Guide de l'utilisateur des microdonnées

ENQUÊTE DE SURVEILLANCE DE L'USAGE DU TABAC AU CANADA

CYCLE 1

FÉVRIER - JUIN 2003

Statistique Canada Statistics Canada

Canada

Table des matières

1.0 Introduction 5

2.0 Contexte 7

3.0 Objectifs 9

4.0 Concepts et définitions 11

5.0 Méthodologie de l'enquête 13
5.1 Champ de l'enquête 13
5.2 Stratification 13
5.3 Plan de sondage et répartition de l'échantillon 13
5.4 Tirage de l'échantillon 14

6.0 Collecte des données 17
6.1 Conception du questionnaire 17
6.2 Collecte et vérification des données 17

7.0 Traitement des données 19
7.1 Saisie des données 19
7.2 Vérification 19
7.3 Création de variables calculées 19
7.4 Pondération 19
7.5 Suppression des renseignements confidentiels 20

8.0 Qualité des données 21
8.1 Taux de réponse des ménages – février à juin 2003 22
8.2 Taux de réponse des personnes - février à juin 2003 23
8.3 Erreurs relatives à l'enquête 25
8.4 Non-réponse totale 25
8.5 Non-réponse partielle 25
8.6 Couverture 25
8.7 Mesure de l'erreur d'échantillonnage 26

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données 27
9.1 Lignes directrices pour l'arrondissement d'estimations 27
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation 28
9.3 Définitions de types d'estimations : catégoriques et quantitatives 28
9.3.1 Estimations catégoriques 28
9.3.2 Estimations quantitatives 28
9.3.3 Totalisation d'estimations catégoriques 29
9.3.4 Totalisation d'estimations quantitatives 29
9.4 Lignes directrices pour l'analyse statistique 30
9.5 Lignes directrices pour la diffusion de coefficients de variation 31
9.6 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des ménages 33
9.7 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des personnes 34

10.0 Tables de variabilité d'échantillonnage approximative..... 35
10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques..... 37
10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques..... 38
10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance..... 43
10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance..... 44
10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t..... 44
10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t..... 45
10.4 Coefficients de variation pour des estimations quantitatives..... 45
10.5 Tables des coefficients de variation - Fichier des ménages..... 46
10.6 Tables des coefficients de variation - Fichier des personnes..... 46

11.0 Pondération..... 47
11.1 Procédures de pondération pour les Fichiers des ménages et des personnes..... 47
11.2 Procédures de pondération pour le Fichier des ménages..... 48
11.3 Procédures de pondération pour le Fichier des personnes..... 49

12.0 Questionnaire..... 51

13.0 Clichés d'enregistrement à valeurs univariées..... 53
13.1 Cliché d'enregistrement à valeurs univariées – Fichier des ménages..... 53
13.2 Cliché d'enregistrement à valeurs univariées – Fichier des personnes..... 53

1.0 Introduction

L'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) a été menée par Statistique Canada entre février et juin 2003, avec la collaboration et l'appui de Santé Canada. Le présent manuel a été produit pour faciliter la manipulation du fichier de microdonnées sur les résultats de l'enquête.

Toutes questions au sujet de l'ensemble de données ou de son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle
Division des enquêtes spéciales
Téléphone : (613) 951-3321 ou appelez sans frais : 1 800 461-9050
Télécopieur : (613) 951-4527
Courriel : des@statcan.ca

Elizabeth Majewski
Division des enquêtes spéciales
2e étage, Immeuble Principal, pré Tunney
Ottawa (Ontario) K1A 0T6
Téléphone : (613) 951-4584
Télécopieur : (613) 951-0562
Courriel : elizabeth.majewski@statcan.ca

Santé Canada

Murray Kaiserman
Bureau de la recherche, de la surveillance et de l'évaluation
Programme de la lutte au tabagisme
Direction générale de la santé environnementale et de la sécurité des consommateurs
Immeuble MacDonald, ind. l'adr. 3507C
123, rue Slater, pièce A723
Ottawa (Ontario) K1A 0K9
Téléphone : (613) 954-5851
Télécopieur : (613) 954-2292
Courriel : Murray_Kaiserman@hc-sc.gc.ca

Judy Snider
Bureau de la recherche, de la surveillance et de l'évaluation
Programme de la lutte au tabagisme
Direction générale de la santé environnementale et de la sécurité des consommateurs
Immeuble MacDonald, ind. l'adr. 3507C
123, rue Slater, pièce A716
Ottawa (Ontario) K1A 0K9
Téléphone : (613) 957-0697
Télécopieur : (613) 954-2292
Courriel : Judy_Snider@hc-sc.gc.ca

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.

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 immédiatement avant l'adoption de cette modification, mesurer exactement l'impact de ce changement était difficile pour Santé Canada et pour les analystes intéressés.

Vu le désir de Santé Canada à suivre de près les conséquences des changements législatifs ainsi que l'effet des politiques antitabac sur les comportements des fumeurs, elle a mis sur pied l'Enquête de surveillance de l'usage du tabac au Canada (ESUTC). Cette enquête a pour but de fournir à Santé Canada ainsi qu'à ses partenaires, des données continues et fiables sur l'usage du tabac et sur des sujets connexes.

Depuis 1999, deux fichiers de l'ESUTC ont été publiés annuellement. Le premier contient des données recueillies durant la période de février à juin et l'autre, de juillet à décembre. De plus, un sommaire de l'année est distribué. Le fichier actuel comprend les données pour la période de février à juin 2003.

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 ainsi qu'annuellement. 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 qui 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.

Âge exact en comparaison à l'âge de la liste

L'information sur l'âge du répondant est obtenue à partir de deux sources : d'un répondant du ménage qui fourni l'âge de tous les membres du ménage (âge de la liste) et, à la fin de l'interview, de la part de la personne répondante elle-même à qui on demande sa date de naissance. L'âge exact (DVAGE) provient de la date de naissance et l'âge de la liste est utilisé uniquement lorsque la date de naissance n'est pas disponible. Habituellement, il y a des divergences entre l'âge de la liste et l'âge exact. Alors que l'âge de la liste détermine l'enchaînement des questions et sert à la vérification du fichier, les variables de l'âge dans le fichier définitif révisé, correspondent à l'âge exact.

5.0 Méthodologie de l'enquête

L'Enquête de surveillance de l'usage du tabac au Canada a été réalisée entre le 1er février et le 30 juin 2003 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

La population cible pour l'Enquête de surveillance de l'usage du tabac au Canada était toute personne âgée de 15 ans ou plus vivant au Canada à l'exception des personnes suivantes :

1) les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, et
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

Afin de s'assurer que des personnes de partout au Canada étaient 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 d'une région métropolitaine de recensement (RMR) et une strate d'une région autre qu'une RMR étaient délimitées. Dans l'Île-du-Prince-Édouard, il n'y avait qu'une strate. En Ontario et au Québec, il y avait une troisième strate pour Toronto et une troisième strate pour Montréal. Les RMR sont des régions définies aux fins du recensement et correspondent approximativement à des villes de 100 000 habitants ou plus.

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 et 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***. On cherche 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 téléphone 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 âgées de 25 ans ou plus ont été sous-échantillonnés. On a estimé qu'au total, presque 130 000 numéros de téléphone 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 40 %.

