



Statistics Statistique  
Canada Canada

**ENQUÊTE SUR LE TABAGISME EN  
COLOMBIE-BRITANNIQUE  
ENQUÊTE SUR LA SANTÉ (ETCB)  
(2006)**

**FICHIER MAÎTRE  
GUIDE DE L'UTILISATEUR**

**Décembre 2008**

## TABLE DES MATIÈRES

<b>1.</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>2.</b>	<b>CONTEXTE.....</b>	<b>2</b>
2.1	LE MODÈLE DE DONNÉES.....	2
<b>3.</b>	<b>OBJECTIF.....</b>	<b>7</b>
<b>4.</b>	<b>CONTENU DE L'ENQUÊTE.....</b>	<b>7</b>
4.1	ESSAIS QUALITATIFS.....	7
4.2	ESSAIS SUR LE TERRAIN.....	8
4.3	CONTENU DU QUESTIONNAIRE FINAL.....	8
<b>5.</b>	<b>PLAN D'ÉCHANTILLONNAGE.....</b>	<b>9</b>
5.1	POPULATION CIBLE.....	9
5.2	BASE DE SONDAGE ET STRATÉGIE D'ÉCHANTILLONNAGE.....	9
5.3	RÉPARTITION DE L'ÉCHANTILLON SUR LA PÉRIODE DE COLLECTE DES DONNÉES.....	9
<b>6.</b>	<b>COLLECTE DES DONNÉES.....</b>	<b>10</b>
6.1	INTERVIEW ASSISTÉE PAR ORDINATEUR.....	10
6.2	DÉVELOPPEMENT DE L'APPLICATION DE L'ETCB.....	11
6.3	FORMATION DES INTERVIEWEURS.....	12
6.4	L'INTERVIEW.....	13
6.5	OPÉRATION SUR LE TERRAIN.....	14
6.6	CONTRÔLE DE LA QUALITÉ ET GESTION DE LA COLLECTE.....	14
<b>7.</b>	<b>TRAITEMENT DES DONNÉES.....</b>	<b>15</b>
7.1	VÉRIFICATION.....	15
7.2	CODAGE.....	15
7.3	CRÉATION DE VARIABLES DÉRIVÉES ET GROUPÉES.....	15
7.4	PONDÉRATION.....	15
<b>8.</b>	<b>PONDÉRATION.....</b>	<b>16</b>
8.1	PONDÉRATION DE L'ÉCHANTILLON.....	16
8.1.2	AJUSTEMENT DES CRITÈRES DE SÉLECTION.....	16
8.1.3	POSTSTRATIFICATION DES POIDS DE L'ESCC EN UTILISANT LES GROUPES D'ÂGE DE L'ETCB.....	17
8.1.4	NON-RÉPONSE À L'ETCB.....	17
8.1.5	POSTSTRATIFICATION.....	18
8.2	POIDS BOOTSTRAP.....	18
<b>9.</b>	<b>QUALITÉ DES DONNÉES.....</b>	<b>19</b>
9.1	TAUX DE RÉPONSE.....	19
9.2	ERREURS DANS LES ENQUÊTES.....	19
9.2.1	ERREURS NON DUES À L'ÉCHANTILLONNAGE.....	19
9.2.2	ERREURS D'ÉCHANTILLONNAGE.....	20
<b>10.</b>	<b>LIGNES DIRECTRICES POUR LA TOTALISATION, L'ANALYSE ET LA DIFFUSION.....</b>	<b>20</b>
10.1	LIGNES DIRECTRICES POUR L'ARRONDISSEMENT D'ESTIMATIONS.....	20
10.2	LIGNES DIRECTRICES POUR LA PONDÉRATION DE L'ÉCHANTILLON EN VUE DE LA TOTALISATION.....	21
10.2.1	DÉFINITIONS : ESTIMATIONS DE TYPE NOMINAL, ESTIMATIONS QUANTITATIVES.....	22
10.2.2	TOTALISATION D'ESTIMATIONS DE TYPE NOMINAL.....	23
10.2.3	TOTALISATION D'ESTIMATIONS QUANTITATIVES.....	23

10.3	LIGNES DIRECTRICES POUR L'ANALYSE STATISTIQUE .....	24
10.4	LIGNES DIRECTRICES POUR LA DIFFUSION .....	24
<b>11.</b>	<b>VARIANCES ET COEFFICIENTS DE VARIATION .....</b>	<b>26</b>
<b>12.</b>	<b>UTILISATION DU FICHIER .....</b>	<b>26</b>
12.1	FICHIER MAÎTRE .....	26
12.2	UTILISATION DE LA VARIABLE DE POIDS .....	26
12.3	RÈGLES D'AFFECTATIONS DES NOMS DES VARIABLES .....	27
12.3.1	STRUCTURE DES COMPOSANTES DU NOM DES VARIABLES DANS L'ETCB .....	27
12.3.2	STRUCTURE DES COMPOSANTES DU NOM DES VARIABLES DANS L'ESCC (À TITRE DE RÉFÉRENCE) .....	27
12.3.3	POSITIONS 1 À 3 : NOM DE LA VARIABLE/SECTION DU QUESTIONNAIRE .....	28
12.3.4	POSITION 4 : CYCLE ET NOM DE L'ENQUÊTE .....	28
12.3.5	POSITION 5 : TYPE DE VARIABLE .....	29
12.3.6	POSITIONS 6 À 8 : NOM DE LA VARIABLE .....	29
12.4	TOTALISATIONS PERSONNALISÉES .....	29



## 1. Introduction

L'Enquête sur le tabagisme en Colombie-Britannique (ETCB) est une enquête transversale conçue en vue de recueillir des renseignements sur les antécédents d'usage du tabac, les antécédents de mobilité et la propension à prendre des risques des résidents de la Colombie-Britannique. Il s'agit d'une enquête de suivi fondée sur le Cycle 3.1 (2005) de l'Enquête sur la santé dans les collectivités canadiennes (ESCC).

Les données de l'ETCB ont été recueillies auprès des :

- personnes de 18 ans et plus vivant dans des ménages privés occupés;
- Répondants non interviewés par personne interposée vivant en Colombie-Britannique au moment où l'interview du Cycle 3.1 de l'ESCC a été effectuée et ayant accepté le couplage de leurs renseignements.

Étaient exclues de l'ESCC les habitants des réserves indiennes et des terres de la Couronne, les résidents des établissements, les membres à temps plein des Forces canadiennes et les personnes vivant dans certaines régions éloignées. L'ETCB a été réalisée de février 2006 à mai 2006.

Le présent document a pour but de faciliter la manipulation du fichier maître de l'ETCB et de décrire la méthodologie utilisée.

Pour toute question concernant l'ensemble de données ou son utilisation, s'adresser à:

Totalisations spéciales ou renseignements généraux sur les données:

Services personnalisés à la clientèle, Division de la statistique de la santé:

Courriel:

613-951-1746

[hd-ds@statcan.ca](mailto:hd-ds@statcan.ca)

## 2. Contexte

L'objectif principal de l'Enquête sur le tabagisme en Colombie-Britannique (ETCB) était la collecte de renseignements sur l'historique d'usage du tabac et l'historique de mobilité des habitants de la Colombie-Britannique, ainsi que sur leur propension à prendre des risques. L'Enquête a été parrainée par le Ministère de la Santé de la Colombie-Britannique. L'ETCB était une enquête transversale de suivi du Cycle 3.1 de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) réalisé en 2005.

### 2.1 Le modèle de données

On a demandé aux répondants des renseignements sur ses antécédents d'usage du tabac à partir du moment où il a commencé à fumer jusqu'à l'heure actuelle (le moment de la tenue de l'enquête) ou jusqu'au moment où il aurait cessé de fumer.

La première question du module BCS demandait au répondant s'il avait fumé une centaine de cigarettes ou plus (environ quatre paquets) durant sa vie (BCSF\_100). Si le répondant répondait oui, on lui posait ensuite une série de questions sur :

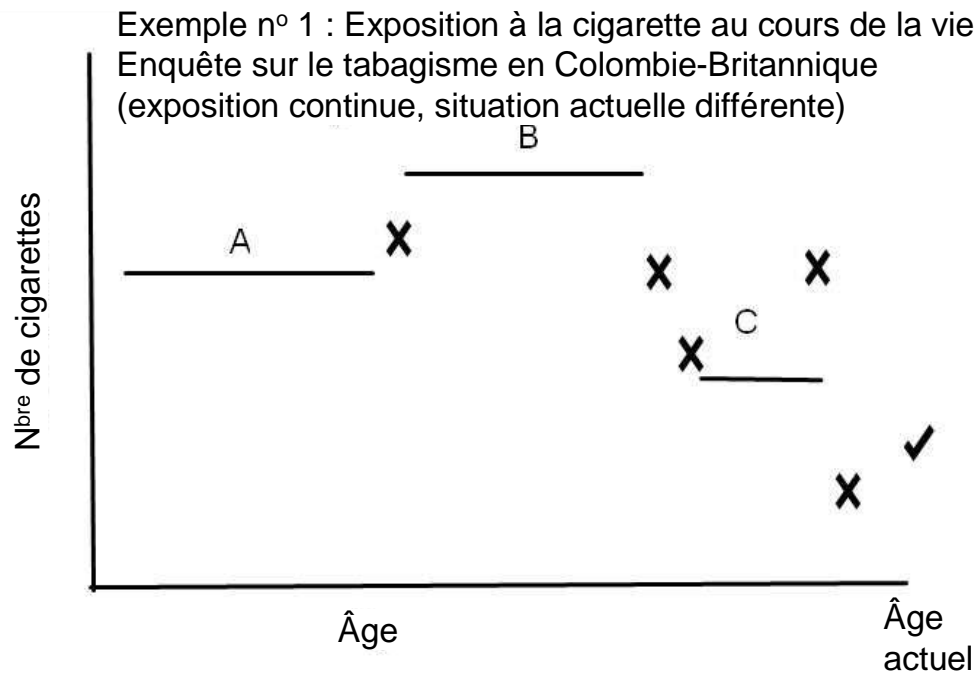
- l'âge auquel il a commencé à fumer au moins une cigarette par mois (BCSF\_102).
- le nombre de jours dans un mois où il fumait au moins une cigarette (BCSF\_103).
- combien de cigarettes il fumait habituellement chaque jour (BCSF\_104).

