOYEN pour ces enregistrements en totalisant les poids originaux du fichier de microdonnées pour ces enregistrements et le divisant par le nombre de répondants ayant répondu SEX=homme
- pour chacun de ces répondants, calculer un poids RECALCULÉ égal au poids original de ce répondant divisé par le poids MOYEN
- effectuer l'analyse de ces répondants en utilisant le poids RECALCULÉ.

Cependant, la stratification et le groupement du plan de sondage ne sont pas considéré dans cette méthode et l'estimation de la variance calculé de cette façon est probablement sous-évaluée.

Le calcul d'estimations de variance vraiment significatives requiert une connaissance approfondie du plan de l'enquête. Ces détails ne peuvent être fournit dans ce fichier de microdonnées à cause d e la confidentialité. Les variances qui prend en compte le plan de l'enquête peuvent être calculé pour plusieurs données par Statistique Canada sur une base de coûts recouvrables.

9.4 Lignes directrices pour la diffusion en fonction de la variabilité d'échantillonnage

Avant de diffuser et (ou) de publier des estimations tirées de ce fichier de microdonnées, les utilisateurs doivent commencer par établir le niveau de qualité de l'estimation. Les niveaux de qualité sont *acceptable, médiocre et inacceptable*. La qualité des données est affectée par des erreurs d'échantillonnage et des erreurs d'observation tel que discuté au chapitre 8. Cependant, le niveau de qualité d'une estimation est déterminé seulement en fonction de l'erreur d'échantillonnage refléter par son coefficient de variation comme indiqué au tableau suivant. Les utilisateurs devraient lire le chapitre 8 pour mieux comprendre les caractéristiques relatives à la qualité de ces données.

Il faut commencer par établir le nombre de répondants qui ont fourni les données entrant dans le calcul de l'estimation. Si ce nombre est inférieur à 30, l'estimation pondérée doit être considérée de qualité inacceptable.

Pour les estimations pondérées basées sur des échantillons composés de 30 enquêtés ou plus, les utilisateurs doivent calculer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité ci-dessous. Elles sont appliquées aux estimations pondérées arrondies.

Tous les estimations peuvent faire l'objet de diffusion. Toutefois, celles de qualité de médiocre ou inacceptable devraient être accompagnées d'un avertissement mettant en garde les prochains utilisateurs.

Lignes directrices relatives à la variabilité d'échantillonnage

Niveau de qualité	Lignes directrices
1. Acceptable	Caractéristiques des estimations: un échantillon de 30 ou plus, et faibles coefficients de variation compris entre 0.0% - 16.5% Aucune mise en garde nécessaire.
2. Médiocre	Caractéristiques des estimations: un échantillon de 30 ou plus, et des coefficients de variation élevés compris entre 16.6% - 33.3%. On doit désigner ces estimations en utilisant la lettre M (ou un autre identificateur semblable). Elles doivent faire l'objet d'un avertissement mettant les prochains utilisateurs en garde contre les hauts taux d'erreur associés aux estimations.
3. Inacceptable	Caractéristiques des estimations: un échantillon inférieur à 30, ou des coefficients de variation très élevés, supérieurs à 33.3%. Statistique Canada recommande de ne pas diffuser d'estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit alors désigner les estimations en utilisant la lettre I (ou un autre identificateur semblable et les diffuser avec l'avertissement suivant: "Nous informons l'utilisateur que ....(précisez les données) ... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de publier cet avertissement.

10.0 Tables de variabilité d'échantillonnage approximative

Tout d'abord, afin de permettre aux utilisateurs d'avoir facilement accès à des coefficients de variation qui s'appliqueraient à une gamme étendue d'estimations de type nominal réalisées à partir de cet fichier de microdonnées, on a préparé un ensemble de «tables à consulter», appelées Tables de variabilité d'échantillonnage approximative. Ces tables permet aux utilisateurs d'obtenir un coefficient de variation basé sur la taille de l'estimation calculé à partir des données de l'enquête.

Les coefficients de variation (c.v.) dans ces tables sont obtenus au moyen de la formule de la variance utilisée pour l'échantillonnage aléatoire simple en y incorporant un facteur reflétant la structure en grappes à plusieurs degrés du plan d'échantillonnage. Pour obtenir ce facteur, appelé effet du plan, on a commencé par calculer les effets du plan pour une gamme étendue de caractéristiques, puis on a choisi parmi les nombres ainsi obtenus une valeur modérée à employer dans les tables à consulter qui serait ensuite appliquée à l'ensemble des caractéristiques.

Le tableau suivant indique l'effet de plan, la taille de l'échantillon et la population par province qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative pour le fichier "Ménages".