Les objectifs visés ainsi que les numéros de téléphone requis ont été ajustés en conséquence.

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 des cinq prochains 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 s'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 âgée 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 selon le 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 sélectionnée 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 du groupe 25 ans et plus 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 du groupe 25 ans et plus 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 utilisé pour le Cycle 1 de 2003 de l'Enquête de surveillance de l'usage du tabac au Canada contient plusieurs questions qui étaient légèrement différentes ou qui n'ont pas été posées au cours du Cycle 2 de 2002. Des spécifications définissant les limites valides et garantissant la cohérence d'une question à l'autre ont été intégrées dans l'application de l'interview téléphonique assistée par ordinateur (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

Les interviews ont été menées chaque mois de février à juin 2003.

Les données ont été recueillies à l'aide des techniques de l'interview téléphonique assistée par ordinateur. Le système ITAO contient plusieurs modules génériques qui s'adaptent rapidement à la plupart des types d'enquêtes. 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 l'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 est mis en œuvre pour conserver le même groupe d'intervieweurs chaque mois. Cela réduit le temps de formation et produit 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 résumé des phases de traitement inhérentes à la production de ces fichiers.

7.1 Saisie des données

Les données ayant été recueillies à l'aide de l'interview téléphonique assistée par ordinateur (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 les fichiers mensuels en un seul fichier. Toutes les valeurs « hors-limites » des fichiers de données ont été remplacées par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) renfermaient des réponses ont été le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Les erreurs qui englobaient un manque d'information à l'intérieur de questions auxquelles on aurait dû répondre ont été le second type d'erreurs traitées. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

7.3 Création de variables calculées

Un certain nombre d'éléments d'information inclus dans le fichier de microdonnées ont été calculés en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Le nombre moyen de cigarettes fumées quotidiennement et le nombre d'années que le répondant fume sont des exemples de variables calculées.

7.4 Pondération

Le principe qui sous-tend une estimation dans le cas d'un échantillon probabiliste veut que chaque personne incluse dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Chaque personne incluse dans un échantillon aléatoire simple de 2 % de la population représente, par exemple, 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes au Canada qui fument tous les jours doit être estimé, cette opération s'effectue en sélectionnant les enregistrements renvoyant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique (Q010 = 1) et en additionnant les poids inscrits dans ces enregistrements. La pondération des ménages et des personnes se fait séparément tous les six mois.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.5 Suppression des renseignements confidentiels

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » décrits ci-dessus diffèrent sous un certain nombre d'aspects importants des « fichiers principaux » de l'enquête que conserve Statistique Canada. Ces différences sont le résultat de mesures prises pour protéger l'anonymat des répondants à un questionnaire d'enquête. Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion d'information dont le chapitre 9.0 du présent document fournit un aperçu.

Identificateurs géographiques

Fichier des ménages et Fichier des personnes :

Les fichiers principaux des données de l'enquête incluent des identificateurs géographiques explicites pour la province et la strate (région métropolitaine de recensement (RMR), région autre qu'une RMR, Toronto ou Montréal). Les fichiers de microdonnées à grande diffusion de l'enquête comprennent seulement un identificateur pour la province.

Fichier des personnes :

Depuis le 1er cycle de 2002 le fichier principal de données comprend les trois premiers chiffres du code postal du répondant.

Composition du ménage selon l'âge

Fichier des ménages et Fichier des personnes :

On peut obtenir la composition du ménage selon l'âge, c'est-à-dire le nombre de membres du ménage (limité à deux) dans les groupes d'âge suivants : 0 à 14 ans, 15 à 24 ans, 25 à 44 ans et 45 ans et plus.

État matrimonial du répondant

Fichier des personnes :

La variable détaillée sur l'état matrimonial (six catégories) est accessible uniquement dans le fichier principal, tandis que dans le fichier de microdonnées à grande diffusion, cette variable a été groupée en trois catégories.

Autres modifications apportées au Fichier des ménages et au Fichier des personnes

Pour un petit nombre d'enregistrements (moins de 20) du Fichier des ménages, la variable démographique a dû être recodée pour éviter l'identification potentielle des répondants résultant d'une combinaison inhabituelle de caractéristiques. Le Fichier des personnes a également fait l'objet d'un recodage similaire.

8.0 Qualité des données

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

Le taux de numéros de téléphone résolus et le taux de succès s'appliquent à la fois au Fichier des ménages et au Fichier des personnes.

- Taux de numéros de téléphone résolus, un numéro de téléphone résolu étant un numéro résidentiel, d'affaires ou non admissible confirmés.

- Taux de succès, un succès étant un numéro résidentiel confirmé appartenant à un ménage contenant des **données valides sur le ménage** ou contenant des **données valides sur la personne**.

Ce taux de réponse des ménages ne s'applique qu'au Fichier des ménages.

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

Le taux de réponse des ménages, le taux d'achèvement des listes, le taux de réponse des personnes et le taux de réponse combiné des ménages ne s'appliquent que pour le Fichier des personnes.

- Taux de réponse des ménages, une réponse étant un numéro de téléphone résidentiel confirmé ou des **données valides sur la personne** existaient avec une liste remplie.

- Taux d'achèvement des listes, une réponse étant un ménage pour lequel l'âge de toutes les personnes sur la liste a été fourni.

- Taux de réponse des personnes, une réponse étant un enregistrement qui contient l'âge de toutes les personnes inscrites sur la liste ainsi qu'un rapport sur les cas de réponse des ménages correspondants et des **données valides sur la personne**.

- Taux de réponse combiné des ménages égal le taux de réponse des ménages multiplié par le taux de réponse des personnes.

Taux de numéros de téléphone résolus et taux de succès selon la province

Provinces	**Nombre total de n^{os} de téléphone produits**	**Total des n^{os} résolus**	**Taux de n^{os} de téléphone résolus (%)**	**Nombre total de ménages**	**Taux de succès (%)**
Terre-Neuve-et-Labrador	6 903	6 721	97,4	2 301	34,2
Île-du-Prince-Édouard	6 077	5 802	95,5	2 399	41,3
Nouvelle-Écosse	6 817	6 564	96,3	2 699	41,1
Nouveau-Brunswick	8 998	8 652	96,2	3 006	34,7
Québec	6 005	5 916	98,5	3 029	51,2
Ontario	6 570	6 263	95,3	2 390	38,2
Manitoba	6 872	6 872	100,0	2 903	42,2
Saskatchewan	6 669	6 669	100,0	2 596	38,9
Alberta	5 370	5 370	100,0	2 559	47,7
Colombie-Britannique	5 995	5 717	95,4	2 601	45,5
Canada	**66 276**	**64 546**	**97,4**	**26 483**	**41,0**

8.1 Taux de réponse des ménages – février à juin 2003

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 3 289 (12,4 %) ménages non répondants.