Après cette première série de questions, on a demandé au répondant s'il y a déjà eu une période d'au moins une année où il a arrêté de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique. Si le répondant n'avait jamais changé ses habitudes de consommation de cigarettes (répondu non à BCSF\_107), on ne continuait pas avec cette section du questionnaire. S'il avait répondu qu'il avait changé ses habitudes de consommation de cigarettes (répondu oui à BCSF\_107), on lui posait jusqu'à neuf fois une série de questions sur l'évolution de ses habitudes de consommation au cours d'une période donnée. Chacun des changements devait avoir duré au moins un an. Les neuf itérations possibles de cette section comprenaient les questions suivantes :

- y avait-il une période d'au moins un an où il a cessé de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique (BCSF\_107)
- son âge au moment où ses habitudes ont changé (BCSF\_108)
- combien de jours par mois il fumait au moins une cigarette (BCSF\_110)
- le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_111)

Dans les trois exemples ci-dessous, chaque lettre représente une période différente des antécédents d'usage du tabac du répondant.

- La lettre A représente la première série de questions dans le module BCS. Le répondant avait fumé 100 cigarettes ou plus durant sa vie (BCSF\_100 = OUI) et a donné une réponse indiquant l'âge auquel il a commencé à fumer et le nombre de cigarettes qu'il fumait habituellement un nombre de jours donné par mois pendant une période d'au moins un an (BCSF\_102, BCSF\_103, BCSF\_104).
- Les lettres B et C représentent des périodes où les habitudes de consommation de cigarettes ont changé. Si le répondant a répondu OUI à BCSF\_107, alors il y a eu une période d'au moins un an où il a arrêté de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique. On lui a ensuite posé des questions sur l'âge auquel ce changement s'est produit et le nombre de cigarettes qu'il fumait habituellement durant un nombre donné de jours par mois pendant une période d'au moins un an (BCSF\_108, BCSF\_110, BCSF\_111).
- La lettre X représente des périodes de moins d'un an où un changement a eu lieu. Aucune information n'a été recueillie pour ces périodes de changements.



Dans l'exemple n° 1, A, B et C représentent des périodes d'au moins un an où le répondant fumait un nombre donné de cigarettes pendant un nombre donné de jours par mois. Les questions ont été posées pour chaque période après la réponse initiale (A) jusqu'à un maximum de neuf périodes (B, C, etc.). Dans cet exemple, le répondant a fourni des renseignements sur trois périodes où ses habitudes de consommation de cigarettes ont changé et qui ont duré pendant au moins un an. Les X représentent les périodes de moins d'un an où la consommation de cigarettes était plus variable.

Dans cet exemple, la lettre A représente l'information suivante :

- L'âge auquel le répondant a commencé à fumer au moins une cigarette par mois (BCSF\_102) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_104).

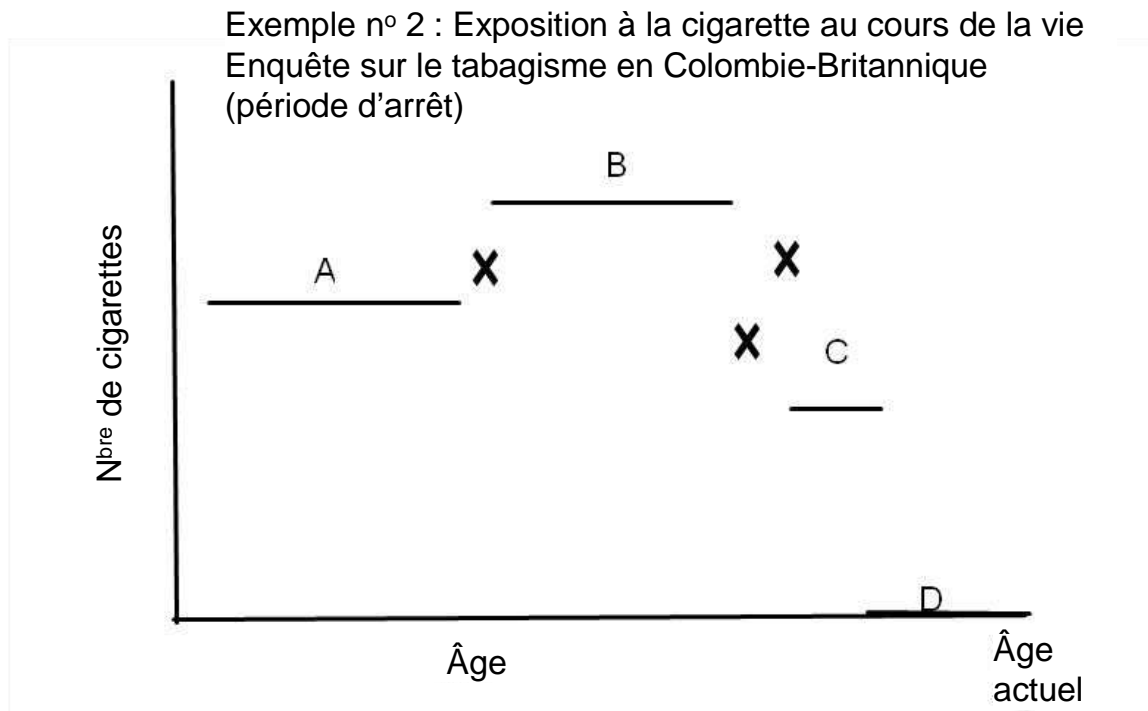
Les lettres B et C représentent l'information suivante :

- Une période d'au moins un an durant laquelle le répondant a cessé de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique (BCSF\_107 = OUI). Pour chaque période de changement des habitudes de consommation de cigarettes (B et C), l'âge auquel ses habitudes ont changé (BCSF\_108) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_111) sont illustrés ci-dessus.

Les secteurs dans le modèle marqués d'un X représentent un changement dans les habitudes de consommation de cigarettes qui a duré moins d'un an.

Le nombre de cigarettes fumées au moment de l'entrevue est représenté par une coche.





Dans l'exemple n° 2, A, B et C représentent des périodes d'au moins un an où le répondant fumait un nombre donné de cigarettes pendant un nombre donné de jours par mois. D représente une période d'au moins un an où le répondant a cessé de fumer (il ne fumait toujours pas au moment de l'entrevue). Les questions ont été posées pour chaque période après la réponse initiale (A) jusqu'à un maximum de neuf périodes (B, C, etc.). Dans cet exemple, le répondant a fourni de l'information pour quatre périodes correspondant à une évolution de ses habitudes de consommation de cigarettes qui a duré pendant au moins un an. Les X représentent des périodes de moins d'un an où la consommation de cigarettes était plus variable.

Dans cet exemple, la lettre A représente l'information suivante :

- L'âge auquel il a commencé à fumer au moins une cigarette par mois (BCSF\_102) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_104).

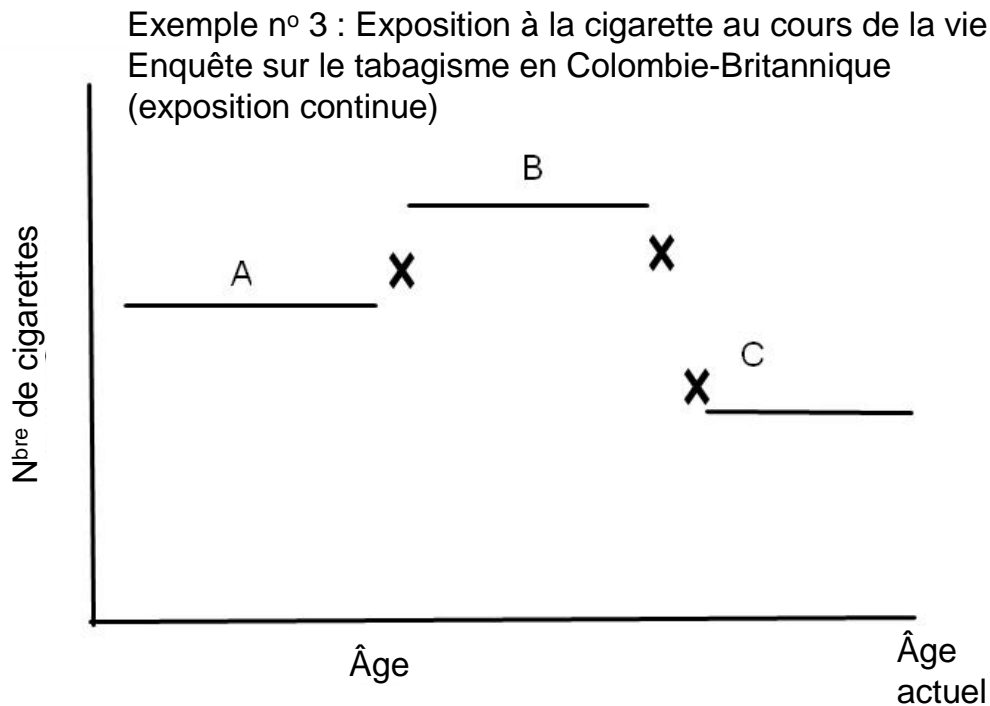
Les lettres B et C représentent l'information suivante :

- Une période d'au moins un an où il a cessé de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique (BCSF\_107 = OUI). Pour chaque période de changement de ses habitudes de consommation de cigarettes (B et C), l'âge auquel ses habitudes ont changé (BCSF\_108) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_111) sont illustrés ci-dessus.

La lettre D représente l'information suivante :

- La période où le répondant avait cessé de fumer pendant au moins un an (BCSF\_107).