Province	Effet de plan	Taille d'échantillon	Population
Terre-Neuve	1.08	2,050	194,167
Île-du-Prince-Édouard	1.12	2,303	51,965
Nouvelle-Écosse	1.16	2,436	362,821
Nouveau-Brunswick	1.18	1,972	285,052
Québec	1.15	2,315	3,016,518
Ontario	1.10	2,099	4,322,459
Manitoba	1.14	2,327	427,501
Saskatchewan	1.15	2,100	385,715
Alberta	1.06	2,307	1,098,250
Colombie-Britannique	1.18	2,526	1,563,839
Canada	2.54	22,435	11,708,286

Voici la table des effets du plan, des tailles de l'échantillon et des chiffres de population par province employés pour produire les tables de variabilité d'échantillonnage approximative pour le fichier “Personnes”.

Province	Groupe d’âge	Effet du plan	Taille échantillon	Population
Terre Neuve	Tous	1.95	1096	441202
	15-19	1.72	354	41226
	20-24	1.73	195	39513
	25+	1.37	547	360463
Île-du-Prince-Edouard	Tous	1.86	1159	109524
	15-19	1.40	323	10236
	20-24	1.79	258	9595
	25+	1.47	578	89693
Nouvelle-Écosse	Tous	2.20	1134	760694
	15-19	1.60	307	63736
	20-24	1.65	259	62959
	25+	1.55	568	633999
Nouveau-Brunswick	Tous	1.87	975	610747
	15-19	1.47	261	51338
	20-24	1.63	216	51597
	25+	1.40	498	507812
Québec	Tous	1.95	1090	5966329
	15-19	1.49	293	478366
	25-64	1.65	262	502436
	25+	1.39	535	4985527
Ontario	Tous	2.00	947	9293460
	15-19	1.42	252	764939
	20-24	1.44	249	757572
	25+	1.31	446	7770948
Manitoba	Tous	1.84	978	894006
	15-19	1.49	230	80342

	20-24	1.52	225	78108
	25+	1.46	523	735556
Saskatchewan	Tous	1.77	963	790190
	15-19	1.41	262	78879
	20-24	1.50	189	73848
	25+	1.36	512	637453
Alberta	Tous	1.88	1073	2342150
	15-19	1.56	271	221450
	20-24	2.09	246	222033
	25+	1.40	556	1898666
Colombie-Britannique	Tous	1.97	1078	3287502
	15-19	1.59	328	270075
	20-24	2.33	218	267192
	25+	1.31	532	2750235
Canada	Tous	4.93	10493	24495804
	15-19	3.67	2881	2060588
	20-24	2.98	2317	2064864
	25+	3.55	5295	20370352

Il convient de mentionner que tous les coefficients de variation sont approximatifs dans les tables de variabilité d'échantillonnage approximative et ne doivent donc pas être considérés comme des valeurs officielles. On peut se procurer, contre recouvrement des coûts, des estimations de la variance réelle de variables particulières auprès de Statistique Canada. L'utilisation d'estimations de la variance réelle peut permettre aux utilisateurs de diffuser des estimations que l'on ne pourrait diffuser autrement, c'est-à-dire des estimations avec des coefficients de variation classés «inacceptables».

N'oubliez pas : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est considéré comme de qualité inacceptable et devrait être accompagnée d'une mise en garde quelle que soit la valeur de son coefficient de variation. Ceci est dû au fait que les formules utilisées pour obtenir une estimation de la variance ne donnent pas de bons résultats pour de petits échantillons.

10.1 Comment utiliser les tables de c.v. pour les estimations de type nominal

Les règles suivantes doivent permettre à l'utilisateur de calculer à partir des tables de variabilité d'échantillonnage les coefficients de variation approximatifs d'estimations relatives au nombre, à la proportion ou au pourcentage de personnes dans la population enquêtée possédant une caractéristique donnée ainsi que des rapports et des écarts entre ces estimations.

Règle 1. Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Sur la table de variabilité d'échantillonnage correspondant à la région géographique appropriée, repérez le nombre estimé dans la colonne à l'extrême gauche de la table (colonne intitulée «Numérateur du pourcentage») et suivez les astérisques (le cas échéant) de gauche à droite jusqu'au premier nombre. Ce nombre constitue le coefficient de variation approximatif.

Règle 2. Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion ou d'un pourcentage estimé dépend à la fois de la grandeur de cette proportion ou de ce pourcentage et de la grandeur du total sur lequel est fondé cette proportion ou ce pourcentage. Les proportions ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque cette proportion ou ce pourcentage est fondé sur un sous-ensemble de la population. Par exemple, la proportion “d’anciens fumeurs qui ont cessé pour des problèmes de santé” est une mesure plus fiable que le nombre estimatif total d’anciens fumeurs qui ont cessé pour des problèmes de santé. (Il faut noter que dans les tables, la valeur des coefficients de variation décroît de gauche à droite sur une même ligne.)