Taux de réponse des ménages selon la province

Provinces	**Nombre total de ménages**	**Ménages répondants**	**Taux de réponse des ménages (%)**
Terre-Neuve-et-Labrador	2 301	2 092	90,9
Île-du-Prince-Édouard	2 399	2 161	90,1
Nouvelle-Écosse	2 699	2 434	90,2
Nouveau-Brunswick	3 006	2 711	90,2
Québec	3 029	2 597	85,7
Ontario	2 390	2 148	89,9
Manitoba	2 903	2 494	85,9
Saskatchewan	2 596	2 180	84,0
Alberta	2 559	2 076	81,1
Colombie-Britannique	2 601	2 301	88,5
Canada	**26 483**	**23 194**	**87,6**

Taux de réponse des ménages selon le mois d'enquête

Mois de l'enquête	**Nombre total de ménages**	**Ménages répondants**	**Taux de réponse des ménages (%)**
Février	5 160	4 615	89,4
Mars	5 365	4 758	88,7
Avril	5 266	4 654	88,4
Mai	5 279	4 556	86,3
Juin	5 413	4 611	85,2
Total	**26 483**	**23 194**	**87,6**

8.2 Taux de réponse des personnes - février à juin 2003

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

- Le numéro de téléphone de 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 était 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 a répondu au moins aux questions clés concernant les habitudes de fumer.

Dans 12 890 ménages, des données sur le ménage ont été recueillies mais personne n'a été sélectionné pour continuer l'ESUTC. (Pour plus d'information, voir (Tirage de l'échantillon) à la section 5.4.) Dans les ménages restants, on a sélectionné une personne dans 8 824 ménages et deux personnes dans 1 564 ménages. Il y a eu seulement 158 (1,3 %) personnes qui ont refusé de remplir le questionnaire et 1 235 (10,3 %) personnes qui n'ont pas répondu pour d'autres raisons.

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

Provinces	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 combiné des ménages (%)
Terre-Neuve-et-Labrador	2 301	2 155	93,7	2 100	97,4	91,3
Île-du-Prince-Édouard	2 399	2 248	93,7	2 166	96,4	90,3
Nouvelle-Écosse	2 699	2 512	93,1	2 445	97,3	90,6
Nouveau-Brunswick	3 006	2 785	92,6	2 714	97,5	90,3
Québec	3 029	2 693	88,9	2 609	96,9	86,1
Ontario	2 390	2 189	91,6	2 152	98,3	90,0
Manitoba	2 903	2 543	87,6	2 505	98,5	86,3
Saskatchewan	2 596	2 226	85,7	2 191	98,4	84,4
Alberta	2 559	2 118	82,8	2 087	98,5	81,6
Colombie-Britannique	2 601	2 367	91,0	2 309	97,5	88,8
Canada	**26 483**	**23 836**	**90,0**	**23 278**	**97,7**	**87,9**

Taux de réponse des personnes selon la province

Provinces	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes (%)
Terre-Neuve-et-Labrador	1 204	1 024	85,0
Île-du-Prince-Édouard	1 225	1 056	86,2
Nouvelle-Écosse	1 216	1 066	87,7
Nouveau-Brunswick	1 467	1 285	87,6
Québec	1 317	1 169	88,8
Ontario	1 043	963	92,3
Manitoba	1 243	1 108	89,1
Saskatchewan	1 145	993	86,7
Alberta	1 116	970	86,9
Colombie-Britannique	976	925	94,8
Canada	**11 952**	**10 559**	**88,3**

Taux de réponse des personnes selon le mois d'enquête

Mois de l'enquête	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponse des personnes (%)
Février	2 369	2 126	89,7
Mars	2 486	2 238	90,0
Avril	2 387	2 110	88,4
Mai	2 328	2 038	87,5
Juin	2 382	2 047	85,9
Total	**11 952**	**10 559**	**88,3**

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

Groupes d'âge	Nombre cible de répondants	Total des personnes sélectionnées	Total des personnes répondantes	Taux de réponses des personnes (%)
15 à 19	2 500	3 165	2 791	88,2
20 à 24	2 500	2 718	2 345	86,3
25 et plus	5 000	6 069	5 423	89,4
Total	**10 000**	**11 952**	**10 559**	**88,3**

8.3 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de ménages. Des estimations légèrement différentes auraient pu être obtenues si un recensement exhaustif avait été effectué en reprenant par exemple le même questionnaire, les mêmes intervieweurs et superviseurs et les mêmes méthodes de traitement que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement exhaustif réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas reliées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être saisies de façon incorrecte sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprennent la formation poussée des intervieweurs concernant les procédures de l'enquête et de l'application de l'interview téléphonique assistée par ordinateur (ITAO), l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, et l'évaluation de 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. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant ou que le répondant a refusé de participer à l'enquête. Les non-réponses totales ont été traitées en ajustant le poids des ménages ou des personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu.

8.5 Non-réponse partielle

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque le répondant n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée. Des codes dans le fichier de microdonnées indiquent les non-réponses partielles, c'est-à-dire refus, ne sait pas.

8.6 Couverture

Tel qu'il est mentionné à la section 5.1 (Champ de l'enquête), moins de 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

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. La présente section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et dont le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées à employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison cependant de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) 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 que, d'après les résultats de l'enquête, l'on estime que 19,9 % des Canadiens fument actuellement la cigarette, et l'on constate que l'erreur-type de cette estimation est de 0,0074. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,0074}{0,199} \right) X\ 100\ \%\ =\ 3,7\ \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouve au chapitre 10.0.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada ainsi que des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs à respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier 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, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.

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

c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, 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, le dernier chiffre à conserver est augmenté de 1.

d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.

e) Dans les cas, où, en raison de limitations d'ordre techniques ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.

f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles 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 (ESUTC) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'échantillonnage approprié.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du champ du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et 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 types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'ESUTC.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes qui fument actuellement des cigarettes ou la proportion de fumeurs quotidiens qui ont tenté de cesser de fumer constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée une estimation d'un agrégat.

Exemples de questions catégoriques :

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 : Ma propre santé / Grossesse, a eu un bébé / Inquiétude au sujet de la santé des membres de la famille / Ne voulait plus fumer, mauvaise odeur, dégoûtant / Moins de stress dans la vie, emploi différent / Maladie liée à la cigarette ou décès d'un parent ou d'un ami / Pressions pour cesser de fumer exercée par la famille ou des amis / Le coût / Restrictions sur les endroits où il est permis de fumer / Voulait simplement cesser de fumer, aucune raison en particulier / Autre (*précisez*)

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures d'une tendance centrale de quantités reposant sur certains ou sur tous les membres de la population visée par l'enquête. Elles

comprennent aussi expressément 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 cigarettes fumées, les samedis, par personne. Le numérateur $\left(\hat{X}\right)$ est une estimation du nombre total de cigarettes fumées les samedis et son dénominateur $\left(\hat{Y}\right)$ est le nombre de personnes ayant déclaré avoir fumé les samedis.

Exemples de questions quantitatives :

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

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

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique 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 de la forme $\hat{X} / \hat{Y}$ en :

a) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur $\left(\hat{X}\right)$,

b) additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur $\left(\hat{Y}\right)$, puis en

c) divisant l'estimation a) par celle de b) $\left(\hat{X} / \hat{Y}\right)$.