Le secteur dans le modèle marqué d'un X représente un changement dans les habitudes de consommation de cigarettes qui a duré moins d'un an.



Dans l'exemple n° 3, A, B et C représentent des périodes d'au moins un an où le répondant fumait un nombre donné de cigarettes pendant un nombre donné de jours par mois. Des questions ont été posées pour chaque période après la réponse initiale (A) jusqu'à un maximum de neuf périodes (B, C, etc.). Dans cet exemple, le répondant a fourni de l'information pour trois périodes correspondant à une évolution de ses habitudes de consommation de cigarettes qui a duré pendant au moins un an. Les X représentent des périodes d'au moins un an où la consommation de cigarettes était plus variable.

Dans cet exemple, la lettre A représente l'information suivante :

- L'âge auquel il a commencé à fumer au moins une cigarette par mois (BCSF\_102) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_104).

Les réponses en B et C représentent l'information suivante :

- Une période d'au moins un an où il a cessé de fumer, changé le nombre de cigarettes qu'il fumait ou changé le nombre de jours où il fumait durant un mois typique (BCSF\_107 = OUI). Pour chaque période de changement dans ses habitudes de consommation de cigarettes (B et C), l'âge auquel ses habitudes ont changé (BCSF\_108) et le nombre de cigarettes qu'il fumait habituellement chaque jour (BCSF\_111) sont illustrés ci-dessus.

Le secteur dans le modèle marqué d'un X représente un changement dans les habitudes de consommation de cigarettes qui a duré moins d'un an.

### **3. Objectif**

L'objectif principal de l'Enquête sur le tabagisme en Colombie-Britannique (ETCB) était la collecte de renseignements sur l'historique d'usage du tabac et l'historique de mobilité des habitants de la Colombie-Britannique, ainsi que sur leur propension à prendre des risques.

### **4. Contenu de l'enquête**

La présente section décrit le processus général de consultation suivi pour élaborer le contenu de l'enquête et résume le contenu final choisi pour l'ETCB.

Le contenu a été élaboré par la Division de la statistique de la santé en s'appuyant sur les objectifs et les exigences énoncés par le client. Celui-ci a été consulté tout au long de l'élaboration et de la mise à l'essai du questionnaire.

#### **4.1 Essais qualitatifs**

Deux essais qualitatifs ont été réalisés pour évaluer le contenu et l'enchaînement des questions du questionnaire de l'ETCB. Le premier a eu lieu en novembre 2004 à Ottawa, en Ontario. Environ 25 répondants, formant un groupe représentatif de fumeurs et d'anciens fumeurs, y compris des personnes fumant tous les jours et à l'occasion, ont participé à des interviews sur place. La base de sondage utilisée pour sélectionner ces répondants était l'échantillon du Cycle 2.1 (2003) de l'ESCC. Toutes les interviews qualitatives ont été réalisées par des représentants du Centre d'information sur la conception des questionnaires (CICQ) et observées par l'équipe du projet de l'ETCB de la Division de la statistique de la santé.

L'une des constatations importantes faites durant l'essai de novembre était que les répondants avaient de la difficulté à se rappeler leurs divers profils d'usage du tabac au cours du temps. Afin de faciliter la collecte des données sur ces antécédents, il a été déterminé que le fait d'établir un schéma chronologique linéaire dans le questionnaire aiderait les répondants à se souvenir de leurs périodes d'usage du tabac. Par exemple, au début de la section destinée à recueillir les antécédents d'usage du tabac, on a demandé aux répondants quand ils avaient fumé leur première cigarette, puis on a posé des questions fondées sur les faits pour déterminer les périodes successives d'usage du tabac ou de renoncement au tabac (c.-à-d., le schéma chronologique linéaire).

Un deuxième et dernier test qualitatif a été réalisé en février 2005 à Vancouver, en Colombie-Britannique. Environ 25 répondants, formant un groupe représentatif d'anciens fumeurs et de fumeurs, y compris des personnes fumant tous les jours et à l'occasion, ont participé à des interviews sur place. La base de sondage utilisée pour sélectionner ces répondants était l'échantillon du Cycle 2.1 (2003) de l'ESCC. Toutes les interviews qualitatives ont été réalisées par des représentants du Centre d'information sur la conception des questionnaires (CICQ) et observées par l'équipe du projet de l'ETCB de la Division de la statistique de la santé.

Le résultat principal était que l'approche du schéma chronologique donnait de bons résultats et que la plupart des répondants la considéraient comme un outil très utile pour rassembler leurs

souvenirs sur les moments où ils avaient fumé, et quelles quantités, au cours de diverses périodes de leur vie.

## 4.2 Essais sur le terrain

En novembre 2005, on a procédé à l'essai pilote de l'ETCB. La base de sondage utilisée pour sélectionner environ 300 répondants pour les interviews était l'échantillon de répondants du Cycle 2.1 (2003) de l'ESCC. L'objectif principal était de déterminer si le contenu élaboré et testé avec succès durant les essais qualitatifs (interviews sur place) au mois de février précédent pourrait aussi être administré avec succès par interview téléphonique. Les données recueillies durant l'essai sur le terrain ont indiqué qu'aussi bien les répondants que les intervieweurs estimaient que l'ETCB se prêtait bien à l'administration du questionnaire par interview téléphonique.

## 4.3 Contenu du questionnaire final

Cette section du guide décrit les modules de la composante C2 (Questionnaire) de l'ETCB et fournit des renseignements sur la pertinence de ces modules. Le questionnaire C2 de l'ETCB a été subdivisé en quatre modules, destinés chacun à recueillir un type particulier d'information:

<b>Module</b>	<b>Contenu</b>
<b>BCS</b>	<b>Historique de l'usage du tabac en Colombie-Britannique/exposition à la cigarette au cours de la vie :</b> Ce module était conçu pour déterminer si les répondants avaient fumé ou non une centaine de cigarettes ou plus au cours de leur vie, ainsi que le schéma chronologique des quantités consommées, de la fréquence de consommation ainsi que des marques fumées. Les questions du module ont été posées à tous les répondants.
<b>BCM</b>	<b>Mobilité :</b> Ce module avait pour but de recueillir des renseignements sur la durée de résidence des répondants à l'intérieur et à l'extérieur de la Colombie-Britannique. Les questions du module n'ont été posées qu'aux répondants qui étaient des anciens fumeurs et/ou des fumeurs au moment de l'enquête.
<b>BCD</b>	<b>Sécurité et conduite :</b> Ce module visait à recueillir des renseignements sur les habitudes de conduite d'un véhicule et sur les mesures de sécurité prises par les conducteurs ainsi que par les passagers d'un véhicule à moteur. Les questions de ce module ont été posées à tous les répondants.
<b>BCA</b>	<b>Administration :</b> Ce module a permis de recueillir des renseignements administratifs, y compris la permission de procéder au couplage de l'information recueillie avec les données des cycles antérieurs de l'ESCC et des dossiers de santé administratifs de la province, le numéro valide de cartes-santé et la langue de l'interview. Les questions de ce module ont été posées à tous les répondants.

## **5. Plan d'échantillonnage**

### **5.1 Population cible**

La population cible comprend les résidents de la Colombie-Britannique de 18 ans et plus qui vivaient dans des logements privés occupés au moment de l'interview du Cycle 3.1 de l'ESCC en 2005. Les habitants des réserves indiennes et des terres de la Couronne, les personnes vivant en établissement, les membres à temps plein des Forces canadiennes et les résidents de certaines régions éloignées ne font pas partie du champ d'observation de l'enquête.

### **5.2 Base de sondage et stratégie d'échantillonnage**

La base de sondage pour le tirage de l'échantillon de l'ETCB a été créée en utilisant les répondants au Cycle 3.1 (2005) de l'Enquête sur la santé dans les collectivités canadiennes (ESCC). Comme les données de l'ESCC devraient être couplées aux données de l'ETCB, seules les personnes ayant consenti au couplage des données qu'elles avaient fournies au Cycle 3.1 de l'ESCC ont été incluses dans la base de sondage de l'ETCB. En outre, seuls les répondants à l'ESCC âgés de 18 ans et plus et vivant en Colombie-Britannique au moment de l'interview du Cycle 3.1 ont été contactés pour l'ETCB. Aucune personne ayant répondu par personne interposée au Cycle 3.1 de l'ESCC n'a été incluse dans la base de sondage de l'ETCB, à cause de la nature personnelle et des demandes de remémoration du questionnaire de cette enquête. Dans le cas de l'ESCC, les interviews par personnes interposées sont permises dans les situations où le répondant n'a pas la capacité de participer lui-même à l'interview à cause de limitations physiques ou mentales. Enfin, toutes les personnes n'ayant participé que partiellement à l'interview du Cycle 3.1 de l'ESCC ont été exclues de la base de sondage de l'ETCB, et ce pour deux raisons. Premièrement, la question concernant la permission de coupler les données du Cycle 3.1 de l'ESCC a été posée dans le dernier module du questionnaire et la plupart des interviews partielles n'ont pas atteint ce stade. Deuxièmement, puisque les données de l'ESCC allaient être couplées à celles de l'ETCB, il n'était pas souhaitable que de grandes proportions de données manquent. Toutes les unités constituant la base de sondage de l'ETCB sont incluses dans l'échantillon de cette enquête.

Puisqu'il s'agissait d'une enquête auprès de répondants ciblés, seul le membre particulier d'un ménage qui avait répondu entièrement à l'interview de Cycle 3.1 de l'ESCC était admissible au suivi pour l'ETCB.

### **5.3 Répartition de l'échantillon sur la période de collecte des données**

En Colombie-Britannique, l'échantillon du Cycle 3.1 de l'ESCC provenait de 16 régions sociosanitaires (RSS) couvrant la province complète et a été réparti proportionnellement à la racine carrée de la taille de la population de la RSS. Chaque RSS comptait de 500 à 1 600 répondants. La répartition de l'échantillon de l'ESCC selon l'âge et le sexe était très semblable à la répartition réelle dans la province, à l'exception d'une légère sur-représentation des personnes de 65 ans et plus.