Lorsque la proportion ou le pourcentage est fondé sur la population totale de la région géographique à laquelle la table s'applique, le coefficient de variation de la proportion ou du pourcentage est égal à celui du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est fondé sur un sous-ensemble de la population totale (par exemple, les personnes qui appartiennent à un groupe d'âge ou un sexe particulier), il faut se reporter à la proportion ou au pourcentage (en haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). Le coefficient de variation se trouve à l'intersection de la ligne et de la colonne appropriées.

Règle 3. Estimations des différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est à peu près égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. L'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est donc :

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{X}_1$ représente l'estimation 1, $\hat{X}_2$ l'estimation 2, et a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement. Le coefficient de variation de $\hat{d}$ est donné par $s_{\hat{d}}/\hat{d}$. Cette formule donne un résultat exact en ce qui a trait à la différence entre des caractéristiques distinctes et non corrélées, mais elle ne donne que des résultats approximatifs dans les autres cas.

Règle 4. Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, il faut convertir le rapport en pourcentage et appliquer la règle 2. Cette situation s'appliquerait notamment au cas où le dénominateur est le nombre «de fumeur» et le numérateur, le nombre de «fumeurs quotidiens».

Si le numérateur n'est pas un sous-ensemble du dénominateur, (par exemple, le rapport du nombre de «fumeurs quotidiens» au nombre de «non-fumeurs»), l'écart-type du rapport entre les estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation pris séparément multipliée par R. L'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est donc :

$$s_{\hat{R}} = \hat{R}\sqrt{a_1^2 + a_2^2}$$

où a_1 et a_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 $s_{\hat{R}}/\hat{R}$. La formule tend à surestimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélées positivement et à sous-estimer l'erreur si $\hat{X}_1$ et $\hat{X}_2$ sont corrélées négativement.

Règle 5. Estimations des différences entre des rapports

Dans ce cas, on combine les règles 3 et 4. On commence par calculer les coefficients de variation des deux rapports au moyen de la règle 4, puis le coefficient de variation de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de c.v. pour des estimations de type nominal

Les exemples «concrets» ci-dessous sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

Exemple 1. Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime que, durant la période de référence, à 5,952,498 le nombre de personnes qui sont des fumeurs actuels (DVSST1 = ‘1') au Canada. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

(1) Reportez-vous à la table de c.v. pour le CANADA.

(2) L'agrégat estimé (5,952,498) ne se trouve pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»), il faut donc utiliser le nombre qui s'en rapproche le plus, c'est-à-dire 6,000,000.

(3) Le coefficient de variation pour un agrégat estimé est la première inscription, autre que des astérisques, sur cette ligne, c'est-à-dire 3.8%.

(4) Le coefficient de variation approximatif de l'estimation est donc 3.8 %.

Il n'y a donc pas lieu de faire une mise en garde lorsque l'on publie le résultat, soit qu'il y a eu 5,952,498 fumeurs durant la période de référence.

Exemple 2. Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime à 3,192,069/12,051,027=26.5 % le pourcentage d’hommes qui fumaient au Canada durant la période de référence. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

(1) Reportez-vous à la table de c.v. pour le CANADA.(Le tableau CANADA est utilisé parce qu'il est le plus petit tableau qui contient le domaine de l'estimation, tous les hommes au Canada).

(2) Parce que l'estimation est un pourcentage basé sur un sous-ensemble de la population totale (c.-à-d., les hommes), il faut utiliser à la fois le pourcentage (26.5%) et la partie numérateur du pourcentage (3,192,069) pour déterminer le coefficient de variation.

(3) Le numérateur, 3,192,069 ne figure pas dans la colonne de gauche (la colonne «Numérateur du pourcentage»), il faut donc utiliser le nombre qui s'en rapproche le plus, soit 3,000,000. De même, l'estimation du pourcentage ne figure pas parmi les en-têtes de colonnes, il faut donc utiliser le nombre qui s'en rapproche le plus, 25.0 %.

(4) Le nombre qui se trouve à l'intersection de la ligne et de la colonne utilisées, soit 5.4 %, est le coefficient de variation à employer.

(5) Le coefficient de variation approximatif de l'estimation est donc 5.4 %.

Il n'y a donc pas lieu de faire une mise en garde lorsque l'on publie le résultat, soit que 26.5 % des hommes fument actuellement.

Exemple 3. Estimations des différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime à 2,760,429/12,444,778=22.2 % le pourcentage de femmes qui fument actuellement et à 3,192,069/12,051,027=26.5 % le pourcentage d'hommes qui fument au Canada. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

(1) À l'aide de la table de c.v. pour le CANADA utilisée de la même façon que dans l'exemple 2, vous établissez à 5.5 % le c.v. de l'estimation pour les femmes et à 5.4 % le c.v. de l'estimation pour les hommes.

