Guide de l'utilisateur des microdonnées

Enquête sur les expériences de la maternité

2006

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 Population visée **13**
5.2 Base de sondage **13**
5.2.1 Sous-dénombrement et surdénombrement de la population **13**
5.3 Plan de sondage **14**
5.3.1 Stratification **14**
5.3.2 Répartition **14**
5.4 Taille de l'échantillon **15**

6.0 Collecte des données 17
6.1 Non-réponse **18**

7.0 Traitement des données 19
7.1 Saisie des données **19**
7.2 Vérification **19**
7.3 Codage des questions ouvertes **20**
7.4 Imputation **20**
7.5 Création de variables dérivées **20**
7.6 Pondération **22**

8.0 Qualité des données 23
8.1 Taux de réponse **23**
8.2 Erreurs relatives à l'enquête **24**
8.2.1 Base de sondage **24**
8.2.2 Collecte des données **24**
8.2.3 Traitement des données **26**
8.2.4 Non-réponse **28**
8.2.5 Mesure de l'erreur d'échantillonnage **30**

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données 31
9.1 Lignes directrices pour l'arrondissement d'estimations **31**
9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation **32**
9.3 Définitions de types d'estimations : catégoriques et quantitatives **32**
9.3.1 Estimations catégoriques **32**
9.3.2 Estimations quantitatives **32**
9.3.3 Totalisation d'estimations catégoriques **33**
9.3.4 Totalisation d'estimations quantitatives **33**
9.4 Lignes directrices pour l'analyse statistique **34**
9.5 Lignes directrices pour la diffusion de coefficients de variation **35**
9.6 Seuils pour la diffusion des estimations pour l'Enquête sur les expériences de la maternité **37**

10.0 Tables de variabilité d'échantillonnage approximative 39
10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques **40**
10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques **41**
10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance **45**
10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance **46**
10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t **46**
10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t **47**
10.4 Coefficients de variation pour des estimations quantitatives **47**
10.5 Tables des coefficients de variation **47**

11.0 Pondération 49

12.0 Questionnaire 53

13.0 Cliché d'enregistrement à valeurs univariées 55

1.0 Introduction

Statistique Canada a mené l'Enquête sur les expériences de la maternité (EEM) à l'automne 2006 avec l'appui et la collaboration de l'Agence de la santé publique du Canada. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou 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

Agence de la santé publique du Canada

Section de la santé maternelle et infantile
Division de surveillance de la santé et de l'épidémiologie
Immeuble Jeanne Mance, 10e étage
200, promenade de l'Églantine, I-A 1910D,
Ottawa (Ontario) K1A 0K9
Téléphone : 613-941-2395
Télécopieur : 613-941-9927
Courriel : MES@phac-aspc.gc.ca

2.0 Contexte

L'Enquête sur les expériences de la maternité (EEM) est une initiative du Système canadien de surveillance périnatale (SCSP) dont la gestion est assurée par la Section de la santé maternelle et infantile de l'Agence de la santé publique du Canada. Dans le cadre de cette initiative, le SCSP a établi 52 indicateurs de santé périnatale qui devraient être surveillés dans le cadre d'un système de surveillance périnatale national complet, mais les sources existantes de données au niveau national comme les statistiques de l'état civil, les données sur les hospitalisations et les enquêtes en place sur la santé de la population ne permettaient au SCSP de diffuser des données que sur 27 de ces indicateurs.

En 1999, celui-ci a chargé un groupe de travail multidisciplinaire appelé Groupe d'étude de l'Enquête sur l'expérience de la maternité de concevoir une nouvelle enquête devant procurer l'information manquante. Le Groupe a dirigé les travaux de détermination du contenu et d'élaboration du questionnaire et travaillé avec Statistique Canada à la mise en place de la première enquête nationale sur la maternité.

Les travaux d'élaboration avec Statistique Canada ont débuté en 1999 et un projet pilote a eu lieu en novembre 2002. Les grandes recommandations formulées pour l'enquête principale étaient de réduire le contenu et d'employer l'application des interviews téléphoniques assistées par ordinateur (ITAO) comme mode de collecte. On a aussi recommandé la tenue d'un autre projet pilote portant sur la version modifiée.

Étant donné que l'enquête principale n'a pu être réalisée comme on le prévoyait au départ, ce second essai a aussi été retardé. On a également décidé en cours de route d'exploiter une base de sondage différente, à savoir le Recensement de la population de 2006 au lieu des registres des naissances. Le second projet pilote a eu lieu à l'automne 2005.

C'est à l'automne 2006 que Statistique Canada a réalisé l'enquête pour le compte de l'Agence de la santé publique du Canada. Près de 6 500 mères de bébés de 5 à 14 mois dans les provinces et les territoires ont été interrogées au téléphone sur la grossesse, l'accouchement et le post-partum.