La période de collecte des données de l'ETCB s'est étendue de février 2006 à mai 2006. Pendant cette période, les données ont été recueillies auprès de deux échantillons. Un premier échantillon d'environ 5 900 enregistrements, qui représentaient les répondants les plus admissibles ayant participé à l'interview du Cycle 3.1 de l'ESCC de janvier 2005 à juin 2005, a été affecté aux intervieweurs à la fin de janvier 2006. Un deuxième échantillon d'environ 5 700 enregistrements, comprenant tous les autres répondants au Cycle 3.1 de l'ESCC admissibles, a été affecté aux intervieweurs à la fin de mars 2006. En tout, 11 587 enregistrements ont été fournis en vue d'une interview éventuelle dans le cadre de l'ETCB.

Les renseignements pour la prise de contact n'étaient pas disponibles pour tous les répondants au Cycle 3.1 de l'ESCC. Certains de ceux faisant partie de la base de sondage aréolaire ne possédaient pas, ou n'avaient pas fourni un numéro de téléphone valide. Comme l'ETCB a été réalisée par téléphone, tous les répondants pour lesquels aucun numéro de téléphone n'était disponible ont été désignés pour le dépistage. Différentes sources ont été utilisées pour essayer de trouver les numéros de téléphone et pour prendre contact avec les personnes sélectionnées pour participer à l'ETCB. Il convient de souligner que le dépistage visait non seulement les répondants sans numéro de téléphone, mais aussi ceux qui avaient déménagé entre le moment de l'interview de l'ESCC et celui de l'interview de l'ETCB. En plus de ces cas posant problème, certains répondants avaient refusé de donner leur vrai nom durant l'interview de l'ESCC et un nom fictif avait été utilisé pour les identifier dans cette enquête. Afin d'aider les intervieweurs à trouver les personnes qu'ils recherchaient, la liste complète des répondants ayant un nom fictif, y compris tous les membres de leur ménage, leur a été fournie. Pour chaque membre de ces ménages, l'âge, la relation avec le répondant, le sexe et l'état matrimonial ont été indiqués. Ces cas particuliers ont été affectés à des intervieweurs chevronnés, étant donné qu'ils étaient plus difficiles à traiter.

## **6. Collecte des données**

### **6.1 Interview assistée par ordinateur**

La collecte des données de l'ETCB a eu lieu de février à mai 2006. Durant cette période, 9 316 interviews valides ont été réalisées par interview téléphonique assistée par ordinateur (ITAO).

L'interview assistée par ordinateur (IAO) offre de grands avantages par rapport aux autres méthodes de collecte des données. D'abord, la technique est étayée d'un système de gestion des cas et d'une fonctionnalité de transmission de données. Le système de gestion des cas enregistre automatiquement l'information de gestion importante sur chaque tentative visant un cas particulier et produit des rapports aux fins de la gestion du processus de collecte. L'IAO comporte aussi un ordonnanceur automatique d'appels, c'est-à-dire un système central qui optimise l'horaire des rappels et le calendrier des rendez-vous utilisés pour appuyer la collecte par ITAO.

Le système de gestion des cas achemine les applications de questionnaire et les fichiers d'échantillons du Bureau central de Statistique Canada aux bureaux régionaux de collecte (dans le cas de l'ITAO). Les données destinées au Bureau central sont acheminées en sens inverse. Par

souci de confidentialité, les données sont encodées avant la transmission. Elles sont ensuite désencodées une fois sauvegardées sur un ordinateur sécurisé distinct, sans accès à distance.

Deuxièmement, l'IAO permet de concevoir une interview personnalisée à l'intention de chaque répondant en fonction de leurs caractéristiques particulières et de leurs réponses à l'enquête.

Notamment, l'application :

- saute automatiquement les questions qui ne s'appliquent pas au répondant;
- applique automatiquement des règles de vérification pour repérer les réponses incohérentes ou non incluses dans la fourchette de valeurs permises et affiche des messages-guides à l'écran quand une entrée est invalide, ce qui permet au répondant de recevoir une rétroaction immédiate et à l'intervieweur, de corriger toute incohérence;
- personnalise automatiquement le libellé des questions, y compris les périodes de référence et les pronoms, en fonction de facteurs comme l'âge et le sexe du répondant, la date de l'interview et les réponses aux questions précédentes.

## **6.2 Développement de l'application de l'ETCB**

Une application d'interview téléphonique assistée par ordinateur (ITAO) a été utilisée pour réaliser l'ETCB. Au moment de son développement, une nouvelle application standardisée pour les composantes/procédures d'entrée et de sortie dans le cas de répondants ciblés venait d'être créée à Statistique Canada. Une certaine personnalisation des composantes Entrée et Sortie a été permise pour l'ETCB. L'application comportait les composantes Entrée, C2 (contenu de l'enquête) et Sortie.

Les composantes Entrée et Sortie comprennent des séries standard de questions conçues pour permettre à l'intervieweur de prendre contact avec le répondant, de confirmer la participation de ce dernier, de procéder au dépistage (au besoin) et de déterminer l'état du cas. La composante C2 comprend les modules proprement dits de l'ETCB qui constituent la partie principale de l'application. L'essai pilote de l'application et du contenu de l'enquête a été effectué en novembre 2005.

L'application ITAO a été modifiée ou améliorée sur la base des résultats de l'essai pilote. Une fois ces modifications effectuées, la mise à l'essai finale de l'application complète a débuté. Elle s'est déroulée en trois étapes d'essai interne, à savoir l'essai modulaire, l'essai intégré et l'essai de bout-en-bout.

L'essai modulaire consiste à mettre à l'essai indépendamment chaque module de contenu afin de vérifier que les instructions « passez à », la logique d'enchaînement et le texte, dans les deux langues officielles, sont spécifiés correctement. Les instructions « passez à » et la logique d'enchaînement entre modules ne sont pas testées à cette étape, car chaque module est considéré comme un questionnaire autonome. Quand plusieurs responsables des essais ont vérifié tous les modules, ces derniers sont regroupés avec les composantes Entrée et Sortie pour former une application intégrée. Cette nouvelle application intégrée passe alors à l'étape suivante des essais.

L'essai intégré porte sur l'ensemble des modules testés à l'étape précédente, regroupés en applications intégrées avec les composantes Entrée et Sortie. Cette deuxième étape des essais a pour but de s'assurer que des renseignements clés, par exemple l'âge et le sexe, sont transmis du fichier d'échantillon aux composantes C2 et Sortie des applications. Elle confirme également que les variables qui influent sur les instructions « passez à » et la logique d'enchaînement sont transmises correctement de module en module à l'intérieur de la composante C2. Étant donné qu'à ce moment, le fonctionnement des applications est essentiellement identique à ce qu'il sera sur le terrain, tous les scénarios possibles auxquels feront face les intervieweurs sont simulés afin d'assurer la fonctionnalité rigoureuse. Ces scénarios servent à tester divers aspects des composantes Entrée et Sortie, y compris la prise de contact, la confirmation que le répondant correct a été découvert, la question de savoir si un cas fait partie du champ de l'enquête et la prise de rendez-vous.

L'essai de bout en bout situe l'application entièrement intégrée dans un environnement de collecte simulé. L'application est chargée dans des ordinateurs connectés à un serveur d'essai. Ensuite, des données sont recueillies, transmises et extraites en temps réel, comme cela se ferait sur le terrain. Cette dernière étape des essais permet d'expérimenter tous les aspects techniques de la saisie, de la transmission et de l'extraction des données pour chacune des applications de l'ETBC. Il s'agit aussi de la dernière occasion de déceler des erreurs dans les composantes Entrée, C2 et Sortie.

### **6.3 Formation des intervieweurs**

Une équipe de représentants du Bureau central de Statistique Canada s'est rendue à Edmonton en février 2006 pour donner une séance de formation au chef de projet du bureau régional et à une petite équipe d'intervieweurs sélectionnés pour l'enquête. Les membres de l'équipe de l'enquête ont participé à ces séances de formation afin d'offrir un soutien supplémentaire et de répondre à toutes questions ou préoccupations qui auraient pu être soulevées.

L'objet de ces séances était de permettre aux intervieweurs d'être confortable avec l'utilisation de l'application de l'ETCB, ainsi que de familiariser ceux-ci avec le contenu. La formation était centrée sur :

- les buts et objectifs de l'enquête
- la méthodologie de l'enquête;
- les fonctionnalités des applications;
- l'examen du contenu du questionnaire, y compris des exercices;
- la simulation d'interviews difficiles et de situations de non-réponse;
- la gestion de l'enquête.

La formation visait en priorité à réduire au minimum les cas de non-réponse. À cette fin, les intervieweurs ont participé à des exercices qui consistaient à persuader des répondants réticents de participer à l'enquête.



## **6.4 L'interview**

Les unités d'échantillonnage sélectionnées dans la base de sondage ont été interviewées par la méthode d'ITAO à partir d'une centrale d'appels. Un intervieweur principal affecté à la même centrale d'appels assurait la surveillance des intervieweurs.

Afin de s'assurer de la qualité des données recueillies, les intervieweurs ont reçu l'instruction d'essayer, par tous les moyens possibles, de mener l'interview avec le répondant sélectionné en privé. Dans les situations où cela a été impossible, le répondant a été interviewé et des indicateurs ont été enregistrés dans le fichier maître pour signaler que l'intervieweur avait établi qu'une autre personne que le répondant était présente durant l'interview (BCAF\_N10) et indiquer s'il estimait que les réponses du répondant avaient été influencées par la présence de cette autre personne (BCAF\_N11).

De nombreuses techniques, y compris les suivantes, ont été mises en œuvre afin de parvenir à un taux de réponse optimal.