(2) Selon la règle 3, l'erreur-type pour une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

où $\hat{X}_1$ est l'estimation 1 (hommes), $\hat{X}_2$ est l'estimation 2 (femmes) et a_1 et a_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d}$ = (.265-.222) = .043 est :

$$s_{\hat{d}} = \sqrt{[(.265)(.054)]^2 + [(.222)(.055)]^2}$$

$$= \sqrt{(.000023) + (.000015)}$$

$$= .019$$

(3) Le coefficient de variation de $\hat{d}$ est donné par $s_{\hat{d}}/\hat{d}$ = .019/.043 = 0.442

(4) Le coefficient de variation approximatif de la différence entre les estimations est donc 44.2 %. Cette estimation est considérée de qualité inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être accompagnée d'un avertissement quant à la qualité.

Exemple 4. Estimations de rapports

Supposons qu'un utilisateur estime à 278,594 le nombre de fumeurs actuels de sexe féminin agés de 15 à 19 ans et à 254,751 le nombre de fumeurs actuels de sexe masculin agés de 15 à 19 ans . L'utilisateur est intéressé à comparer l'estimation des femmes à celle des hommes sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

(1) Tout d'abord, cette estimation est une estimation de rapport, où le numérateur de l'estimation (= $\hat{X}_1$) est le nombre de fumeurs actuels de sexe féminin agés de 15 à 19 ans . Le dénominateur de l'estimation (= X_2) est le nombre de fumeurs actuels de sexe masculin agés de 15 à 19 ans .

(2) Reportez-vous à la table de c.v. pour le CANADA 15-19.

(3) Le numérateur de cette estimation de rapport est 278,594. Le chiffre qui se rapproche le plus de ce nombre est 300,000. On détermine le coefficient de variation pour cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit, 8.6 %.

(4) Le dénominateur de cette estimation de rapport est 254,751. Le chiffre qui se rapproche le plus de ce nombre est 250,000. On détermine le coefficient de variation de cette estimation en trouvant la première inscription, autre que des astérisques, sur cette ligne, soit, 9.4 %.

(5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est,

$$a_{\hat{R}} = \sqrt{a_1^2 + a_2^2}$$

où a_1 et a_2 sont les coefficients de variation de $\bar{x}_1$ et $\bar{x}_2$ respectivement.

C'est-à-dire,

$$a_{\hat{R}} = \sqrt{(.086)^2 + (.094)^2}$$

$$= 0.127$$

Le rapport femmes:hommes fumeurs agés de 15 à 19 ans est égal à 278,594/254,751, soit 1.09. Le coefficient de variation de cette estimation est égal à 12.7%, l'estimation peut être publiée sans mise en garde.

10.2 Comment utiliser les tables de c.v. pour calculer des limites de confiance

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est une façon d'énoncer la probabilité que la valeur vraie de la population se situe dans une plage de valeurs données. Par exemple, un intervalle de confiance à 95 % peut être décrit comme suit :

Si l'échantillonnage de la population se répète à l'infini, chacun des échantillons donnant un nouvel intervalle de confiance pour une estimation alors, dans 95 % des cas, l'intervalle contiendra la valeur vraie de la population.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'en procédant à un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre

une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance pour une estimation, $\hat{X}$, sont généralement exprimés sous forme de deux nombres, l'un étant inférieur à l'estimation et l'autre supérieur à celle-ci, sous la forme ($\hat{X}$-k, $\hat{X}$+k), où k varie selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance pour une estimation à partir des tables de variabilité d'échantillonnage aPproximative, en commençant par trouver, dans la table appropriée, le coefficient de variation de l'estimation $\hat{X}$, puis en utilisant la formule ci-dessous pour obtenir l'intervalle de confiance IC correspondant :

$$IC_{\hat{X}} = \{\hat{X} - (t)(\hat{X})(a_{\hat{X}}), \hat{X} + (t)(\hat{X})(a_{\hat{X}})\}$$

où $a_{\hat{X}}$ est le coefficient de variation trouvé pour $\hat{X}$

t = 1 si l'on désire un intervalle de confiance à 68 %
t = 1.6 si l'on désire un intervalle de confiance à 90 %
t = 2 si l'on désire un intervalle de confiance à 95 %
t = 3 si l'on désire un intervalle de confiance à 99 %.

NOTA : Les lignes directrices qui s'appliquent à la diffusion des estimations s'appliquent aussi aux intervalles de confiance. Par exemple, si l'estimation est de qualité inacceptable, alors l'intervalle de confiance est aussi de qualité inacceptable

10.2.1 Exemple d'utilisation de tables de c.v. pour obtenir des limites de confiance

Voici comment on calculerait un intervalle de confiance à 95 % pour la proportion estimée de fumeurs actuels qui sont des hommes (d'après l'exemple 2):

$\hat{X}$ = 26.5% (ou, sous forme de proportion, .265)

t = 2

$a_{\hat{X}}$ = 5.4% (.054 sous forme de proportion) est le coefficient de variation de cette estimation tel que déterminé à partir des tables.