3.0 Objectifs

D'une façon générale, l'Enquête sur les expériences de la maternité (EEM) visait à combler les lacunes statistiques en recueillant des données auprès des mères ayant récemment accouché sur d'importants indicateurs de la santé périnatale, p. ex. :

- les vues de ces nouvelles mères sur les soins reçus;
- les informations qu'elles ont reçues sur la grossesse, l'accouchement et le post-partum;
- les pratiques et facteurs influant sur les expériences de la maternité.

Le but fixé en définitive était de réunir une information permettant d'accroître le bien-être des femmes enceintes, des nouvelles mères et des bébés au Canada. D'un intérêt tout particulier étaient les mères ayant des besoins spéciaux ou dont la grossesse représentait un risque, soit :

- les mères adolescentes, c'est-à-dire ayant moins de 20 ans à la naissance de leur bébé;
- les immigrantes récentes, c'est-à-dire les femmes établies au Canada cinq ans au plus avant la tenue de l'enquête;
- les mères autochtones.

Dans ses 25 modules, le questionnaire s'enquérait des antécédents de reproduction, des caractéristiques sociodémographiques, des comportements et des attitudes des mères avant la grossesse, ainsi que des divers aspects de leurs expériences de la grossesse, du travail, de l'accouchement et du post-partum. Il abordait plus précisément les sujets suivants :

Grossesse :

- sentiments des mères en ce qui concerne la grossesse et le choix du moment;
- ingestion d'acide folique;
- usage du tabac, de l'alcool et des drogues;
- indice de masse corporelle (IMC) avant la grossesse et gain pondéral en période de grossesse;
- événements stressants et violence physique;
- informations sur la grossesse et l'accouchement;
- accès et recours aux soins prénataux.

Travail et accouchement :

- naissance en établissement ou à la maison;
- soignant préposé à l'accouchement;
- présence du conjoint ou du compagnon pendant le travail et l'accouchement;
- évaluation générale de l'expérience de l'accouchement;
- naissance par voie vaginale :
    - mesures prises;
    - administration d'analgésiques;
    - maîtrise non médicamenteuse de la douleur;
    - durée du travail;
- césarienne :
    - prévue – pour des raisons médicales ou non;
    - imprévue – accouchement tenté ou non par voie vaginale;
- âge gestationnel et poids du bébé à la naissance.

Post-partum :

- contact de la mère avec le nouveau-né dans les 24 heures suivant l'accouchement;
- mise aux soins intensifs et séjour à l'hôpital;
- allaitement;
- accès aux soins de santé pour la mère et son bébé;
- problèmes de santé physique dus à l'accouchement;
- dépression post-partum;

- satisfaction à l'égard des soignants avant, pendant et après l'accouchement;
- retour à l'emploi.

4.0 Concepts et définitions

Ce chapitre donne un aperçu des concepts et des définitions d'intérêt pour les utilisateurs. Les utilisateurs sont priés de se reporter au chapitre 12.0 de ce document où figure une copie du (des) questionnaire(s) d'enquête réellement employé(s).

Accouchement prématuré – Travail qui débute avant la fin de la 37e semaine de grossesse – c.-à-d. plus de trois semaines avant la date prévue pour l'accouchement (RH_Q07).

Amniocentèse – Test de diagnostic par lequel une petite quantité du liquide amniotique entourant le bébé dans l'utérus est prélevée à l'aide d'une seringue introduite dans l'abdomen de la femme (SI_Q09).

Analgésie épidurale (ou anesthésie rachidienne) – Type d'anesthésie administrée sous forme d'injection dans le dos (dans la colonne vertébrale par un espace entre deux os) afin d'engourdir le bas du dos (BB_Q01, PM_Q08).

Ballon d'accouchement – Large ballon flexible utiliser par la femme enceinte en travail pour s'asseoir dessus et faire bouger son pelvis et ses hanches (PM_Q01F).

Déchirures périnéales – Déchirures spontanées du périnée qui peuvent surgir pendant l'accouchement (BB_Q03).

Dépression post-partum – Le sentiment de la tristesse, un manque d'intérêt, dépression et une anxiété pendant deux semaines ou plus et qui peuvent durer plusieurs mois. L'Échelle de dépression postnatale d'Édimbourg (Cox et Holden, 1987) est incluse dans la présente enquête (ES_Q01 à ES_Q11).

Doula/aide rémunérée – Personne qui apporte son soutien ou qui supervise lors du travail. Elle détient une formation non médicale (laquelle varie) afin de prêter son appui aux femmes et à leur famille pendant la grossesse, l'accouchement et la période post-partum. Les doulas ne sont pas considérées comme prodiguant des soins de santé (PC_Q05A).

Épisiotomie – Incision que le médecin ou la sage-femme effectue au niveau du périnée ou du vagin afin d'élargir la filière pelvi-génitale et de faciliter l'accouchement (BB_Q03).