9.3.4 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. Pour obtenir, par exemple, une estimation du nombre total de cigarettes fumées le samedi, multipliez la valeur déclarée à la question Q090SAT (nombre de cigarettes fumées le samedi) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements portant la mention Q090SAT < 96 (tous les répondants qui ont donné une réponse à ce champ).
Pour obtenir une moyenne pondérée de la forme $\hat{X} / \hat{Y}$, le numérateur $\left(\hat{X}\right)$ est calculé comme une estimation quantitative et le dénominateur $\left(\hat{Y}\right)$ est calculé comme une estimation catégorique. Pour estimer, par exemple, le nombre moyen de cigarettes fumées le samedi,

a) estimez le nombre total de cigarettes fumées le samedi $\left(\hat{X}\right)$, tel qu'il est décrit ci-dessus,

b) estimez le nombre de personnes $\left(\hat{Y}\right)$ incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements ayant la mention Q090SAT < 96, puis

c) divisez l'estimation a) par l'estimation b) $\left(\hat{X} / \hat{Y}\right)$.

9.4 Lignes directrices pour l'analyse statistique

L'Enquête de surveillance de l'usage du tabac au Canada repose sur un plan d'échantillonnage complexe comportant une stratification, plusieurs degrés de sélection et des probabilités inégales de sélection des répondants. L'utilisation des données tirées d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influencent les procédures d'estimation et de calcul de la variance qu'il faudrait utiliser. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent l'utilisation de poids, la signification ou la définition du poids inclus dans ces procédures diffère de ce qui convient à la base de sondage d'une enquête-échantillon, de telle sorte que dans bien des cas les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont piètres. Des variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faut effectuer l'analyse de tous les répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

1) sélectionner tous les répondants du fichier qui ont déclaré SEXE = homme;

2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements, puis le diviser par le nombre de répondants ayant déclaré SEXE = homme;

3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original des personnes divisé par le poids MOYEN;

4) effectuer l'analyse portant sur ces répondants en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations des variances calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations des variances plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées en raison de la confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'Enquête de surveillance de l'usage du tabac au Canada, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable, médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage influencent la qualité des données comme il en a été question au chapitre 8.0. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable.

Pour les estimations pondérées fondées sur les tailles d'échantillons de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations arrondies pondérées.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	Estimations : d'une taille d'échantillon de 30 ou plus et de faibles coefficients de variation de l'ordre de 0,0 % à 16,5 %. Aucune mise en garde n'est requise.
2) Médiocre	Estimations : d'une taille d'échantillon de 30 ou plus et de coefficients de variation élevés de l'ordre de 16,6 % à 33,3 %. Ces estimations devraient être signalées par la lettre M (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.
3) Inacceptable	Estimations : d'une taille d'échantillon inférieur à 30, ou de coefficients de variation très élevés, supérieur à 33,3 %. Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre I (ou d'un quelconque identificateur similaire) et être accompagnées de la mise en garde suivante : « Nous informons l'utilisateur que ces estimations (désignées avec la lettre I)... ne respectent pas les normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions qui reposeront sur ces données ne seront ni fiables ni valables très probablement. »

9.6 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des ménages

La taille minimum des estimations pour les ménages au niveau provincial est indiquée dans le tableau ci-dessous. Les estimations plus petites que la taille minimum indiquée dans la colonne « Inacceptable » doivent être signalées de façon pertinente.

Tableau des seuils pour la diffusion des estimations – Fichier des ménages

Provinces	**Acceptable CV 0,0 % - 16,5 %**	**Médiocre CV 16,6 % - 33,3 %**			**Inacceptable CV > 33,3 %**	
Terre-Neuve-et-Labrador	4 500 & plus	1 000	à <	4 500	moins de	1 000
Île-du-Prince-Édouard	1 000 & plus	500	à <	1 000	moins de	500
Nouvelle-Écosse	6 500 & plus	1 500	à <	6 500	moins de	1 500
Nouveau-Brunswick	4 500 & plus	1 000	à <	4 500	moins de	1 000
Québec	54 000 & plus	13 500	à <	54 000	moins de	13 500
Ontario	91 000 & plus	22 500	à <	91 000	moins de	22 500
Manitoba	7 500 & plus	2 000	à <	7 500	moins de	2 000
Saskatchewan	6 500 & plus	1 500	à <	6 500	moins de	1 500
Alberta	25 500 & plus	6 500	à <	25 500	moins de	6 500
Colombie-Britannique	31 000 & plus	7 500	à <	31 000	moins de	7 500
Canada	**53 000 & plus**	**13 000**	**à <**	**53 000**	**moins de**	**13 000**

9.7 Seuils pour la diffusion des estimations pour l'Enquête de surveillance de l'usage du tabac au Canada - Fichier des personnes

Le tableau suivant spécifie la taille minimum des estimations selon la province et le groupe d'âge. Les estimations plus petites que la taille minimum présentées dans la colonne « Inacceptable » doivent être signalées de façon pertinente.

Tableau des seuils pour la diffusion des estimations – Fichier des personnes

Provinces	**Groupes d'âge**	**Acceptable CV 0,0 % - 16,5 %**	**Médiocre CV 16,6 % - 33,3 %**	**Inacceptable CV > 33,3 %**
Terre-Neuve-et-Labrador	Tous	30 500 & plus	8 000 à < 30 500	moins de 8 000
	15-19	5 500 & plus	1 500 à < 5 500	moins de 1 500
	20-24	7 500 & plus	2 000 à < 7 500	moins de 2 000
	25+	33 500 & plus	9 000 à < 33 500	moins de 9 000
Île-du-Prince-Édouard	Tous	8 000 & plus	2 000 à < 8 000	moins de 2 000
	15-19	1 500 & plus	500 à < 1 500	moins de 500
	20-24	2 000 & plus	500 à < 2 000	moins de 500
	25+	7 500 & plus	2 000 à < 7 500	moins de 2 000
Nouvelle-Écosse	Tous	58 000 & plus	15 000 à < 58 000	moins de 15 000
	15-19	10 000 & plus	3 000 à < 10 000	moins de 3 000
	20-24	11 000 & plus	3 000 à < 11 000	moins de 3 000
	25+	61 000 & plus	16 000 à < 61 000	moins de 16 000
Nouveau-Brunswick	Tous	35 000 & plus	9 000 à < 35 000	moins de 9 000
	15-19	6 000 & plus	1 500 à < 6 000	moins de 1 500
	20-24	9 000 & plus	2 500 à < 9 000	moins de 2 500
	25+	37 000 & plus	9 500 à < 37 000	moins de 9 500
Québec	Tous	347 500 & plus	89 000 à < 347 500	moins de 89 000
	15-19	75 000 & plus	21 000 à < 75 000	moins de 21 000
	20-24	83 000 & plus	23 500 à < 83 000	moins de 23 500
	25+	407 500 & plus	106 500 à < 407 500	moins de 106 500
Ontario	Tous	743 500 & plus	193 500 à < 743 500	moins de 193 500
	15-19	140 500 & plus	39 500 à < 140 500	moins de 39 500
	20-24	176 000 & plus	52 000 à < 176 000	moins de 52 000
	25+	815 500 & plus	216 500 à < 815 500	moins de 216 500
Manitoba	Tous	54 000 & plus	14 000 à < 54 000	moins de 14 000
	15-19	12 000 & plus	3 500 à < 12 000	moins de 3 500
	20-24	14 500 & plus	4 000 à < 14 500	moins de 4 000
	25+	60 000 & plus	15 500 à < 60 000	moins de 15 500
Saskatchewan	Tous	53 500 & plus	14 000 à < 53 500	moins de 14 000
	15-19	12 500 & plus	3 500 à < 12 500	moins de 3 500
	20-24	15 000 & plus	4 500 à < 15 000	moins de 4 500
	25+	57 000 & plus	15 000 à < 57 000	moins de 15 000
Alberta	Tous	170 500 & plus	44 000 à < 170 500	moins de 44 000
	15-19	42 500 & plus	12 000 à < 42 500	moins de 12 000
	20-24	51 500 & plus	15 000 à < 51 500	moins de 15 000
	25+	191 000 & plus	50 500 à < 191 000	moins de 50 500
Colombie-Britannique	All	288 500 & plus	75 500 à < 288 500	moins de 75 500
	15-19	49 000 & plus	14 000 à < 49 000	moins de 14 000
	20-24	96 500 & plus	32 000 à < 96 500	moins de 32 000
	25+	312 500 & plus	83 500 à < 312 500	moins de 83 500
Canada	**Tous**	**465 500 & plus**	**116 000 à < 465 500**	**moins de 116 000**
	15-19	**89 500 & plus**	**22 500 à < 89 500**	**moins de 22 500**
	20-24	**128 000 & plus**	**33 000 à < 128 000**	**moins de 33 000**
	25+	**514 000 & plus**	**128 500 à < 514 000**	**moins de 128 500**