$IC_{\hat{x}}$ = (.265 - (2) (.265) (.054), .265 + (2) (.265 (.054)

$IC_{\hat{x}}$ = (.265- .029, 265 + .029

IC_x = (.236, .294)

On peut dire, avec une probabilité de 95 %, qu'entre 23.6 % et 29.4 % des hommes fument actuellement.

10.3 Comment utiliser les tables de c.v. pour effectuer un test t

On peut aussi utiliser les erreurs-types pour effectuer des tests d'hypothèses, une technique qui permet de faire la distinction entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des rapports, etc. Les test peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Soit $\mathbf{X_1}$ et $\mathbf{X_2}$ des estimations basées sur un échantillon pour deux caractéristiques qui nous intéressent. Supposons que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $s_{\hat{d}}$

$$\text{Si } t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}}$$

est compris entre -2 et 2, alors on ne peut tirer de conclusion à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0.05. C'est-à-dire que les caractéristiques sont significatives.

10.3.1 Exemple d'utilisation des tables de c.v. pour effectuer un test t

Supposons que nous désirons tester, au niveau de signification de 5 %, l'hypothèse qu'il y a une différence entre la proportion de fumeurs de sexe masculin et la proportion de fumeurs de sexe féminin. Dans l'exemple 3, nous avons trouvé que l'erreur-type de la différence entre ces deux estimations était égal à 0.043. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}} = \frac{.265 - .222}{.019} = \frac{.043}{.019} = 2.26$$

Puisque t = 2.26 et est plus grand que 2, on doit conclure qu'il existe une différence significative au niveau de 0.05 entre les deux estimations.

10.4 Coefficients de variation d'estimations quantitatives

Pour calculer l'erreur d'échantillonnage d'estimations quantitatives, il faudrait élaborer des tables spéciales, ce qui n'a pas été fait puisque la majorité des variables de l'Enquête de surveillance sur l'usage du tabac au Canada sont surtout de type nominal.

Néanmoins, le coefficient de variation d'un total quantitatif est généralement supérieur à celui de l'estimation de type nominal correspondante (c'est-à-dire l'estimation du nombre de personnes contribuant à l'estimation quantitative). Si l'estimation de type nominal correspondante est de qualité inacceptable, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de cigarettes fumées le lundi serait plus élevé que le coefficient de variation de la proportion de fumeurs actuels. Par conséquent, si le coefficient de variation de la proportion est de qualité inacceptable, ce sera aussi le cas pour le coefficient de variation de l'estimation quantitative correspondante.

On peut calculer, au besoin, les coefficients de variation d'estimations de ce genre pour une estimation particulière au moyen d'une technique appelée pseudo-réplication, qui consiste à diviser les enregistrements des fichiers de microdonnées en sous-groupes (ou sous-échantillons) et à calculer la variabilité de l'estimation d'un sous-échantillon à l'autre. Les utilisateurs désireux de calculer le coefficient de variation d'estimations quantitatives peuvent demander conseil à Statistique Canada en ce qui concerne la manière de répartir les enregistrements en sous-échantillons appropriés et les formules à utiliser pour ces calculs.

10.5 Seuils pour la diffusion des estimations relatives à l'ESUTC - fichier ménages

La taille minimum des estimations au niveau provincialest indiquée dans le tableau ci-dessous. Les estimations plus petites que la taille indiquée dans la colonne “Inacceptable” doivent être publiées avec une mise en garde tel que décrit dans le tableau des qualités des estimations.

Province	Acceptable CV < 16.15	Médiocre CV 16.5 - 33.0	Inacceptable CV > 33.0
Terre-Neuve	3,500 & +	1,000-3,500	moins de 1,000
Île-du-Prince-Édouard	1,000 & +	0-1,000	moins de 0
Nouvelle-Écosse	6,000 & +	1,500-6,000	moins de 1,500
Nouveau-Brunswick	6,000 & +	1,500-6,000	moins de 1,500
Québec	54,000 & +	13,500-54,000	moins de 13,500
Ontario	81,500 & +	20,500-81,500	moins de 20,500
Manitoba	7,500 & +	2,000-7,500	moins de 2,000
Saskatchewan	7,500 & +	2,000-7,500	moins de 2,000
Alberta	18,000 & +	4,500-18,500	moins de 4,500
Colombie-Britannique	26,500 & +	6,500-26,500	moins de 6,500
Canada	48,500 & +	12,000-48,500	moins de 12,000

10.6 Seuils pour la diffusion des estimations relatives à l'ESUTC - fichier personnes

Le tableau suivant spécifie la taille minimum des estimations par province et groupe d'âge. Les estimations plus petite que la taille minimum présentées dans la colonne "Inacceptable" doivent être accompagnées d'une mise en garde appropriée.