Extraction (ou naissance) à vide – Également appelée naissance ou délivrance par aspiration. Procédure où le médecin aide le bébé à naître en plaçant une ventouse sur la tête du bébé afin de l'attirer délicatement vers l'extérieur (VB_Q01B).

Fausse couche – Expulsion du fœtus de l'utérus avant qu'il puisse survivre de manière autonome (en général avant la 20e semaine ou 5e mois de la grossesse). Est parfois aussi appelée avortement spontané. Une grossesse se terminant avec l'expulsion de l'oeuf clair est aussi à considérer comme une fausse couche (RH_Q02 et RH_Q10).

Grossesse ectopique – Aussi désignée de grossesse tubaire. Grossesse où le fœtus se niche à l'extérieur de l'utérus, le plus souvent dans les trompes de Fallope (par exemple, l'œuf fertilisé ne descend pas dans l'utérus). Une telle grossesse nécessite souvent une opération d'urgence dans l'abdomen (RH_Q10).

Jeune mère – Une mère ayant moins de 20 ans à la naissance du bébé sélectionné.

Les immigrants récents font référence aux immigrants reçus qui sont arrivés au Canada au cours des cinq années ayant précédé un recensement donné. Dans le cas du Recensement de 2006, les immigrants récents sont des immigrants reçus qui sont arrivés au Canada entre le 1er janvier 2001 et le jour du recensement, soit le 16 mai 2006.

Mère autochtone – Une mère qui a déclaré être un membre des Premières nations, une métisse ou une Inuite (SD_Q06).

Mort-né – Bébé né sans aucun signe de vie après la 20^{e} semaine ou 5^{e} mois de grossesse (RH_Q02 et RH_Q08).

MSN – Mort subite du nourrisson, aussi appelée mort au berceau. Mort soudaine et inexpliquée d'un nouveau-né apparemment en bonne santé (PI_Q03).

Prestations de maternité - Les prestations de maternité sont payées par le Régime d'assurance-emploi ou au Québec par le gouvernement provincial et sont versées à la mère qui donne naissance, pendant 15 semaines au plus (WA_Q06).

Prestations parentales – Les prestations parentales sont payées par le Régime d'assurance-emploi ou au Québec par le gouvernement provincial et sont versées à l'un ou l'autre des parents pendant qu'ils prennent soin de leur nouveau-né, pendant 35 semaines au plus (WA_Q07).

Soignants – Il s'agit de professionnels de la santé ayant reçu une formation, comme :

- Obstétricien
- Gynécologue
- Un médecin qui est à la fois obstétricien et gynécologue (OBGYN)
- Médecin de famille ou omnipraticien
- Sage-femme
- Infirmier ou infirmier praticien

Remarque : Une doula n'est PAS une soignante.

Soins prénataux – Visites que fait une femme enceinte à un médecin, à une infirmière ou à un autre fournisseur de soins de santé avant le travail et l'accouchement pour subir un examen et obtenir des conseils au sujet de sa grossesse. Les visites pour une échographie ou un test sanguin ne sont pas considérées comme étant des consultations pour des soins prénataux (PC_Q01).

5.0 Méthodologie de l'enquête

Ce chapitre décrit la population visée, la base de sondage et le plan d'échantillonnage de l'Enquête sur les expériences de la maternité (EEM).

5.1 Population visée

La population visée par l'EEM est formée de mères biologiques âgées de 15 ans et plus à la naissance vivante d'un seul enfant au Canada et qui, au moment de l'enquête, vivaient aussi avec leur bébé. Les bébés en question devaient être nés entre le 15 février et le 15 mai 2006 dans les provinces et entre le 1er novembre 2005 et le 1er février 2006 dans les territoires. Les mères qui se trouvaient dans des réserves indiennes ou des logements collectifs au Recensement de 2006 ont été exclues de l'enquête.

5.2 Base de sondage

La base de sondage de l'EEM a été tirée du Recensement de la population de 2006, qui est réalisé à l'aide d'un questionnaire abrégé (formule 2A) et d'un questionnaire complet (formule 2B/2D). Chaque ménage au Canada reçoit l'un ou l'autre. Dans la plupart des régions du pays, on choisit au hasard un ménage sur cinq auquel est destiné le questionnaire 2B/2D. Dans les provinces, on a constitué la base de sondage par le seul échantillon de la formule 2A. Dans les territoires, on a pris les deux formules. Enfin, on a exclu les occupantes des réserves indiennes et des logements collectifs.

La base de sondage de l'EEM a été constituée en plusieurs étapes. On s'est reporté à la date de naissance communiquée au recensement pour repérer les bébés nés entre les dates cibles. On n'a pas retenu les bébés de naissances multiples connus. Ensuite, on a déterminé quelle était la mère de chaque bébé dans la base de sondage. Comme le recensement renseigne uniquement sur les liens entre les membres du ménage et la première personne énumérée dans le questionnaire (personne 1), il a souvent été nécessaire d'examiner à la fois le lien de chaque membre féminin avec la personne 1 et le lien du bébé avec cette même personne pour déduire quel membre du ménage était probablement la mère. Si on ne voyait personne dans le ménage qui puisse être la mère, on retranchait le bébé de la base de sondage. La base de sondage définitive comprenait 58 972 femmes.