### **Lettre d'introduction**

Avant le début de chacune des deux périodes de collecte, les ménages échantillonnés ont reçu des lettres d'introduction qui expliquaient l'objet de l'enquête. Elles énonçaient notamment l'importance de l'enquête et offraient des exemples de l'utilisation prévue des données tirées de l'ETCB.

### **Prise de contact**

Les intervieweurs ont reçu l'instruction de faire toutes les tentatives raisonnables pour obtenir les interviews. Lorsque l'appel avait eu lieu à un moment inopportun, l'intervieweur fixait le moment d'un rappel qui convenait au répondant. De nombreux rappels ont été effectués, à divers moments de la journée et à des jours différents.

### **Conversion des cas de refus**

Aux personnes qui refusaient dès le premier contact de participer à l'enquête, le bureau régional de Statistique Canada envoyait une lettre insistant sur l'importance de l'enquête et de la collaboration du ménage. Ensuite, un intervieweur principal, un surveillant de projet ou un autre intervieweur rappelait le répondant pour essayer de le convaincre de l'importance de sa participation à l'enquête.

### **Obstacles linguistiques**

Pour parer au problème de langue susceptible de nuire aux interviews, le bureau régional de Statistique Canada a embauché des intervieweurs qui parlaient un grand nombre de langues. Au besoin, les cas étaient transférés à un intervieweur capable de remplir le questionnaire dans la

langue voulue. Le chinois et le punjabi sont les deux langues qui, selon le bureau régional d'Edmonton, ont posé le plus souvent des difficultés.

### **Interviews par procuration**

Les interviews par procuration n'étaient pas permises.

### **6.5 Opération sur le terrain**

L'échantillon de l'ETCB était réparti en deux périodes de collecte des données chevauchantes de deux mois. Le bureau régional de collecte des données a reçu l'instruction d'utiliser les quatre premières semaines de chaque période de collecte pour interviewer la majorité de l'échantillon, puis de consacrer les quatre semaines suivantes aux interviews restantes et au suivi des cas de non-réponse. À la deuxième semaine de chaque période de collecte, des tentatives devaient avoir été effectuées relativement à tous les cas.

Il incombait au superviseur du projet, à l'intervieweur principal et à l'équipe de soutien technique du bureau régional d'Edmonton de s'occuper de la transmission des cas au Bureau central. Tous les cas achevés ont été transmis chaque nuit au Bureau central de Statistique Canada.

### **6.6 Contrôle de la qualité et gestion de la collecte**

Durant le cycle de collecte des données de l'ETCB, plusieurs méthodes ont été appliquées pour s'assurer de la qualité des données et de l'optimisation de la collecte. Elles comportaient, entre autres, des mesures internes de vérification du rendement de l'intervieweur et une série de rapports de contrôle des diverses cibles de collecte et de la qualité des données.

Les intervieweurs par ITAO ont été sélectionnés au hasard pour la validation. Cette dernière comportait la surveillance des interviews téléphoniques par des intervieweurs principaux en vue de s'assurer que l'intervieweur suivait les techniques et procédures prévues (c.-à-d. qu'il lisait le libellé des questions tel qu'il figurait dans les applications, qu'il ne posait pas de question pour susciter des réponses, etc.).

Les responsables ont produit une série de rapports dont l'objet était de contrôler et de gérer efficacement les objectifs de collecte et de mettre au jour les problèmes posés par la collecte.

Des rapports cumulatifs précisant les taux de réponse et de couplage ont été produits à la fin de chaque période. À la fin de la collecte des données, le taux de réponse était de 80 %. En tout, 87 % de répondants à l'ETCB ont donné leur permission pour le couplage de leurs données, tandis que 72 % ont donné la permission pour le couplage et ont déclaré un numéro d'assurance-santé valide. Par ailleurs, des rapports personnalisés ont été établis et utilisés pour examiner des questions particulières de qualité des données qui se sont manifestées au cours de la collecte.

## **7. Traitement des données**

### **7.1 Vérification**

La vérification des données a été exécutée en grande partie durant la collecte des données par l'application d'interview assistée par ordinateur (IAO). Les intervieweurs ne pouvaient pas entrer de valeurs hors-normes et les erreurs d'enchaînement faisaient l'objet de l'instruction de contrôle programmée « passez à ». Par exemple, l'IAO s'assurait de ne pas poser au répondant les questions non pertinentes.

En réponse à certaines données incompatibles ou inhabituelles, l'application a émis des messages d'avertissement, mais aucune mesure correctrice n'a été prise au moment de l'interview. On a plutôt mis au point, au besoin, des contrôles destinés à être appliqués au Bureau central après la collecte des données. Les incohérences ont été le plus souvent corrigées en attribuant à l'une ou aux deux variables en question la valeur « non déclaré ».

### **7.2 Codage**

Des catégories de réponses précodées ont été fournies pour toutes les variables appropriées. Les intervieweurs ont reçu une formation durant laquelle ils ont appris à classer les réponses recueillies dans la catégorie appropriée.

### **7.3 Création de variables dérivées et groupées**

Afin de faciliter l'analyse des données et de réduire au minimum le risque d'erreur, deux variables du fichier ont été dérivées en utilisant des items figurant dans le questionnaire de l'ETCB. En général, le cinquième caractère du nom d'une variable dérivée est un « D ». Dans les deux cas, une ou plusieurs variables ont été combinées pour en créer une nouvelle. Pour des renseignements sur les règles d'affectation des noms, consulter la section 12.3.

### **7.4 Pondération**

Le principe qui sous-tend l'estimation fondée sur un échantillon probabiliste tel que celui de l'ETCB est que chaque personne comprise dans l'échantillon « représente », outre elle-même, plusieurs autres personnes qui n'y figurent pas. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne en représente 50. Dans la terminologie en usage ici, nous dirons que nous avons attribué à chaque personne un facteur de pondération de 50.

L'étape de la pondération est celle où l'on calcule le poids d'échantillonnage de chaque personne échantillonnée. Ce poids figure dans le fichier de données et doit être utilisé pour calculer des estimations significatives d'après les données de l'enquête. Par exemple, si l'on doit évaluer le nombre de personnes qui fument tous les jours, on le fait en choisissant dans l'échantillon les enregistrements des personnes qui présentent cette caractéristique et en faisant la somme des facteurs de pondération appliqués à ces enregistrements.

Des précisions sur la façon de calculer les poids d'échantillonnage figurent à la section 8.

## 8. Pondération

Pour que les estimations produites à partir de données d'enquête soient représentatives de la population étudiée, et non pas seulement de l'échantillon proprement dit, l'utilisateur doit incorporer les facteurs de pondération, appelés ici les poids de sondage, dans ses calculs. Un poids de sondage est attribué à chaque personne incluse dans l'échantillon final, c'est-à-dire l'échantillon de personnes ayant répondu à l'enquête. Ce poids correspond au nombre de membres de la population représentés par le répondant.

Comme il est décrit à la section 5, la base de sondage de l'ETCB est fondée sur les répondants au Cycle 3.1 (2005) de l'ESCC. Le point de départ du processus de pondération de l'ETCB est le poids principal de l'ESCC avant poststratification. Pour plus de renseignements sur ce poids, consulter le guide de l'utilisateur des données du Cycle 3.1 de l'ESCC.

### 8.1 Pondération de l'échantillon

Le tableau 8.1 donne un aperçu des divers ajustements qui font partie de la stratégie de pondération de l'ETCB, dans l'ordre où ils sont appliqués.

**Tableau 8.1 : Étape de pondération pour l'ETCB**

BC1 – Ajustement des critères de sélection
BC2 – Poststratification des poids de l'ESCC en utilisant les groupes d'âge de l'ETCB
BC3 – Non-réponse à l'ETCB
BC4 – Poststratification finale

#### 8.1.2 Ajustement des critères de sélection

Puisque l'ETCB est une enquête supplémentaire fondée sur le Cycle 3.1 (2005) de l'ESCC, le point de départ du processus de pondération a été le poids principal de l'ESCC avant la poststratification. La première étape consistait à éliminer du fichier des poids de l'ESCC toutes les unités de l'ESCC ne résidant pas en Colombie-Britannique, puisqu'elles n'étaient représentatives d'aucun membre de la population cible de l'ETCB. Étant donné que le plan de l'ESCC était stratifié dans chaque province et que chaque ajustement de l'ESCC a été exécuté dans la province, les poids de l'ESCC pour les répondants résidant en dehors de la Colombie-Britannique étaient indépendants de ceux des répondants vivant en Colombie-Britannique, ce qui a permis de supprimer simplement ces unités du fichier.

L'étape suivante consistait à supprimer tous les cas d'interviews par personne interposée et d'interviews partielles de l'ESCC, ainsi que les répondants qui n'avaient pas autorisé le couplage des données qu'ils avaient fournies. Pour ce groupe de répondants à l'ESCC qui ont été exclus de la base de sondage de l'ETCB, le poids de l'ESCC initial représentaient d'autres personnes vivant

en Colombie-Britannique. Par conséquent, on a réparti leurs poids entre les répondants restants de l'ESCC. Pour cela, on a créé des groupes de réponses homogènes (GRH) en se servant du logiciel KnowledgeSeeker<sup>1</sup>.

L'ajustement a été calculé dans chaque GRH, de la façon suivante :

$$adjBC1 = \frac{\text{Somme des poids avant poststratification de l'ESCC pour toutes les unités}}{\text{Somme des poids pour toutes les interviews non réalisées par personne interposée, achevées avec permission de coupler les données de l'ESCC}}$$

Le poids wgtBC1 a été calculé sous la forme  $wgti2b * adjBC1$ , où wgti2b est le poids principal avant poststratification de l'ESCC. Après le calcul de l'ajustement, les unités exclues ont été éliminées du fichier et l'ajustement a été appliqué aux unités restantes.