Tableau de limites pour diffusion- fichier personnes

Province	Groupe d'âge	Acceptable cv < 16.15	Médiocre cv entre 16.5-33.0	Inacceptable cv >33.0
Terre-Neuve	Tous	27,000 & +	7,000-27,000	moins de 7,000
	15-19	6,000 & +	1,500-6,000	moins de 1,500
	20-24	9500 & +	3,000-9,500	moins de 3,000
	25+	30,500 & +	8,000-30,500	moins de 8,000

Île-du-Prince-Édouard	Tous	6,000 & +	1,500-6,000	moins de 1,500
	15-19	1,500 & +	500-1,500	moins de 500
	20-24	2,000 & +	500-2,000	moins de 500
	25+	7,500 & +	2,000-7,500	moins de 2,000
Nouvelle-Écosse	Tous	50,500 & +	13,000-50,500	moins de 13,000
	15-19	10,000 & +	3,000-10,000	moins de 3,000
	20-24	12,000 & +	3,500-12,000	moins de 3,500
	25+	58,000 & +	15,000-58,000	moins de 15,000
Nouveau-Brunswick	Tous	40,000 & +	10,500-40,000	moins de 10,500
	15-19	9,000 & +	2,500-9,000	moins de 2,500
	20-24	11,000 & +	3,500-11,000	moins de 3,500
	25+	47,500 & +	12,500-47,500	moins de 12,500
Québec	Tous	368,000 & +	94,500-368,000	moins de 94,500
	15-19	75,500 & +	21,000-75,500	moins de 21,000
	20-24	94,500 & +	27,000-94,500	moins de 27,000
	25+	434,500 & +	114,000-434,500	moins de 114,000
Ontario	Tous	669,000 & +	173,500-669,000	moins de 173,500
	15-19	131,000 & +	37,000-131,000	moins de 37,000
	20-24	132,500 & +	38,500-132,500	moins de 37,500
	25+	756,500 & +	200,000-756,500	moins de 200,500
Manitoba	Tous	58,000 & +	15,000-58,000	moins de 15,000
	15-19	15,500 & +	4,500-15,500	moins de 4,500
	20-24	15,500 & +	4,500-15,500	moins de 4,500
	25+	68,500 & +	18,000-68,500	moins de 18,000
Saskatchewan	Tous	50,000 & +	13,000-50,000	moins de 13,000
	15-19	13,000 & +	3,500-13,000	moins de 3,500
	20-24	16,500 & +	5,000-16,500	moins de 5,000
	25+	56,500 & +	15,000-56,500	moins de 15,000
Alberta	Tous	141,500 & +	36,500-114,500	moins de 36,500
	15-19	38,500 & +	11,000-38,500	moins de 11,000
	20-24	53,000 & +	16,000-53,000	moins de 16,000
	25+	160,500 & +	42,000-160,500	moins de 42,000

Colombie-Britannique	Tous	207,000 & +	53,500-207,000	moins de 53,500
	15-19	41,000 & +	11,500-41,000	moins de 11,500
	20-24	75,500 & +	23,500-75,500	moins de 23,500
	25+	228,000 & +	59,500-228,000	moins de 59,500
Canada	Tous	415,500 & +	103,500-415,500	moins de 103,500
	15-19	92,000 & +	23,500-92,000	moins de 23,500
	20-24	93,000 & +	23,500-93,000	moins de 23,500
	25+	489,500 & +	122,500-489,500	moins de 122,500

10.7 Tables de c.v. - fichier ménage

Voir le fichier 00_C1_CVMN_FRA.PDF pour les tables de cv pour le fichier ménage du Cycle 1.

10.8 Tables de c.v. - fichier Personnes

Voir le fichier 00_C1_CVPR_FRA.PDF pour les tables de cv pour le fichier personne du Cycle 1.

11.0 Pondération

Pour le fichier de microdonnées, des poids statistiques ont été placés pour chaque enregistrement pour représenter le nombre de personnes que chaque enregistrement échantilloné représente. Un poids a été calculé pour chaque ménage et sur un fichier différent un poids pour chaque personne.

La pondération pour le premier cycle de l'Enquête de surveillance sur l'usage du tabac au Canada comprend plusieurs étapes: calcul du poids de base, le facteur de compensation pour la non-réponse, un ajustement pour la sélection d'une ou deux personnes dans le ménage, l'élimination des enregistrements hors plan et finalement un ajustement pour rendre les estimations de la population consistantes avec les totaux connus Province-âge-sexe de la population estimée du recensement pour les personnes de 15 ans et plus.