5.2.1 Sous-dénombrement et surdénombrement de la population

Étant une enquête postcensitaire, l'EEM a une base de sondage où se retrouvent un grand nombre de problèmes de dénombrement censitaire. Si des recensés manquaient, il y avait sous-dénombrement et, si des gens étaient comptés deux fois, il y avait surdénombrement. Les erreurs de déclaration et de saisie des dates de naissance (tout comme d'autres renseignements dans la base de sondage) pouvaient causer à la fois un sous-dénombrement et un surdénombrement.

Un certain sous-dénombrement était ainsi causé par des délais d'enquête qui obligeaient à créer la base de sondage avant que les données du recensement ne soient entièrement traitées. On corrigeait ce sous-dénombrement par calage (voir le chapitre 11.0).

La population cible était aussi en surdénombrement dans la base de sondage de l'EEM, parce que le recensement ne recueillait pas tous les renseignements permettant d'établir si une mère faisait partie du champ de l'EEM (si elle était la mère biologique d'un bébé

né au Canada). On corrigeait le surdénombrement en repérant les mères hors du champ de l'enquête pendant la collecte des données et en les retranchant de l'échantillon et, par extension, de la population estimée faisant partie du champ de l'enquête.

Un léger sous-dénombrement était enfin possible à cause des bébés qui vivaient avec leur mère une partie du temps, mais qui ne figuraient pas dans la base de sondage, ayant été recensés dans un autre ménage.

5.3 Plan de sondage

Le plan de sondage est un échantillonnage de deuxième phase. À la première phase, il y a l'échantillon du recensement qui a servi à créer la base de sondage de l'EEM. À la deuxième phase, il y a un échantillonnage aléatoire simple stratifié sans remise.

5.3.1 Stratification

On a d'abord stratifié par la province ou le territoire au moment du recensement et par l'âge maternel au moment de l'accouchement (moins de 20 ans et 20 ans et plus). On a sous-stratifié dans le cas des mères de 20 ans et plus :

- dans toutes les provinces selon un âge maternel plus précis (tranches d'âge 20 à 29 et 30 et plus);
- dans toutes les provinces sauf les trois qui sont les moins peuplées selon la résidence de la mère dans une région métropolitaine de recensement (RMR);
- dans les trois provinces les plus populeuses selon la présence d'autres enfants dans le ménage.

Les strates formées de jeunes mères (de moins de 20 ans) étaient généralement trop petites pour être subdivisées sauf en Ontario et au Québec où elles l'ont été selon la résidence de la mère dans une RMR.

5.3.2 Répartition

On a adopté la méthode de Kish de répartition de populations finies pour partager l'échantillon entre les provinces et les territoires. On a choisi un paramètre de 0,7, juste milieu entre une répartition égale et une répartition proportionnelle. Il y a d'abord eu partage au niveau des strates et ensuite à celui des sous-strates en proportion des effectifs de ces dernières. On a suréchantillonné les jeunes mères pour qu'elles soient mieux représentées dans l'échantillon. On a enfin majoré l'échantillon en prévision de la non-réponse, puisqu'on prévoyait un taux de réponse de 75 % pour les provinces et de 70 % pour les territoires.

5.4 Taille de l'échantillon

Le tableau qui suit montre la taille de l'échantillon pour chaque province et territoires.

Provinces et territoires	**Taille de l'échantillon**
Terre-Neuve-et-Labrador	348
Île-du-Prince-Édouard	225
Nouvelle-Écosse	416
Nouveau-Brunswick	400
Québec	1 567
Ontario	2 622
Manitoba	465
Saskatchewan	432
Alberta	814
Colombie-Britannique	871
Yukon	80
Territoires du Nord-Ouest	136
Nunavut	166
Canada	**8 542**

6.0 Collecte des données

Quatre bureaux régionaux de Statistique Canada (Halifax, Sherbrooke, Sturgeon Falls et Winnipeg) se sont chargés de la collecte des données dans le cadre de l'Enquête sur les expériences de la maternité (EEM). De plus, le bureau régional de Toronto s'est vu confier les cas où l'interview devait avoir lieu dans des langues choisies autres que le français et l'anglais.

Au total, une soixantaine d'intervieweuses, toutes des femmes ayant une expérience considérable à mener des interviews, et 10 cadres et surveillants ont été affectés à l'enquête. Ils ont eu droit à deux jours de formation en classe pour apprendre à connaître les concepts de cette enquête et l'application informatique employée.