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient approximatif de variation fondé sur la taille de l'estimation calculée à partir des données d'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ces dernières une valeur modérée (habituellement le 75^{e} percentile) à utiliser à l'intérieur de ces tables de CV qui s'appliqueraient ensuite à l'ensemble entier de caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête de surveillance de l'usage du tabac au Canada (ESUTC) pour le Fichier des ménages.

Fichier des ménages

Provinces	**Effet du plan**	**Taille de l'échantillon**	**Population**
Terre-Neuve-et-Labrador	1,30	2 092	197 409
Île-du-Prince-Édouard	1,19	2 161	53 940
Nouvelle-Écosse	1,15	2 434	370 187
Nouveau-Brunswick	1,10	2 711	293 187
Québec	1,25	2 597	3 113 119
Ontario	1,19	2 148	4 553 928
Manitoba	1,20	2 494	431 024
Saskatchewan	1,01	2 180	382 897
Alberta	1,25	2 076	1 167 577
Colombie-Britannique	1,22	2 301	1 624 565
Canada	**2,77**	**23 194**	**12 187 833**

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province et le groupe d'âge qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'ESUTC pour le Fichier des personnes.

Fichier des personnes

Provinces	Groupes d'âge	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	Tous	2,07	1 024	440 404
	15-19	1,38	292	37 490
	20-24	1,49	216	36 611
	25+	1,41	516	366 303
Île-du-Prince-Édouard	Tous	2,14	1 056	113 356
	15-19	1,33	294	10 435
	20-24	1,42	219	9 804
	25+	1,32	543	93 117
Nouvelle-Écosse	Tous	2,36	1 066	772 242
	15-19	1,55	302	64 126
	20-24	1,51	254	61 640
	25+	1,45	510	646 476
Nouveau-Brunswick	Tous	2,10	1 285	617 513
	15-19	1,35	351	49 837
	20-24	1,70	274	49 536
	25+	1,38	660	518 140
Québec	Tous	1,92	1 169	6 111 588
	15-19	1,51	279	453 260
	20-24	1,48	279	509 588
	25+	1,43	611	5 148 740
Ontario	Tous	2,15	963	9 809 759
	15-19	1,41	245	805 552
	20-24	1,45	189	800 484
	25+	1,59	529	8 203 722
Manitoba	Tous	1,92	1 108	905 744
	15-19	1,30	276	82 457
	20-24	1,50	243	78 345
	25+	1,40	589	744 942
Saskatchewan	Tous	1,98	993	785 351
	15-19	1,37	263	76 321
	20-24	1,55	211	71 280
	25+	1,39	519	637 750
Alberta	Tous	1,93	970	2 501 664
	15-19	1,50	243	230 472
	20-24	1,57	215	242 609
	25+	1,45	512	2 028 583
Colombie-Britannique	Tous	2,32	925	3 419 246
	15-19	1,61	272	274 769
	20-24	2,86	204	284 223
	25+	1,50	449	2 860 255
Canada	**Tous**	**5,35**	**10 559**	**25 476 868**
	15-19	**3,44**	**2 817**	**2 084 720**
	20-24	**3,98**	**2 304**	**2 144 120**
	25+	**3,67**	**5 438**	**21 248 028**

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée permettrait aux utilisateurs de passer d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée devrait être considérée *inacceptable* et devrait être signalée de façon pertinente, 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 coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles 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. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est 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 estimée 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 repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé est plus fiable que le nombre estimé d'anciens fumeurs qui ont cessé pour des problèmes actuels de santé. (Remarquez que dans les tables, la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex., comme ses membres d'un sexe ou d'un groupe d'âge particulier), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des 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 considérée séparément. C'est-à-dire que l'erreur-type d'une différence $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ est :

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

où $\hat{X}_1$ est l'estimation 1, $\hat{X}_2$ est l'estimation 2 et α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et $\hat{X}_2$ respectivement. Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}} / \hat{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de fumeurs et le numérateur, le nombre de fumeurs quotidiens.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de fumeurs quotidiens comparativement au nombre de non-fumeurs, l'erreur-type du rapport des estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément multipliée par $\hat{R}$. C'est-à-dire que l'erreur-type d'un rapport $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ 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 $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur, si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés positivement et à la sous-estimer si $\hat{X}_1$ et $\hat{X}_2$ sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier annuel de 2002 et 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 5 414 335 le nombre de personnes qui étaient 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 des coefficients de variation pour le CANADA – Tous les âges du Fichier des personnes.

Enquête de surveillance sur l'usage du tabac au Canada 2002 - février à décembre - Fichier des personnes

Tables de la variabilité d'échantillonnage approximative pour le Canada - Tous les âges