11.1 Pondération pour les fichiers Ménage et Personne

1. Calcul du poids pour le numéro de téléphone

Chaque numéro de téléphone dans l'échantillon a reçu un poids de base, W1, égal à l'inverse de sa probabilité de sélection.

$$W1 = \frac{\left(\begin{array}{l}\text{nombre total de numéros de téléphone pouvant etre échantillonnés pour}\\ \text{Province - Strate}\end{array}\right)}{\left(\begin{array}{l}\text{nombre de numéros de téléphone échantillonés}\\ \text{pour Province - Strate}\end{array}\right)}$$

Il y avait 58,396 numéros de téléphone dans l'échantillon avec des poids.

2. Ajustement pour les numéros de téléphone non-résolus

Il y a eu 923 numéros de téléphone qui n'ont pas été résolus, laissant 57,473 numéros résolus. Nous n'avons pas pu déterminé si ces numéros non-résolus appartenaient à un ménage, une entreprise ou étaient hors service. Chaque numéro avait un indicateur si on s'attendait que le numéro serait une résidence, une entreprise ou de genre inconnu. L'ajustement pour les numéros non-résolus a été fait par province-strate et le genre de numéro.

Pour chaque genre de ligne-province-strate,

$$W2 = W1 * \left(\frac{\sum W1 \text{ pour numéros résolus } + \sum W1 \text{ pour numéros non-résolus}}{\sum W1 \text{ pour numéros résolus}} \right)$$

3. Éliminer les numéros de téléphone hors du champs de l'enquête

Les numéros de téléphone correspondants à une entreprise, hors service ou hors du champs de l'enquête, comme les chalets, ont été éliminés après l'ajustement pour la non-réponse du téléphone. Veuillez prendre note que si les données pour le **ménage ou pour une personne** étaient présentes le numéros de téléphone était considéré comme étant celui d'un ménage. 32,520 numéros de téléphone étaient hors du champs de l'enquête et 24,953 numéros appartenaient à un ménage.

4. Ajustement pour la non-réponse pour le nombre de lignes téléphoniques dans le ménage

Le nombre de lignes téléphoniques dans le ménage a été calculé. Si le nombre de lignes était inconnu et que des données existaient pour le **ménage ou pour une personne**, la valeur de 1 a été imputé pour conserver des données. Après l'imputation, il restait 1,661 numéros de téléphone pour lesquels le nombre de lignes manquait toujours. Donc, il y avait 23,292 ménages dont l'information a été calculé ou imputé. L'ajustement a été fait au niveau province-strate.

$$W3 = W2 * \left(\frac{\sum W2 \text{ pour ménages avec nombre de lignes } + \sum W2 \text{ pour ménages avec nombre de lignes manquant}}{\sum W2 \text{ pour ménages avec nombre de lignes}} \right)$$

5. Calcul du poids des ménages avec de multiples lignes téléphoniques

Les poids des ménages avec plus d'une ligne téléphonique ont été ajustés à la baisse pour prendre en compte le fait que ces ménages avaient une plus grande probabilité de sélection. Le poids pour chaque ménage a été divisé par le nombre de lignes téléphoniques du ménage.

$$W4 = \frac{W3}{\text{Nombre de lignes téléphoniques dans le champs d'enquête dans le ménage}}$$

11.2 Pondération pour le fichier Ménage

6. Ajustement pour les ménages non-répondants

Les répondants du ménage ont répondu aux questions sur leurs habitudes de fumer. Si ces questions n'ont pas été complétées, refusées ou remplies partiellement, le ménage a été considéré comme non-répondant. Il y a eu 857 non-répondants. Il reste donc 22,435 ménages pondérés par province-strate.

$$W5 = W4 * \left(\frac{\sum \text{W4 pour ménages répondants} + \sum \text{W4 pour ménages non-répondants}}{\sum \text{W4 pour ménages répondants}} \right)$$

7. Ajustement pour totaux externes connus par strate

On a calculé un ajustement aux poids des ménages pour chaque enregistrement par province, strate et mois pour s'assurer que les estimations du nombre de ménages étaient consistantes avec des totaux de ménage externes. Le facteur d'ajustement par province-strate-mois (P-S-M) est défini comme:

$$W6 = W5 * \left(\frac{\text{Totaux externes des ménages pour P-S-M}}{\text{Total des poids W5 pour les ménages répondants dans l'échantillon par P-S-M}} \right)$$

Le poids des ménages, W6, obtenu après cette étape, est final et apparaît sur le fichier de microdonnées du ménage.