Dans les provinces, on a utilisé l'application ITAO (interviews téléphoniques assistées par ordinateur). Dans les territoires, là où il était impossible d'interviewer au téléphone, on interviewait sur place avec la version papier du questionnaire. La plupart des interviews ont eu lieu au téléphone dans les territoires. Il y a eu environ 30 cas d'interviews sur place avec le questionnaire papier. Les questions et la formule de consentement au partage des données ont été traduites en trois dialectes inuits.

Préalablement aux interviews, les bureaux régionaux ont posté une lettre de présentation aux mères échantillonnées. Il s'agissait de présenter l'enquête et de solliciter la collaboration des intéressées. On mentionnait un numéro sans frais où obtenir plus de renseignements et on donnait l'adresse du site Web de Statistique Canada où les participantes pouvaient mieux se renseigner sur l'EEM. On a joint à cet envoi le dépliant *Ce que disent les mères : l'Enquête sur les expériences de la maternité* réalisé par l'organisme de parrainage, l'Agence de la santé publique du Canada. Ce document décrivait brièvement l'enquête, en affirmait l'importance et incitait les mères à y prendre part. Dans les provinces, on a distribué un dépliant bilingue et, dans le Nord, une version spéciale en anglais et en syllabique de Baffin avec des photos de mères et de bébés inuits.

La collecte des données a débuté dans les provinces le 23 octobre 2006 comme prévu. Elle s'est terminée dans les délais le 8 décembre aux bureaux de Halifax et de Sherbrooke, mais on l'a prolongée de deux jours (jusqu'au 10 décembre) dans les bureaux de Sturgeon Falls et de Winnipeg pour ajouter au nombre d'interviews, notamment de jeunes mères en Ontario et en Colombie-Britannique.

La collecte a été retardée dans les territoires jusqu'au début de novembre, parce que la base de sondage tirée du recensement n'était pas prête et que les échantillons territoriaux ne pouvaient être constitués à temps. La collecte a duré jusqu'à la fin de janvier et a profité des interviews sur place dans d'autres enquêtes postcensitaires alors menées sur le terrain.

Les immigrantes récentes, qui étaient arrivées au Canada au cours des cinq dernières années, comptaient parmi les groupes d'intérêt spécial aux fins de l'enquête. Pour brosser un tableau précis de leurs expériences, il fallait réduire au minimum le nombre de cas de non-interview en raison de la langue. Les cas codés en ce sens par les bureaux régionaux étaient pris en charge dans la mesure du possible par des intervieweuses du bureau régional de Toronto, qui emploie un certain nombre de travailleurs versés dans les langues d'un grand nombre d'immigrants récents. Les intéressées ont aussi reçu la formation pour l'EEM et se sont vu confier des interviews auprès des immigrantes ayant besoin d'une interprétation entière ou partielle. Les intervieweuses devaient traduire les questions et les réponses au besoin. Elles disposaient d'une liste des termes de l'enquête dans les langues jugées les plus utiles d'après les données du recensement et d'autres données d'enquête. Les intervieweuses du bureau de Toronto ont effectué 150 interviews dans 13 langues. Il est cependant resté des cas où on n'a pas réussi à joindre les mères qui avaient été reconnues comme ayant besoin d'un service d'interprétation (30 cas environ). Enfin, une erreur de planification a fait qu'une cinquantaine de cas qui auraient dû être renvoyés au bureau de Toronto l'ont été trop tard pour que les interviews aient lieu.

En moyenne, les interviews téléphoniques duraient 45 minutes.

6.1 Non-réponse

L'application informatique présentait des caractéristiques spéciales permettant d'optimiser le taux de réussite des tentatives de prise de contact. Les 25 appels prévus par cas étaient répartis entre la semaine et le week-end et entre la matinée, l'après-midi et le soir en fonction de l'enchaînement des tâches des mères de jeunes enfants.

On s'est efforcé de joindre les membres des groupes offrant un intérêt particulier pour les utilisateurs des données. Comme le fichier-échantillon renseignait sur l'âge de la personne que l'on supposait être la mère du bébé sélectionné, il était possible de reconnaître les jeunes mères, c'est-à-dire les femmes ayant moins de 20 ans à la naissance de leur bébé. On pouvait repérer indirectement un autre groupe, celui des immigrantes récentes, même si le fichier-échantillon ne révélait pas ce renseignement, en combinant les données sur la première langue apprise (autre que le français ou l'anglais) et sur la résidence à Vancouver ou à Toronto, où les immigrantes récentes étaient proportionnellement les plus nombreuses. Les mères sélectionnées qui étaient repérées de cette manière pouvaient recevoir une priorité de prise de contact (selon les taux de réussite des tentatives d'entrée en contact avec les groupes cibles pendant la collecte des données).

La non-réponse à l'Enquête sur les expériences de la maternité tient surtout à l'incapacité de prendre contact avec un certain nombre de mères. On a fait les efforts voulus pour repérer les mères qui avaient déménagé ou dont les données de recensement étaient inexactes, mais un nombre appréciable de cas n'étaient pas encore résolus à la fin de la période de collecte.