### 8.1.3 Poststratification des poids de l'ESCC en utilisant les groupes d'âge de l'ETCB

À ce stade, les seules unités encore dans le fichier de données et non incluses dans la base de sondage de l'ETCB étaient les répondants à l'ESCC de moins de 18 ans. Avant d'éliminer ces unités du fichier, les données ont été poststratifiées en utilisant les projections démographiques provinciales de 2005 pour deux groupes d'âge, à savoir les 12 à 17 ans et les 18 ans et plus. Les projections démographiques étaient fondées sur le Recensement de 2001. La moyenne des estimations mensuelles de 2005 pour chaque groupe d'âge a servi de total connu. L'ajustement a été calculé dans chaque poststrate de la façon suivante :

$$adjBC2 = \frac{\text{Estimation démographique pour la province : groupe d'âge du répondant}}{\text{Somme des wgtBC1 pour la province : groupe d'âge du répondant}}$$

Le poids wgtBC2 a été calculé sous la forme  $wgtBC1 * adjBC2$ . Après la poststratification, les répondants à l'ESCC de moins de 18 ans ont été éliminés du fichier. Donc, les unités restantes représentaient l'échantillon complet de l'ETCB et les poids étaient représentatifs de la population cible de l'enquête.

### 8.1.4 Non-réponse à l'ETCB

Bien que la population cible de l'ETCB comprenait tous les répondants à l'ESCC, ces personnes n'ont pas toutes répondu à l'enquête. À la présente étape, les poids des unités non répondantes à l'ETCB ont été répartis entre les unités répondantes. Cela comporte la répartition des cas en double de l'ESCC, qui ont été traités comme des non-répondants dans l'ETCB. Le fichier de l'échantillon de l'ETCB contenait les cas que l'on savait être des doubles dans l'ESCC. Ces cas en double ont eu lieu quand un répondant a été choisi dans la base aréolaire ainsi que dans la base de sondage téléphonique de l'ESCC et qu'il a décidé de répondre deux fois à l'enquête. Souvent, ces personnes ont donné de réponses différentes aux deux interviews et celles-ci ont été réalisées à des périodes différentes de l'année. Par conséquent, on a retenu les deux versions dans le fichier du Cycle 3.1 de l'ESCC et on leur a attribué des poids distincts.

---

<sup>1</sup> Angoss Software, 2005

Ces cas en double de l'ESCC ont été découverts en se servant de renseignements personnels, comme le nom, l'âge, la date de naissance, le numéro de téléphone et le numéro d'assurance-santé. Afin de réduire au minimum le fardeau de réponse, il n'a été pris contact avec ces personnes qu'une seule fois pour l'ETCB et leur enregistrement en double ont été traités comme une non-réponse. L'ETCB ne comportait pas d'interview partiellement achevée, si bien que seules les unités pour lesquelles l'interview était complète ont été considérées comme des répondants.

Pour cette étape, la non-réponse a été modélisée en se servant de tests du Khi-deux en utilisant le logiciel KnowledgeSeeker. Des GRH ont été créés en utilisant des variables auxiliaires pour lesquelles des données avaient été recueillies dans l'ESCC ainsi que l'ETCB. L'ajustement a été calculé dans chaque GRH, de la façon suivante :

$$adjBC3 = \frac{\text{Somme des wgtBC2 pour toutes les unités}}{\text{Somme des wgtBC2 pour toutes les unités répondantes}}$$

Cet ajustement a été appliqué à wgtBC2 pour obtenir wgtBC3. Après le calcul de l'ajustement, les unités non répondantes (et les cas en double de l'ESCC) ont été éliminés du fichier et l'ajustement a été appliqué aux unités restantes.

### 8.1.5 Poststratification

L'étape finale de l'obtention du poids principal de l'ETCB était la poststratification afin de projeter les chiffres de population de 2005 au niveau provincial. Cette poststratification avait pour but de s'assurer que la somme des poids finals correspondait aux projections démographiques définies au niveau provincial pour huit groupes d'âge, à savoir 18 à 29 ans, 30 à 44 ans, 45 à 64 ans et 65 ans et plus, selon le sexe. Les projections démographiques étaient basées sur les chiffres du Recensement de 2001. La moyenne arrondie des estimations démographiques mensuelles de 2005 pour chaque groupe âge-sexe a été utilisée comme total connu. L'ajustement a été calculé dans chaque poststrate, de la façon suivante :

$$adjBC4 = \frac{\text{Estimation démographique pour la province : groupe âge-sexe du répondant}}{\text{Somme des wgtBC3 pour la province : groupe âge-sexe du répondant}}$$

Le poids final a été calculé sous la forme wgtBC3\*adjBC4. Par conséquent, le poids wgtBC4 correspond au **poids final de l'ETCB** que l'on trouve dans le fichier des données sous le nom de variable WTSE\_M.

## 8.2 Poids bootstrap

Des poids bootstrap coordonnés ont été utilisés pour l'ETCB, en raison de la dépendance de cette dernière à l'égard de l'échantillon de l'ESCC. Donc, le point de départ du calcul de poids bootstrap de l'ETCB a été les 500 répliques tirées du fichier bootstrap principal de l'ESCC avant la poststratification. Chaque réplique bootstrap a été corrigée en utilisant les quatre ajustements énumérés au tableau 8.1.

## 9. Qualité des données

### 9.1 Taux de réponse

À la fin de la collecte des données, le taux de réponse se chiffrait à 80 %. En tout, 87 % des répondants à l'ETCB ont donné leur permission pour le couplage de leurs données, tandis que 72 % ont donné leur permission pour le couplage et ont déclaré un numéro d'assurance-santé valide.

### 9.2 Erreurs dans les enquêtes

L'enquête permet de produire des estimations fondées sur l'information recueillie auprès d'un échantillon de personnes. On aurait obtenu des estimations quelque peu différentes si on avait effectué un recensement complet en utilisant le même questionnaire, les mêmes intervieweurs, les mêmes superviseurs, les mêmes méthodes de traitement, etc. que ceux utilisés pour l'enquête. La différence entre les estimations tirées de l'échantillon et celles qui découlent d'un dénombrement complet effectué dans des conditions semblables s'appellent l'*erreur d'échantillonnage* des estimations.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent être commises à presque toutes les étapes d'une enquête. Il est possible que les intervieweurs comprennent mal les instructions, que les répondants fassent des erreurs en remplissant le questionnaire, que les réponses soient mal saisies et que des erreurs se produisent au moment du traitement et de la totalisation des données. Tous ces exemples représentent des *erreurs non dues à l'échantillonnage*.

#### 9.2.1 Erreurs non dues à l'échantillonnage

Sur un grand nombre d'observations, les erreurs aléatoires ont peu d'effet sur les estimations tirées de l'enquête. Par contre, celles qui se produisent systématiquement introduisent un biais dans les estimations de l'enquête. Beaucoup de temps et d'efforts ont été consacrés à la réduction des erreurs non dues à l'échantillonnage dans l'ETCB. Des mesures d'assurance de la qualité ont été appliquées à chaque étape du cycle de collecte et de traitement des données afin de contrôler la qualité des données. On a notamment fait appel à des intervieweurs hautement qualifiés, qui ont reçu une formation poussée sur les méthodes d'enquête et le questionnaire et que l'on a observés afin de déceler les problèmes. La mise à l'essai de l'application d'IAO et les essais sur le terrain ont également été des mesures essentielles en vue de réduire au maximum les erreurs de collecte de données.

L'effet de la *non-réponse* sur les résultats de l'enquête constitue une source importante d'erreurs non dues à l'échantillonnage dans les enquêtes. L'ampleur de la non-réponse va de la non-réponse partielle (le fait de ne pas répondre à une ou à plusieurs questions) à la non-réponse totale. La non-réponse partielle à l'ETCB était minime; une fois qu'ils commençaient à répondre au questionnaire, les répondants avaient tendance à aller jusqu'au bout et les cas de non-réponse étaient très peu nombreux. Il y a eu non-réponse totale lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire ou que l'intervieweur a été incapable d'entrer en contact

avec elle. On a traité les cas de non-réponse totale en corrigeant les poids des personnes qui ont répondu à l'enquête afin de compenser pour celles qui n'ont pas répondu. Consulter la section 8 pour de plus amples renseignements sur la correction de la pondération pour la non-réponse.

### **9.2.2 Erreurs d'échantillonnage**

Étant donné que les estimations d'une enquête par sondage comportent inévitablement des erreurs dues à l'échantillonnage, de bonnes méthodes statistiques exigent que les chercheurs fournissent aux utilisateurs une certaine indication de la grandeur de cette erreur. La mesure de la grandeur possible des erreurs dues à l'échantillonnage est l'écart-type des estimations calculées d'après les données de l'enquête. Cependant, en raison de la grande diversité des estimations que l'on peut tirer d'une enquête, l'écart-type d'une estimation est habituellement exprimé relativement à l'estimation à laquelle il se rapporte. La mesure ainsi obtenue, appelée coefficient de variation (CV), s'obtient en divisant l'écart-type de l'estimation par l'estimation elle-même et est exprimée en pourcentage de l'estimation.

Par exemple, supposons que l'on estime que 25 % des résidents de la Colombie-Britannique de 18 ans et plus fument régulièrement et que l'écart-type de cette estimation est égal à 0,003. On calcule alors le CV de l'estimation de la façon suivante:

$$(0,003/0,25) \times 100 \% = 1,20 \%$$

Statistique Canada utilise fréquemment les valeurs des CV pour analyser les données et conseille vivement aux utilisateurs produisant des estimations à partir des fichiers de données de l'ETCB de faire de même. Pour plus d'information sur le calcul des CV, voir la section 11. Pour consulter les lignes directrices sur la façon d'interpréter les valeurs du CV, se référer au tableau à la fin de la sous-section 10.4.

## **10. Lignes directrices pour la totalisation, l'analyse et la diffusion**

Cette section de la documentation décrit les lignes directrices que doivent suivre les utilisateurs pour totaliser, analyser, publier ou diffuser de toute autre façon des données tirées du fichier de données de l'enquête. Ces lignes directrices devraient leur permettre de reproduire les chiffres déjà publiés par Statistique Canada et de produire des chiffres non encore publiés conformes aux lignes directrices établies.