11.3 Pondération pour le fichier Personne

6. Ajustement pour les ménages non-répondants

Sur le fichier Personne, les répondants du ménage devait répondre à la liste des membres sans aucun refus pour l'âge. Il y a eu 765 non-répondants. Il y a donc 22,523 poids pour les ménages dans le champs d'enquête qui ont été utilisés et ajustés selon province-strate.

$$W5 = W4 * \left(\frac{\sum \text{W4 pour ménages répondants} + \sum \text{W4 pour ménages non-répondants}}{\sum \text{W4 pour ménages répondants}} \right)$$

7. Calcul du poids de groupe

Tous les ménages répondants dans l'enquête ayant une liste complète des membres (aucun âge manquant) reçoivent un poids de groupe. À partir de la liste des membres, trois indicateurs sont placés pour indiquer la présence d'une personne dans les groupes d'âge 15-19, 20-24 et 25+. Si un ou deux groupes d'âge sont présents, un individu a été sélectionné dans chaque groupe d'âge présent (c.-à-d. la probabilité de sélection du groupe est 1). Donc, le poids n'est pas augmenté. Toutefois, si les trois groupes d'âge sont présents, deux personnes ont été sélectionnées et la probabilité le groupe d'âge est de 2 sur 3 groupe d'âge. Le poids est donc augmenté par son inverse.

Si 1 ou 2 groupes d'âge présent(s) , W6 = W5
Si tous les 3 groupes d'âge présents, W6 = W5 * 3/2

8. Éliminer les ménages pour lesquels personne n'a été sélectionné

Il y avait 12,571 ménages pour lesquels personne n'a été sélectionné pour compléter la portion sur l'usage du tabac ou la personne sélectionnée n'a pas été retenue à cause du sous-échantillonnage des individus. Ces ménages n'avaient donc aucune information au niveau de la personne. Approximativement 70% des personnes de 25 ans et plus n'ont pas été retenues. Il y avait 9,956 ménages pour lesquels nous avons sélectionné un répondant. Dans 8,484 ménages nous avons sélectionné une personne et 1,472 où nous avons sélectionné deux personnes.

9. Calcul du poids des ménages pour les personnes sélectionnées

Ces 8,484 + 2(1,472) = 11,428 personnes sélectionnées sont associées avec les ménages répondants et ont conservées le poids correspondant, W6.

10. Calcul du sous-poids de la personne sélectionnée

Tous les individus dans la population de l'enquête ont reçu un poids. Le poids est augmenté par le nombre de personnes dans le groupe d'âge et l'inverse du facteur de sous-échantillonnage.

$$W7 = W6 * \left(\frac{\text{Nombre de personnes dans le groupe d'âge choisi}}{\text{Facteur de sous-échantillonnage}} \right)$$

Le facteur de sous-échantillonnage est de 1 pour les groupes d'âge 15-19 et 20-24. Le facteur de sous-échantillonnage pour le 25 plus avait été déjà assigné et variait de 39.2% à 48.4% (selon la province).

11. Ajustement pour les individus non-répondants

Sur le fichier Personne, les répondants individuels ont donné l'information des questions sur l'usage du tabac et fourni une date de naissance correspondant à l'âge indiqué dans la liste des membres. Il y a eu 935 non-répondants. Donc, nous avons utilisé 10,493 poids des individus dans le champs d'enquête et ajusté selon la province, le groupe d'âge selon la liste des membres (15-19, 20-24, 25-44,45-64, 65+) et le sexe.

$$W8 = W7 * \left(\frac{\sum W7 \text{ pour personnes répondantes } + \sum W7 \text{ pour personnes non-répondantes}}{\sum W7 \text{ pour personnes répondantes}} \right)$$

12. Ajustement pour totaux externes

On a calculé un ajustement aux poids des personnes pour chaque enregistrement par province, strate et mois pour s'assurer que les estimations de la population étaient consistantes avec des totaux externes de population pour les personnes de 15 ans et plus. Ceci est connu comme la post-stratification. Les totaux externes suivants ont été utilisés:

1) Totaux mensuels de population par province - strate et mois
2) Totaux de population par province, sexe et les groupes d'âge suivants: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64 et 65 et plus. La moyenne de ces totaux a été faite pour la période d'enquête.

La méthode appelée GREG a été utilisé pour modifier les poids afin de s'assurer que les estimations de l'enquête concordaient avec les totaux externes simultanément pour les deux dimensions.

Après cette étape, les poids obtenus pour les personnes sont considérés comme finals et apparaissent au fichier de microdonnées.

12.0 Questionnaire

Voir le fichier 00_C1_QUES_FRA.pdf pour le questionnaire français du Cycle 1.

13.0 Cliché d’article et variables uniques

13.1 Cliché d’article et variables uniques - Ménages

Voir le fichier 00_C1_CLIMN_FRA.pdf pour le cliché d’articles et variables uniques pour le fichier ménage du Cycle 1.

13.2 Cliché d’article et variables uniques - Personne

Voir le fichier 00_C1_CLIPR_FRA.pdf pour le cliché d’articles et les variables uniques pour le fichier personne du Cycle 1.