On a dû estimer le nombre total de cas de non-réponse (par opposition au nombre de cas hors du champ de l'enquête), parce que l'admissibilité devait être confirmée par le volet « tri » de l'interview et qu'il ne pouvait y avoir tri en tout ou en partie s'il n'y avait pas prise de contact.

Malgré la durée de l'interview, les refus de participation ont été rares. Les intervieweuses avaient pour consigne de faire toute tentative raisonnable en vue d'obtenir l'interview en offrant de rappeler au moment convenant à l'intéressée et en expliquant l'importance de la participation. Un autre rappel avait lieu par les soins de l'intervieweuse principale.

7.0 Traitement des données

Le principal produit de l'Enquête sur les expériences de la maternité (EEM) est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

Les intervieweuses saisissent directement les réponses aux questions de l'enquête au moment de l'interview à l'aide d'une version automatisée du questionnaire. L'emploi d'un questionnaire automatisé réduit les délais et coûts de traitement associés à la saisie des données, aux erreurs de transcription et à la transmission des données. On soumet les données des réponses à un cryptage pour en préserver le caractère confidentiel, puis on les transmet par modem au bureau régional concerné de Statistique Canada. Les données sont par la suite transmises à Ottawa, au moyen d'une ligne protégée, pour y être traitées davantage.

Une partie du contrôle se fait au moment de l'interview. Lorsque les renseignements introduits sont hors limites (trop faibles ou trop élevés) des valeurs attendues, ou qu'ils entrent en contradiction avec des renseignements introduits auparavant, l'intervieweur voit paraître à l'écran de l'ordinateur des messages lui demandant de modifier les renseignements. Cependant, pour certaines questions, l'intervieweur a la possibilité de passer outre aux contrôles et de sauter des questions si l'enquêté ne connaît pas la réponse ou refuse de répondre. Pour cette raison, on soumet les données des réponses à d'autres processus de vérification et d'imputation après réception au bureau central.

Dans les territoires, on a effectué une trentaine d'interviews sur place avec un questionnaire papier. On a envoyé les questionnaires remplis au bureau régional de Winnipeg et les réponses ont été entrées dans la version SDAO (saisie des données assistée par ordinateur) du questionnaire électronique de l'EEM. Cette version correspond parfaitement à celle des interviews ITAO, sauf qu'elle est dépouillée des éléments de vérification. Il reste que toutes les données de l'EEM indépendamment de leur mode de saisie ont subi toutes les vérifications nécessaires au stade du traitement.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données 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 à la répondante (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient 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.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles la répondante aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

Voici les codes qu'utilise normalement Statistique Canada à cette fin :

6, 96, 996, etc. = Saut ou enchaînement valide
7, 97, 997, etc. = Ne sait pas
8, 98, 998, etc. = Refus
9, 99, 999, etc. = Non déclaré

7.3 Codage des questions ouvertes

Quelques données élémentaires ont été consignées sur le questionnaire par les intervieweuses sous forme de questions ouvertes. On a inclus dans l'enquête au total 10 questions partiellement ou complètement ouvertes. Les deux questions complètement ouvertes de l'EEM demandaient le nom de l'hôpital ou du centre de naissance et son emplacement. Pour coder ces éléments, on se reportait à une liste de noms d'établissements fournie par l'Agence de la santé publique du Canada.

Huit questions (PC_Q03, SI_Q10, BH_Q03, MH_Q05, PI_Q11, SD_Q10, WA_Q01 et WA_Q08) comportaient une catégorie « Autre – Précisez ». Dans l'ensemble, environ 400 entrées correspondant aux catégories de réponse existantes ont été recodées en cours de traitement. À la question PC_Q03, par exemple, où on demandait « *Quelles sont les raisons qui vous ont empêchée de recevoir des soins prénataux aussi rapidement que vous le vouliez?* », la réponse « Médecin trop occupé pour me prendre plus tôt » a été recodée pour correspondre au premier élément de réponse (qui n'était pas lu à la répondante) « Le médecin/soignant n'était pas disponible ». Par ailleurs, la réponse « Ne voulait pas que les parents le sachent » demeurait codée comme « Autre – précisez ». On n'a pas établi de nouvelles catégories de réponse.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Il n'y a pas eu d'imputation aux fins de l'Enquête sur les expériences de la maternité de 2006.

7.5 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Plusieurs consistent en conversions du système métrique au système impérial (ou vice versa) ou encore en conversions des réponses données selon diverses unités de temps (jours, semaines ou mois) en une même unité, par exemple en semaine. Pour BHD05 et MHD23 par exemple, on convertit en nombre de jours l'âge du bébé au moment de son hospitalisation ou de l'hospitalisation de sa mère lorsque l'âge est déclaré en semaines ou en mois, et ce, par la formule suivante : jours = mois * 4,33 * 7.