NUMÉRATEUR DU POURCENTAGE	POURCENTAGE ESTIMÉ													
('000)	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	197,2	196,3	195,3	192,3	187,1	181,9	176,4	170,8	165,0	159,0	152,8	139,5	108,0	62,4
2	139,4	138,8	138,1	135,9	132,3	128,6	124,8	120,8	116,7	112,5	108,0	98,6	76,4	44,1
3	113,8	113,3	112,7	111,0	108,0	105,0	101,9	98,6	95,3	91,8	88,2	80,5	62,4	36,0
4	98,6	98,1	97,6	96,1	93,6	90,9	88,2	85,4	82,5	79,5	76,4	69,7	54,0	31,2
5	88,2	87,8	87,3	86,0	83,7	81,3	78,9	76,4	73,8	71,1	68,3	62,4	48,3	27,9
:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
:	.	.	.	.	.	.	.	.	.	.	.	.	.	.
75	********	22,7	22,5	22,2	21,6	21,0	20,4	19,7	19,1	18,4	17,6	16,1	12,5	7,2
80	********	21,9	21,8	21,5	20,9	20,3	19,7	19,1	18,5	17,8	17,1	15,6	12,1	7,0
85	********	21,3	21,2	20,9	20,3	19,7	19,1	18,5	17,9	17,2	16,6	15,1	11,7	6,8
90	********	20,7	20,6	20,3	19,7	19,2	18,6	18,0	17,4	16,8	16,1	14,7	11,4	6,6
95	********	20,1	20,0	19,7	19,2	18,7	18,1	17,5	16,9	16,3	15,7	14,3	11,1	6,4
100	********	19,6	19,5	19,2	18,7	18,2	17,6	17,1	16,5	15,9	15,3	13,9	10,8	6,2
125	********	17,6	17,5	17,2	16,7	16,3	15,8	15,3	14,8	14,2	13,7	12,5	9,7	5,6
150	********	16,0	15,9	15,7	15,3	14,8	14,4	13,9	13,5	13,0	12,5	11,4	8,8	5,1
200	********	13,9	13,8	13,6	13,2	12,9	12,5	12,1	11,7	11,2	10,8	9,9	7,6	4,4
250	********	12,4	12,4	12,2	11,8	11,5	11,2	10,8	10,4	10,1	9,7	8,8	6,8	3,9
300	********	********	11,3	11,1	10,8	10,5	10,2	9,9	9,5	9,2	8,8	8,1	6,2	3,6
350	********	********	10,4	10,3	10,0	9,7	9,4	9,1	8,8	8,5	8,2	7,5	5,8	3,3
400	********	********	9,8	9,6	9,4	9,1	8,8	8,5	8,3	8,0	7,6	7,0	5,4	3,1
450	********	********	9,2	9,1	8,8	8,6	8,3	8,1	7,8	7,5	7,2	6,6	5,1	2,9
500	********	********	8,7	8,6	8,4	8,1	7,9	7,6	7,4	7,1	6,8	6,2	4,8	2,8
750	********	********	********	7,0	6,8	6,6	6,4	6,2	6,0	5,8	5,6	5,1	3,9	2,3
1000	********	********	********	6,1	5,9	5,8	5,6	5,4	5,2	5,0	4,8	4,4	3,4	2,0
1500	********	********	********	********	4,8	4,7	4,6	4,4	4,3	4,1	3,9	3,6	2,8	1,6
2000	********	********	********	********	4,2	4,1	3,9	3,8	3,7	3,6	3,4	3,1	2,4	1,4
3000	********	********	********	********	********	3,3	**3,2**	**3,1**	3,0	2,9	2,8	2,5	2,0	1,1
4000	********	********	********	********	********	********	2,8	2,7	2,6	2,5	2,4	2,2	1,7	1,0
5000	********	********	********	********	********	********	**2,5**	2,4	2,3	2,2	2,2	2,0	1,5	0,9
6000	********	********	********	********	********	********	********	2,2	2,1	2,1	2,0	1,8	1,4	0,8
7000	********	********	********	********	********	********	********	********	2,0	1,9	1,8	1,7	1,3	0,7
8000	********	********	********	********	********	********	********	********	********	1,8	1,7	1,6	1,2	0,7
9000	********	********	********	********	********	********	********	********	********	********	1,6	1,5	1,1	0,7
10000	********	********	********	********	********	********	********	********	********	********	1,5	1,4	1,1	0,6
12500	********	********	********	********	********	********	********	********	********	********	********	1,2	1,0	0,6
15000	********	********	********	********	********	********	********	********	********	********	********	********	0,9	0,5

NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.

2) L'agrégat estimé (5 414 335) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 5 000 000.

3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 2,5 %.

4) Le coefficient de variation approximatif de l'estimation est donc 2,5 %. Le résultat selon lequel il y avait 5 414 335 (à être arrondie d'après les lignes directrices tel qu'il est indiqué à la section 9.1) fumeurs durant la période de référence, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime à 2 865 929 / 12 436 728 = 23,0 % la proportion d'hommes qui fument actuellement 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 des coefficients de variation pour le CANADA du Fichier des personnes (voir ci-dessus). Le tableau au niveau du CANADA devrait être 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 fondé sur un sous-ensemble de la population totale (c'est-à-dire les hommes), il faut utiliser à la fois le pourcentage (23,0 %) et la portion numérateur du pourcentage (2 865 929) pour déterminer le coefficient de variation.

3) Le numérateur, 2 865 929, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 3 000 000. De même, l'estimation du pourcentage ne figure sous aucun des titres de colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 25,0 %.

4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 3,1 %, est le coefficient de variation à employer.

5) Le coefficient de variation approximatif de l'estimation est donc 3,1 %. Le résultat selon lequel 23,0 % des hommes fument actuellement, peut être publié sans réserve.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime à 2 548 406 / 12 814 359 = 19,9 % la proportion de femmes qui fument actuellement au Canada et à 2 865 929 / 12 436 728 = 23,0 % la proportion d'hommes qui fument actuellement au Canada. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

1) L'utilisation de la table des coefficients de variation pour le CANADA du Fichier des personnes (voir ci-dessus) de la même façon que celle décrite dans l'exemple 2 donne un CV de l'estimation pour les femmes de 3,2 % et un CV de l'estimation pour les hommes de 3,1 %.

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

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

où $\hat{X}_1$ est l'estimation 1 (hommes), $\hat{X}_2$ est l'estimation 2 (femmes) et α_1 et α_2 sont les coefficients de variation de $\hat{X}_1$ et de $\hat{X}_2$ respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d}$ = 0,230 – 0,199 = 0,031 est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0{,}230)(0{,}031)]^2 + [(0{,}199)(0{,}032)]^2} \\ &= \sqrt{(0{,}00005) + (0{,}00004)} \\ &= 0{,}009\end{aligned}$$

3) Le coefficient de variation de $\hat{d}$ est donné par

$\sigma_{\hat{d}} / \hat{d}$ = 0,009 / 0,031 = 0,290.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 29,0 %. Cette estimation est considérée médiocre et Statistique Canada recommande que cette estimation soit signalée par la lettre M (ou un quelconque identificateur similaire) et qu'elle soit accompagnée d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés à l'estimation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 237 261 le nombre de femmes qui fument actuellement agées de 15 à 19 ans et à 220 511 le nombre d'hommes qui fument actuellement 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 par quotient, où le numérateur de l'estimation ($\hat{X}_1$) est le nombre de femmes qui fument actuellement agées de 15 à 19 ans. Le dénominateur de l'estimation ($\hat{X}_2$) est le nombre d'hommes qui fument actuellement agés de 15 à 19 ans.

2) Reportez-vous à la table des coefficients de variation pour le CANADA – 15 à 19 ans du Fichier des personnes.