### **10.1 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 ce fichier de données correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne leur arrondissement :

- a) les estimations qui figurent dans le corps d'un tableau statistique doivent être arrondies à la centaine près par la méthode d'arrondissement classique. Selon cette méthode, si le premier



- ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre retenu ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) la valeur du dernier chiffre retenu. Par exemple, si l'on veut arrondir à la centaine près de la façon classique une estimation dont les deux derniers chiffres sont compris entre 00 et 49, il faut les remplacer par 00 et ne pas modifier le chiffre précédent (le chiffre des centaines). Si les deux derniers chiffres sont compris entre 50 et 99, il faut les remplacer par 00 et augmenter d'une unité (1) le chiffre précédent;
- b) les totaux partiels de marge et les totaux de marge des tableaux statistiques doivent être calculés à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près selon la méthode d'arrondissement classique;
  - c) les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir d'éléments non arrondis (c'est-à-dire les numérateurs et (ou) dénominateurs), puis arrondis à une décimale par la méthode d'arrondissement classique. Si l'on veut arrondir une estimation à un seul chiffre décimal par cette méthode et que le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à retenir ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente d'une unité (1) le dernier chiffre à retenir;
  - d) les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs éléments correspondants non arrondis, puis arrondies à leur tour à la centaine près (ou à la décimale près) selon la méthode d'arrondissement classique;
  - e) si, en raison de contraintes d'ordre technique ou autre, on applique une autre méthode que l'arrondissement classique, si bien que les estimations qui seront publiées ou diffusées de toute autre façon diffèrent des estimations correspondantes publiées par Statistique Canada, il est vivement conseillé à l'utilisateur d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser;
  - f) des estimations non arrondies ne doivent être publiées ou diffusées de toute autre façon en aucune circonstance. Des estimations non arrondies donnent l'impression d'être beaucoup plus précises qu'elles ne le sont en réalité.

## **10.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation**

Le plan d'échantillonnage utilisé pour cette enquête n'est pas autopondéré. Autrement dit, le poids d'échantillonnage n'est pas le même pour toutes les personnes qui font partie de l'échantillon. Même pour produire des estimations simples, y compris des tableaux statistiques ordinaires, l'utilisateur doit employer le poids d'échantillonnage approprié. Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir du fichier de données ne peuvent pas être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

L'utilisateur ne doit pas non plus perdre de vue qu'en raison du traitement réservé au champ du poids, certains progiciels ne permettent pas d'obtenir des estimations qui coïncident exactement avec celles de Statistique Canada.

### 10.2.1 Définitions : Estimations de type nominal, estimations quantitatives

Avant d'exposer la façon de totaliser et d'analyser les données de l'enquête, il est bon de décrire les deux grandes catégories d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites d'après le fichier de données.

#### Estimations de type nominal:

Les estimations de type nominal sont des estimations du nombre ou du pourcentage de personnes qui, dans la population visée par l'enquête, possèdent certaines caractéristiques ou rentrent dans une catégorie particulière. La fréquence à laquelle les personnes bouclent leurs ceintures quand elles sont au volant est un exemple de ce type d'estimation. Une estimation du nombre de personnes possédant une caractéristique particulière ou manifestant un comportement particulier peut aussi être considérée comme une estimation d'un agrégat.

Exemple de question de type nominal:

*À quelle fréquence attachez-vous votre ceinture de sécurité lorsque vous conduisez un véhicule à moteur? (BCDF\_02)*

- Toujours
- La plupart du temps
- Rarement
- Jamais

#### Estimations quantitatives:

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités qui ont trait à tous les membres de la population observée ou à certains d'entre eux.

Un exemple d'estimation quantitative est le nombre moyen de cigarettes que fument par jour les personnes qui fument tous les jours. Le numérateur correspond à l'estimation du nombre total de cigarettes que fument par jour les personnes qui fument tous les jours et le dénominateur, à l'estimation du nombre de personnes qui fument tous les jours.

Exemple de question quantitative:

*Combien de cigarettes environ fumez-vous habituellement chaque jour? (BCSF\_136)*

|\_|\_| Nombre de cigarettes

### 10.2.2 Totalisation d'estimations de type nominal

Les estimations du nombre de personnes présentant une caractéristique particulière peuvent être calculées d'après le fichier de données en totalisant les poids finals des enregistrements de toutes les personnes possédant la caractéristique d'intérêt.

Pour obtenir les proportions et les rapports de la forme  $\hat{X} / \hat{Y}$ , on doit :

- a) additionner les poids finals des enregistrements contenant la caractéristique voulue pour le numérateur ( $\hat{X}$ );
- b) faire la somme des poids finals des enregistrements contenant la variable étudiée pour obtenir le dénominateur ( $\hat{Y}$ );
- c) diviser l'estimation du numérateur par celle du dénominateur.

### 10.2.3 Totalisation d'estimations quantitatives

Pour obtenir l'estimation d'une somme ou d'une moyenne pour une variable quantitative, on procède aux étapes suivantes (seule l'étape (a) est nécessaire pour obtenir l'estimation pour une somme):

- a) multiplier la valeur de la variable étudiée par le poids final, puis faire la somme de cette quantité pour tous les enregistrements visés pour obtenir le numérateur ( $\hat{X}$ );
- b) faire la somme des poids finals des enregistrements contenant la variable étudiée pour obtenir le dénominateur ( $\hat{Y}$ );
- c) diviser l'estimation du numérateur par celle du dénominateur.

Par exemple, pour obtenir l'estimation du nombre moyen de cigarettes fumées chaque jour au moment où une personne a commencé à fumer parmi les personnes ayant commencé tout de suite à fumer quotidiennement, on calcule d'abord le numérateur ( $\hat{X}$ ) en additionnant les produits de la valeur de la variable BCSF\_104 par le poids WTSF\_M. Puis, on calcule la somme de cette valeur sur l'ensemble des enregistrements dans lesquels figure la valeur « 31 » pour la variable BCSF\_103. Pour calculer le dénominateur ( $\hat{Y}$ ), on totalise les poids finals des enregistrements dans lesquels figure une valeur de « 31 » pour la variable BCSF\_103. En divisant ( $\hat{X}$ ) par ( $\hat{Y}$ ), on obtient le nombre moyen de cigarettes fumées chaque jour, au moment où une personne a commencé à fumer, pour les personnes qui ont fumé quotidiennement dès qu'elles ont commencé à fumer.

Dans cet exemple, les variables BCSF\_103 et BCSF\_104 contiennent les réponses fournies par le répondant aux questions suivantes :

*Quand vous aviez [âge auquel le répondant a commencé à fumer au moins une cigarette par mois] ans, combien de jours dans un mois typique fumiez-vous au moins une cigarette? (BCSF\_103)*

Nombre de jours (si le répondant fumait tous les jours, inscrivez 31)

*Quand vous aviez [âge auquel le répondant a commencé à fumer au moins une cigarette par mois] ans, combien de cigarettes fumiez-vous habituellement chaque jour? (BCSF\_104)*

Nombre de cigarettes

### 10.3 Lignes directrices pour l'analyse statistique

L'ETCB se fonde sur un plan de sondage complexe qui prévoit une stratification et un échantillonnage à plusieurs degrés, ainsi que la sélection des répondants avec probabilités inégales. L'utilisation des données provenant d'une enquête aussi complexe pose des difficultés aux analystes, car le choix des méthodes d'estimation et de calcul de la variance dépend du plan de sondage et des probabilités de sélection.

Nombre de méthodes d'analyse intégrées aux progiciels statistiques permettent d'utiliser des poids, mais la signification et la définition de ces derniers peuvent différer de celles applicables dans le contexte d'une enquête par sondage. Par conséquent, si les estimations calculées au moyen de ces progiciels sont souvent exactes, les variances n'ont, quant à elles, pratiquement aucune signification.

Dans le cas de nombreuses méthodes d'analyse (par exemple la régression linéaire, la régression logistique, l'analyse de la variance), une méthode permet de rendre plus significatifs les résultats produits par les progiciels standard. Elle consiste à réechelonner les poids qui figurent dans les enregistrements de façon à ce que le poids moyen soit égal à un (1). Les résultats produits par les progiciels standard sont ainsi plus raisonnables puisque, même s'ils ne reflètent toujours pas la stratification et la mise en grappes du plan d'échantillonnage, ils tiennent compte de la sélection avec probabilités inégales. On peut effectuer cette transformation en utilisant dans l'analyse un poids égal au poids original divisé par la moyenne des poids originaux pour les unités échantillonnées (personnes) qui contribuent à l'estimation en question.

### 10.4 Lignes directrices pour la diffusion

Avant de diffuser et/ou publier des estimations tirées du fichier de données, l'utilisateur doit déterminer le nombre de répondants dans l'échantillon qui ont la caractéristique d'intérêt (p. ex., le nombre de répondants qui fument si l'on s'intéresse à la proportion de fumeurs dans une

population donnée). Si ce nombre est inférieur à 30, l'estimation non pondérée ne doit pas être diffusée, quelle que soit la valeur de son coefficient de variation. Pour les estimations pondérées basées sur des échantillons d'au moins 30 personnes, l'utilisateur doit calculer le coefficient de variation de l'estimation arrondi et suivre les lignes directrices qui suivent.