Pour d'autres types de variables dérivées, on combine les réponses à deux questions ou plus des répondantes pour qui l'enchaînement des questions a été différent comme dans le cas de SDDEDUC (plus haut niveau de scolarité atteint par la mère) qui est basé sur les questions détaillées sur la scolarité SD_Q11 à SD_Q14.

Comme autre exemple de variable dérivée où on combine les réponses à deux questions ou plus, il y a l'âge gestationnel GEST_AGE (en semaines) qui est calculé d'après la date prévue (pour un âge gestationnel de 40 semaines) et la date d'accouchement déclarée. Pour cette variable, on arrondit vers le bas.

Pour l'indice de masse corporelle (IMC), on compare le poids à la taille des répondantes. C'est une variable dérivée obtenue en divisant le poids en kilogrammes par le carré de la taille en mètres.

On a créé un indice de variable dérivée pour mesurer le nombre d'événements stressants qu'ont connus les répondantes dans les 12 mois ayant précédé l'accouchement. La variable SEDSTRSS est fondée sur les réponses à un ensemble de questions par oui ou non (SE_Q02 à SE_Q14). On a établi le total de réponses affirmatives pour tous les enregistrements ayant des réponses valides à toutes les questions.

L'échelle d'évaluation de dépression post-partum d'Édimbourg a été utilisée dans le questionnaire de l'EEM. Pour chacun des 10 éléments, il y a 4 catégories de réponse parmi lesquelles choisir. L'échelle est de 0, 1, 2 et 3, où 3 indique la dépression la plus profonde. Un résultat total de 13 et plus est généralement considéré comme indiquant une dépression clinique.

Région métropolitaine de recensement (RMR) ou Agglomération de recensement (AR)
La Classification géographique type (CGT) est une classification de régions géographiques utilisée pour la collecte et la diffusion de statistiques.

Selon la classification de 2001, pour former une région métropolitaine de recensement, le centre urbain doit compter au moins 100 000 habitants. Pour former une agglomération de recensement, le centre urbain doit compter au moins 10 000 habitants. Plusieurs changements ont eu lieu depuis 2001 :

Il y a 33 RMR et 111 AR en 2006. Six AR sont devenues des RMR en 2006 :

- Moncton au New Brunswick;
- Barrie, Brantford, Guelph, et Peterborough en Ontario;
- Kelowna en Colombie-Britannique.

Sept nouvelles AR ont été ajoutées :

- Bay Roberts à Terre-Neuve-et-Labrador;
- Miramichi au New Brunswick;
- Centre Wellington et Ingersoll en Ontario;
- Okotoks et Canmore en Alberta;
- Salmon Arm en Colombie-Britannique.

L'AR de Magog en 2001 a été fusionnée à la RMR de Sherbrooke depuis qu'au moins 35 % de la population active occupée qui demeure dans l'AR fait la navette entre la RMR et l'AR, d'après les données sur le lieu de travail tirées du Recensement de 2001. Finalement, deux AR, Gander et Labrador City à Terre-Neuve-et-Labrador, ont été supprimées car la population de leur noyau urbain était inférieure à 10 000 habitants en 2001.

Même si les régions métropolitaines de recensement et les agglomérations de recensement englobent presque 80 % de la population du Canada, ces régions représentent seulement 4 % du territoire. Le concept de **z**one d'**i**nfluence des régions **m**étropolitaines de recensement et des agglomérations de recensement (ZIM) offre une distinction géographique de la région à l'extérieur des RMR et des AR. Les subdivisions de recensement qui ne font pas partie des RMR et des AR sont classées selon une des quatre catégories de ZIM (forte, modérée, faible et sans influence) selon le degré d'influence exercé collectivement sur elles par les RMR et les AR. Les subdivisions de recensement ayant le même degré d'influence (déterminé par le pourcentage de leur population active occupée résidente qui travaille dans le noyau urbain des RMR et des AR) ont tendance à se grouper en zones autour des RMR et des AR.

L'introduction du concept ZIM permet maintenant de classer toutes les subdivisions de recensement dans la nouvelle classification des secteurs statistiques (CSS) en vue de la diffusion des données. En effet, la CSS classe les subdivisions de recensement selon qu'elles fassent

partie d'une région métropolitaine de recensement, d'une agglomération de recensement, d'une zone d'influence des régions métropolitaines de recensement et des agglomérations de recensement (ZIM forte, ZIM modérée, ZIM faible, ou zone sans IM), ou des territoires (Territoire du Yukon, Territoires du Nord-Ouest, Nunavut). L'application de cette classification aux données par subdivision de recensement pourrait dévoiler des particularités cachées en plus d'aider les utilisateurs à étudier la diversité des régions autres que les RMR et les AR.

7.6 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 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 ficher de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Par exemple, si le nombre de mères de moins de 20 ans à la naissance du bébé doit être estimé, on le fait en choisissant les enregistrements pour les membres de l'échantillon présentant cette caractéristique et en sommant les poids de ces enregistrements.