Enquête de surveillance sur l'usage du tabac au Canada 2002 - février à décembre - Fichier des personnes

Tables de la variabilité d'échantillonnage approximative pour le Canada - 15 à 19 ans

NUMÉRATEUR DU POURCENTAGE	POURCENTAGE ESTIMÉ													
('000)	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	95,8	95,3	94,9	93,4	90,9	88,3	85,7	83,0	80,2	77,3	74,2	67,8	52,5	30,3
2	67,7	67,4	67,1	66,0	64,3	62,5	60,6	58,7	56,7	54,6	52,5	47,9	37,1	21,4
3	********	55,0	54,8	53,9	52,5	51,0	49,5	47,9	46,3	44,6	42,9	39,1	30,3	17,5
4	********	47,7	47,4	46,7	45,5	44,2	42,9	41,5	40,1	38,6	37,1	33,9	26,2	15,2
5	********	42,6	42,4	41,8	40,7	39,5	38,3	37,1	35,9	34,6	33,2	30,3	23,5	13,6
6	********	38,9	38,7	38,1	37,1	36,1	35,0	33,9	32,7	31,5	30,3	27,7	21,4	12,4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
95	********	********	********	9,6	9,3	9,1	8,8	8,5	8,2	7,9	7,6	7,0	5,4	3,1
100	********	********	********	9,3	9,1	8,8	8,6	8,3	8,0	7,7	7,4	6,8	5,2	3,0
125	********	********	********	********	8,1	7,9	7,7	7,4	7,2	6,9	6,6	6,1	4,7	2,7
150	********	********	********	********	7,4	7,2	7,0	6,8	6,5	6,3	6,1	5,5	4,3	2,5
200	********	********	********	********	**6,4**	6,2	6,1	5,9	5,7	5,5	5,2	4,8	3,7	2,1
250	********	********	********	********	********	**5,6**	5,4	5,2	5,1	4,9	4,7	4,3	3,3	1,9
300	********	********	********	********	********	5,1	4,9	4,8	4,6	4,5	4,3	3,9	3,0	1,7
350	********	********	********	********	********	********	4,6	4,4	4,3	4,1	4,0	3,6	2,8	1,6
400	********	********	********	********	********	********	4,3	4,1	4,0	3,9	3,7	3,4	2,6	1,5
450	********	********	********	********	********	********	********	3,9	3,8	3,6	3,5	3,2	2,5	1,4
500	********	********	********	********	********	********	********	3,7	3,6	3,5	3,3	3,0	2,3	1,4
750	********	********	********	********	********	********	********	********	********	********	2,7	2,5	1,9	1,1
1000	********	********	********	********	********	********	********	********	********	********	********	2,1	1,7	1,0
1500	********	********	********	********	********	********	********	********	********	********	********	********	********	0,8

NOTA : Pour utiliser ces tables correctement, veuillez vous référer à la documentation reliée aux microdonnées.

3) Le numérateur de cette estimation de rapport est 237 261. Le chiffre qui s'en rapproche le plus est 250 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 5,6 %.

4) Le dénominateur de cette estimation de rapport est 220 511. Le chiffre qui s'en rapproche le plus est 200 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 6,4 %.

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

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

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

C'est-à-dire que :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0{,}056)^2 + (0{,}064)^2} \\ &= \sqrt{0{,}003136 + 0{,}004096} \\ &= 0{,}085\end{aligned}$$

6) Le rapport obtenu des femmes:hommes fumeurs actuels agés de 15 à 19 ans est égal à 237 261 / 220 511, c'est-à-dire 1,08 (à être arrondie d'après les lignes directrices tel qu'il est indiqué à la section 9.1). Le coefficient de variation de cette estimation est 8,5 %, ce qu'on peut diffuser sans réserve.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation sont beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suis :

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

On peut obtenir, à l'aide de l'erreur-type d'une estimation, des intervalles de confiance pour des estimations en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement 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.

On exprime généralement des intervalles de confiance pour une estimation, $\hat{X}$, sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $\left(\hat{X} - k, \hat{X} + k\right)$, où on détermine k suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement des intervalles de confiance pour une estimation à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation $\hat{X}$, puis en utilisant la formule suivante pour le convertir à un intervalle de confiance $\left(IC_{\hat{x}}\right)$:

$$IC_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de $\hat{X}$, et

t = 1 si l'on désire un intervalle de confiance de 68 %;
t = 1,6 si l'on désire un intervalle de confiance de 90 %;
t = 2 si l'on désire un intervalle de confiance de 95 %;
t = 2,6 si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices sur la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée des hommes qui fument actuellement (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$\hat{X}$ = 23,0 % (ou exprimé sous forme de proportion 0,230)

t = 2

$\alpha_{\hat{x}}$ = 3,1 % (0,031 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$IC_{\hat{x}}$ = {0,230 - (2) (0,230) (0,031), 0,230 + (2) (0,230) (0,031)}

$IC_{\hat{x}}$ = {0,230 - 0,014, 0,230 + 0,014}

$IC_{\hat{x}}$ = {0,216, 0,244}

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 21,6 % et 24,4 % des hommes fument actuellement.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

On peut aussi utiliser des erreurs-types pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests 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.

Supposons que $\hat{X}_1$ et $\hat{X}_2$ sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$

se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, 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 la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion d'hommes qui fument actuellement et la proportion de femmes qui fument actuellement. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,009. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0{,}230 - 0{,}199}{0{,}009} = \frac{0{,}031}{0{,}009} = 3{,}44 .$$

Puisque t = 3,44 est plus grand que 2, il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'Enquête de surveillance sur l'usage du tabac au Canada sont principalement de nature catégoriques.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de cigarettes fumées le samedi serait supérieur au coefficient de variation de la proportion correspondante de fumeurs actuels. S'il est impossible, par conséquent, de diffuser le coefficient de variation de la proportion, on ne pourra pas non plus alors communiquer le coefficient de variation de l'estimation quantitative correspondante.

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs désireux de calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'en obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables des coefficients de variation - Fichier des ménages

Voir le fichier ESUTC2003_C1_MN_CVTabsF.pdf pour les tables de coefficient de variation pour le Fichier des ménages du Cycle 1 de 2003.

10.6 Tables des coefficients de variation - Fichier des personnes

Voir le fichier ESUTC2003_C1_PR_CVTabsF.pdf pour les tables de coefficient de variation pour le Fichier des personnes du Cycle 1 de 2003.

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 échantillonné représente. Un poids a été calculé pour chaque ménage et sur un fichier différent un poids pour chaque personne.

La pondération pour le fichier annuel 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 du champ de l'enquête, et finalement
- un ajustement pour rendre les estimations de la population consistantes avec les totaux connus soient province-âge-sexe de la population estimée du recensement pour les personnes de 15 ans et plus.

11.1 Procédures de pondération pour les Fichiers des ménages et des personnes

1. Calcul du poids pour le numéro de téléphone

Chaque numéro de téléphone dans l'échantillon a reçu un poids de base, W_1, égal à l'inverse de sa probabilité de sélection.

$$W_1 = \left(\frac{Nombre\,total\,de\,numéros\,de\,téléphone\,pouvant\,être\,échantillonnés\,pour\,province-strate}{Nombre\,de\,numéros\,de\,téléphone\,échantillonés\,pour\,province-strate} \right)$$

Il y avait 66 276 numéros de téléphone dans l'échantillon avec des poids.

2. Ajustement pour les numéros de téléphone non-résolus

Il y a eu 1 730 numéros de téléphone qui n'ont pas été résolus, laissant 64 546 numéros de téléphone résolus. Il n'a pas été déterminé si ces numéros de téléphone non-résolus appartenaient à un ménage, une entreprise ou étaient hors du champ de l'enquête. Chaque numéro de téléphone avait un indicateur signalisant si l'on s'attendait que le numéro serait une résidence, une entreprise ou de genre inconnu. L'ajustement pour les numéros de téléphone non-résolus a été fait par province-strate et le genre de numéro.