**Tableau 10.1 Lignes directrices relatives à la variabilité d'échantillonnage**

Type d'estimation	CV (en %)	Lignes directrices
Acceptable	$0,0 \leq CV \leq 16,6$	On peut envisager une diffusion générale non restreinte des estimations. Aucune annotation particulière n'est nécessaire.
Marginale	$16,6 < CV \leq 33,3$	On peut envisager une diffusion générale non restreinte des estimations, en y joignant une mise en garde aux utilisateurs quant à la variabilité d'échantillonnage élevée liée aux estimations. Les estimations de ce genre doivent être identifiées par la lettre E (ou d'une autre manière similaire).
Inacceptable	$CV > 33,3$	Statistique Canada recommande de ne pas publier d'estimations dont la qualité est inacceptable. Toutefois, si l'utilisateur choisit de le faire, il doit les marquer de la lettre F (ou d'un autre identificateur semblable) et les diffuser avec l'avertissement suivant: «Nous avisons l'utilisateur que . . . (précisez les données)... ne répondent pas aux normes de qualité de Statistique Canada pour ce programme statistique. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne doivent pas être publiées. Si l'utilisateur choisit de les publier, il est tenu de publier également le présent avertissement. »

## **11. Variances et coefficients de variation**

Le calcul des coefficients de variation exacte n'est pas une tâche aisée, car il n'existe aucune formule mathématique simple permettant de tenir compte de tous les aspects de la base de sondage et de la pondération de l'ETCB. Il faut donc recourir à d'autres méthodes, comme des méthodes de rééchantillonnage, pour estimer les mesures de précision. Parmi ces méthodes, celle du bootstrap est recommandée pour l'analyse des données de l'ETCB.

Le calcul des coefficients de variation (ou de toute autre mesure de précision) par la méthode du bootstrap nécessite l'accès à des renseignements qui sont considérés confidentiels. Ce calcul doit être fait en utilisant le fichier maître. L'accès à ce fichier est décrit à la section 12.

Pour le calcul des coefficients de variation, il est conseillé d'utiliser la méthode du bootstrap. Un programme macro, appelé « Bootvar », a été développé afin de permettre aux utilisateurs d'appliquer plus facilement cette méthode. Le programme Bootvar est offert en formats SAS et SPSS, et est constitué de macros qui calculent les variances de totaux, de ratios, de différences entre ratios, et de régressions linéaires et logistiques.

Certains progiciels statistiques standard permettent d'intégrer les poids d'échantillonnage aux analyses, mais, souvent, les variances produites ne tiennent pas bien compte de la stratification et de la mise en grappes de l'échantillon, contrairement à celles obtenues grâce au programme de calcul de la variance exacte.

## **12. Utilisation du fichier**

Cette section débute par la description du fichier maître et de la variable de poids de ce fichier et donne des explications sur la façon de l'utiliser dans les totalisations. Vient ensuite l'explication des règles d'attribution des noms de variables employées pour l'ETCB et l'ESCC. On aborde ensuite le sujet des totalisations personnalisées, en donnant les coordonnées des personnes ressources.

### **12.1 Fichier maître**

Le fichier maître contient toutes les variables et tous les enregistrements de données de l'enquête recueillies durant une période de collecte. Ces fichiers sont accessibles à Statistique Canada pour usage interne et peuvent aussi faire l'objet de demandes de totalisation personnalisée.

### **12.2 Utilisation de la variable de poids**

La variable de poids WTSF\_M représente le poids d'échantillonnage utilisé dans l'ETCB. Pour un répondant donné, ce poids d'échantillonnage peut être interprété comme étant le nombre de personnes que le répondant représente dans la population. Il doit être utilisé en tout temps dans les calculs d'estimations statistiques, afin de permettre l'inférence à l'échelle de la population. La production de résultats non pondérés n'est pas recommandée. La répartition de l'échantillon, de même que les détails du plan de sondage, peuvent entraîner des résultats biaisés qui ne

représentent pas correctement la population. Pour une description plus détaillée du calcul de ce poids, consulter la section 8 sur la pondération.

### 12.3 Règles d'affectations des noms des variables

Pour l'ETCB, on a adopté des règles d'affectation des noms des variables qui permettent d'utiliser et d'identifier facilement les données en se fondant sur le module et le cycle. Les règles sont semblables à la norme appliquée pour l'ESCC, puisque l'ETCB est un supplément au Cycle 3.1 de cette enquête. Les règles d'affectation des noms des variables satisfont à l'obligation de restreindre les noms de variables à un maximum de huit caractères afin de faciliter leur utilisation dans les produits logiciels analytiques.

#### 12.3.1 Structure des composantes du nom des variables dans l'ETCB

Chacun des huit caractères du nom d'une variable fournit des renseignements sur le type de données que contient la variable.

- Positions 1 et 2 :** Référence à l'Enquête sur le tabagisme en Colombie-Britannique (BC)
- Position 3 :** Référence au module (S = Usage du tabac, M = Mobilité, D = Conduite et sécurité et A = Administration)
- Position 4 :** Cycle de l'enquête (E – en référence au Cycle 3.1 de l'ESCC à partir duquel a été tiré l'échantillon et auquel les données ont été couplées en retour)
- Position 5 :** Type de variable (\_ - question, D – variable dérivée)
- Positions 6 à 8 :** Numéro de la question

Par exemple, le nom de la variable correspondant à la question 103 du module sur l'usage du tabac de l'ETCB (BCSE\_103) à la structure suivante :

- Position 1 et 2 :** BC Provient de l'ETCB
- Position 3 :** S Module sur l'usage du tabac
- Position 4 :** E Cycle 3.1 (rappel)
- Position 5 :** \_ Tiret de soulignement (\_ = données recueillies)
- Position 6 à 8 :** 103 Numéro de la question et option de réponse

#### 12.3.2 Structure des composantes du nom des variables dans l'ESCC (à titre de référence)

Chacun des huit caractères du nom d'une variable fournit des renseignements sur le type de données que contient la variable.

- Positions 1 à 3 :** Nom du module/de la section du questionnaire
- Position 4 :** Cycle de l'enquête
- Position 5 :** Type de variable
- Positions 6 à 8 :** Numéro de la question

Par exemple, la structure du nom de la variable correspondant à la question 202, module Usage du tabac, Cycle 2.1, c'est-à-dire SMKC\_202 est la suivante:

<b>Position 1 à 3 :</b>	SMK	Module sur l'usage du tabac
<b>Position 4 :</b>	E	Cycle 3.1
<b>Position 5 :</b>	_	Tiret de soulignement (_ = données recueillies)
<b>Position 6 à 8 :</b>	202	Numéro de la question et option de réponse

### 12.3.3 Positions 1 à 3 : Nom de la variable/section du questionnaire

On se sert des valeurs suivantes pour la composante du nom de la variable correspondant à la section du questionnaire:

BCS	Historique de l'usage du tabac en Colombie-Britannique/exposition à la cigarette au cours de la vie
BCM	Module de mobilité
BCD	Sécurité et la conduite
BCA	Module de l'administration

### 12.3.4 Position 4 : Cycle et nom de l'enquête

(\*il convient de souligner que cette règle a été modifiée pour la collecte de l'ESCC de 2007, la position 4 ne désignant plus un cycle)

Cycle	Nom de l'enquête
A	<u>Cycle 1.1 : Enquête sur la santé dans les collectivités canadiennes</u>
B	<u>Cycle 1.2: Enquête sur la santé dans les collectivités canadiennes, Santé mentale et bien-être</u>
C	<u>Cycle 2.1: Enquête sur la santé dans les collectivités canadiennes</u>
D	<u>Cycle 2.2: Enquête sur la santé dans les collectivités canadiennes, Nutrition</u>
E	<u>Cycle 3.1: Enquête sur la santé dans les collectivités canadiennes</u>
F	<u>ETCB : _Enquête sur le tabagisme en Colombie-Britannique</u>



### 12.3.5 Position 5 : Type de variable

	Variable collectée	Variable qui figure directement sur le questionnaire.
<b>C</b>	Variable codée	Variable codée à partir d'une ou de plusieurs variables collectées (par exemple, code de la Classification type des industries (CTI)).
<b>D</b>	Variable dérivée	Variable calculée d'après une ou plusieurs variables collectées ou codées, ordinairement pendant le traitement au Bureau central (p. ex., indice de l'état de santé).
<b>F</b>	Variable indicatrice	Variable calculée à partir d'une ou de plusieurs variables collectées (comme variable dérivée), mais ordinairement par l'application informatique de collecte des données, aux fins de son utilisation ultérieure durant l'interview (p. ex., indicateur de travail).
<b>G</b>	Variable groupée	Variables collectées, codées, supprimées ou dérivées, agrégées en un groupe (p. ex., groupes d'âge).

### 12.3.6 Positions 6 à 8 : Nom de la variable

En général, les trois dernières positions correspondent à la numérotation de la variable qui figure sur le questionnaire. On supprime la lettre « Q » utilisée pour représenter le mot "question" et on présente tous les numéros de question au moyen d'un groupe de deux chiffres. Par exemple, la question Q01A d'un questionnaire devient simplement 01A, et la question Q15 devient simplement 15.

Quand certaines questions comportent plusieurs réponses, la position finale dans la séquence du nom de la variable correspond à une lettre. Pour ce genre de questions, de nouvelles variables sont créées dans le but de différencier un « oui » d'un « non » pour chaque possibilité de réponse. Par exemple, si la question Q2 possède 4 réponses possibles, les nouvelles questions seront Q2A pour la première possibilité, Q2B pour la deuxième, Q2C pour la troisième et ainsi de suite. Si les options 2 et 3 seulement sont choisies, Q2A = Non, Q2B = Oui, Q2C = Oui et Q2D = Non.

## 12.4 Totalisations personnalisées

Un moyen de donner accès au fichier maître consiste à offrir à tous les utilisateurs l'option de demander au personnel des Services à la clientèle de la Division de la statistique de la santé de préparer des totalisations personnalisées. Ce service est offert contre le recouvrement des coûts. Il permet aux utilisateurs qui ne savent pas se servir de logiciels de totalisation d'obtenir des résultats personnalisés. Les résultats sont filtrés pour s'assurer qu'ils sont conformes aux normes de confidentialité et de fiabilité avant d'être diffusés. Pour plus de renseignements, s'adresser aux Services à la clientèle par téléphone en composant le 613-951-1746 ou par courriel à l'adresse [hd-ds@statcan.ca](mailto:hd-ds@statcan.ca).