Pour chaque genre de ligne-province-strate,

$$W_2 = W_1 * \left(\frac{\sum W_1\ pour\ numéros\,de\,téléphone\,résolus + \sum W_1\ pour\ numéros\,de\,téléphone\,non-résolus}{\sum W_1\ pour\ numéros\,de\,téléphone\,résolus} \right)$$

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

Les numéros de téléphone correspondants à une entreprise, hors service ou hors du champ de l'enquête, tels 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 numéro de téléphone était considéré comme étant celui d'un ménage. Il y avait 38 063 numéros de téléphone qui étaient hors du champ de l'enquête et 26 483 numéros qui 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 téléphoniques différentes dans le ménage ne pouvait être calculé mais que des données existaient pour le **ménage ou pour une personne**, la valeur de 1 a été imputé pour conserver de bonnes données. Après l'imputation, il restait 2 647 numéros de téléphone pour lesquels le nombre de lignes manquait toujours. Donc, il y avait 23 836 ménages dont l'information a été calculée ou imputée. L'ajustement a été fait au niveau province-strate.

$$W_3 = W_2 * \left(\frac{\sum W_2 \ pour\ ménages\ avec\ nombre\ de\ lignes + \sum W_2 \ pour\ ménages\ avec\ nombre\ de\ lignes\ manquant}{\sum W_2 \ pour\ ménages\ avec\ nombre\ de\ lignes} \right)$$

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

Les poids des ménages avec plus d'une ligne téléphonique (avec différents numéros de téléphone) 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 résidentielles distinctes (maximum 4) du ménage.

$$W_4 = \left(\frac{W_3}{Nombre\ de\ lignes\ téléphoniques\ dans\ le\ champ\ d'enquête\ dans\ le\ ménage} \right)$$

11.2 Procédures de pondération pour le Fichier des ménages

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 642 non-répondants. Donc 23 194 ménages pondérés dans le champ d'enquête ont été utilisés et ajustés par province-strate.

$$W_5 = W_4 * \left(\frac{\sum W_4 \ pour\ ménages\ répondants + \sum W_4 \ pour\ ménages\ non-répondants}{\sum W_4 \ pour\ ménages\ répondants} \right)$$

7. Ajustement pour totaux externes connus de ménages 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 externes connus de ménages. Le facteur d'ajustement par province-strate-mois (P-S-M) est défini comme :

$$W_6 = W_5 * \left(\frac{Totaux\ externes\ connus\ de\ ménages\ pour\ P-S-M}{\sum W_5 \ pour\ les\ ménages\ répondants\ dans\ l'échantillon\ par\ P-S-M} \right)$$

Les poids des ménages, W_6, obtenus après cette étape, sont finals et apparaissent sur le fichier de microdonnées du ménage.

11.3 Procédures de pondération pour le Fichier des personnes

6. Ajustement pour les ménages non-répondants

Dans le Fichier des personnes, on trouve uniquement les enregistrements pour lesquels la liste des membres du ménage a été complétée sans refus de fournir l'âge. Il y a eu 558 non-répondants. Donc 23 278 poids pour les ménages dans le champ de l'enquête ont été utilisés et ajustés selon province-strate.

$$W_5 = W_4 * \left(\frac{\sum W_4 \ pour\ ménages\ répondants + \sum W_4 \ pour\ ménages \ non-répondants}{\sum W_4 \ pour\ ménages\ répondants} \right)$$

7. Calcul du poids de groupe

Tous les ménages répondants dans l'enquête ayant une liste complète des membres (c.-à-d., 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 suivants : 15 à 19, 20 à 24 et 25 ans et plus. 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 d'âge 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 de sélectionner le groupe d'âge est de 2 sur 3 groupes d'âge. Le poids est donc augmenté par son inverse.

Si 1 ou 2 groupes d'âge présent(s), $W_6 = W_5$.

Si tous les 3 groupes d'âge présents, $W_6 = W_5 * 3/2$.

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

Il y avait 12 890 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 ont été éliminés parce qu'ils n'avaient aucune information au niveau de la personne. Environ 70 % des répondants sélectionnés âgés de 25 ans et plus n'ont pas été retenues. Il y avait 10 388 ménages pour lesquels nous avons sélectionné un répondant. Il y avait 8 824 ménages avec une personne sélectionnée et 1 564 avec deux personnes sélectionnées.

9. Calcul du poids des ménages pour les personnes sélectionnées

Ces 8 824 + 2(1 564) = 11 952 personnes sélectionnées sont associées avec des ménages répondants faisant partis du champ de l'enquête et ont conservées le poids correspondant, W_6.

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 sélectionné et l'inverse du facteur de sous-échantillonnage.

$$W_7 = W_6 * \left(\frac{Nombre\ de\ personnes\ dans\ le\ groupe\ d'âge\ choisi}{Facteur\ de\ sous-échantillonnage} \right)$$

Le facteur de sous-échantillonnage est de 1 pour les groupes d'âge 15 à 19 et 20 à 24. Le facteur de sous-échantillonnage pour le groupe d'âge 25 ans et plus avait été déjà assigné et variait de 23,0 % à 30.0 %, selon la province.

11. Ajustement pour les individus non-répondants

Le Fichier des personnes comprend les enregistrements des répondants individuels qui ont donné l'information aux questions sur les habitudes de l'usage du tabac et qui ont fourni une date de naissance correspondant à l'âge indiqué dans la liste des membres. Il y a eu 1 393 non-répondants.

Donc, nous avons utilisé 10 559 poids des individus dans le champ d'enquête et ajusté selon la province, le groupe d'âge selon la liste des membres (15 à 19, 20 à 24, 25 à 44, 45 à 64, 65 ans et plus) et le sexe.

$$W_8 = W_7 * \left(\frac{\sum W_7 \ pour \ personnes \, répondantes + \sum W_7 \ pour \ personnes \, non - répondantes}{\sum W_7 \ pour \ personnes \, répondantes} \right)$$

12. Ajustement pour totaux externes

On a calculé un ajustement aux poids des personnes 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 pour chaque province-strate, et

2) Pour le cycle 1 et le cycle 2 :
   totaux de population selon la province, le sexe et les groupes d'âge suivants : 15 à 19, 20 à 24, 25 à 34, 35 à 44, 45 à 54, 55 à 64, et 65 ans et plus. La moyenne de ces totaux a été faite pour la période d'enquête.

   Pour le sommaire annuel :
   totaux de population selon la province, le sexe et les groupes d'âge suivants : 15 à 19, 20 à 24, 25 à 29, 30 à 34, 35 à 39, 40 à 44, 45 à 49, 50 à 54, 55 à 59, 60 à 64, 65 à 69 et 70 ans et plus. La moyenne de ces totaux a été faite pour la période d'enquête.

La méthode appelée GREG, soit la régression généralisée, a été utilisée 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 de la personne.

12.0 Questionnaire

Voir le fichier ESUTC2003_C1_QuestF.pdf pour le questionnaire français utilisé pour le Cycle 1 de 2003.

13.0 Clichés d'enregistrement à valeurs univariées

13.1 Cliché d'enregistrement à valeurs univariées – Fichier des ménages

Voir le fichier ESUTC2003_C1_MN_LvCds.pdf pour le cliché d'articles avec les variables uniques pour le Fichier des ménages du Cycle 1 de 2003.

13.2 Cliché d'enregistrement à valeurs univariées – Fichier des personnes

Voir le fichier ESUTC2003_C1_PR_LvCds.pdf pour le cliché d'articles avec les variables uniques pour le Fichier des personnes du Cycle 1 de 2003.