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

8.0 Qualité des données

8.1 Taux de réponse

Le tableau qui suit renferme un résumé des taux de réponse au questionnaire de l'Enquête sur les expériences de la maternité (EEM).

Provinces et territoires	Échantillon total	Cas résolus	Cas estimés admissibles	Répondantes	Taux de cas résolus (%)	Taux de réponse (%)
Terre-Neuve-et-Labrador	348	304	335	279	87,4	83,3
Île-du-Prince-Édouard	225	193	220	184	85,8	83,6
Nouvelle-Écosse	416	368	398	344	88,5	86,4
Nouveau-Brunswick	400	335	388	303	83,8	78,1
Québec	1 567	1 314	1 534	1 256	83,9	81,9
Ontario	2 622	2 059	2 534	1 858	78,5	73,3
Manitoba	465	370	448	341	79,6	76,1
Saskatchewan	432	353	421	341	81,7	81,0
Alberta	814	679	799	651	83,4	81,5
Colombie- Britannique	871	682	841	631	78,3	75,0
Yukon	80	62	79	61	77,5	77,2
Territoires du Nord-Ouest	136	107	117	89	78,7	76,1
Nunavut	166	111	130	83	66,9	63,8
Canada	**8 542**	**6 937**	**8 244**	**6 421**	**81,2**	**77,9**

Voici comment se définissent les colonnes de ce tableau :

Échantillon total
Nombre total d'enregistrements choisis dans la base de sondage.

Cas résolus
Il s'agit du nombre de cas où l'examen d'admissibilité-inadmissibilité s'est fait à l'étape de la collecte des données. Ces cas ont été confirmés comme faisant partie du champ de l'enquête ou étant hors du champ.

Cas estimés admissibles
On a estimé le nombre total de cas admissibles ou faisant partie du champ de l'enquête en calculant la proportion de ces cas parmi les cas résolus et en posant que la proportion serait la même parmi les cas non résolus. On a fait cette estimation à l'intérieur des classes de pondération, parce que les proportions variaient selon ces catégories.

Répondantes
Nombre de cas où les données utilisables étaient suffisantes et où la répondante donnait la permission à Statistique Canada de partager les données avec l'Agence de la santé publique du Canada.

$$Taux\,de\,cas\,résolus = \frac{Cas\,résolus}{Échantillon\,total}$$

$$Taux\,de\ réponse = \frac{Répondantes}{Cas\,estimés\,admissibles}$$

8.2 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 complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweuses, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweuses peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies 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 comprenaient le recours à des intervieweuses hautement qualifiés, une formation poussée des intervieweuses concernant les procédures d'enquête et du questionnaire, l'observation des intervieweuses en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

La base de sondage de l'EEM s'est retrouvée avec un grand nombre de problèmes de dénombrement censitaire. Statistique Canada effectue des études de couverture après chaque recensement. Les résultats provisoires du Recensement de 2006 sont diffusés en mars 2008 et les résultats définitifs, en septembre de la même année. On a estimé à 3,1 % le taux de sous-dénombrement au Recensement de 2001. Les utilisateurs des données doivent se rendre compte que la partie en sous-dénombrement censitaire peut présenter des caractéristiques différentes de celles de toute la population. Il est aussi question de l'exhaustivité de la base de sondage à la section 5.2.1.

8.2.2 Collecte des données

La formation des intervieweuses a consisté à lire le Manuel de l'intervieweur EEM, à assister à une séance de formation en classe d'environ deux jours et à faire les exercices de cas de l'EEM à l'écran. L'information fournie aux intervieweuses comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses.

Voici les principaux éclaircissements et consignes non compris dans le Manuel de l'intervieweur qui ont été fournis à l'occasion de la séance de formation en classe :

- Aux questions CB_Q04 *(Au cours des trois mois précédant votre grossesse de (^nom du bébé), avez-vous pris des multivitamines contenant de l'acide folique ou des suppléments d'acide folique?)* et CB_Q06 *(Au cours des trois premiers mois de votre grossesse de (^nom du bébé), avez-vous pris des multivitamines contenant de l'acide folique ou des suppléments d'acide folique?)*, une réponse par l'affirmative implique que la femme savait (sans simplement le supposer) qu'elle prenait de l'acide folique.

- La question MH_Q22 *(Mis à part pour le travail et l'accouchement, avez-vous passé une nuit à l'hôpital depuis la naissance de (^nom du bébé)?)* parle d'une hospitalisation à cause d'ennuis de santé de la mère.

- À la question BB_Q05 *(Pendant le travail ou l'accouchement, avez-vous eu des complications ou des problèmes de santé qui ont nécessité des soins spéciaux, des tests supplémentaires ou un séjour à l'hôpital?)*, on devait répondre par l'affirmative si on avait subi une césarienne imprévue et par la négative si celle-ci était prévue et s'était déroulée normalement. On a rappelé aux intervieweuses que les femmes qui avaient subi une césarienne, mais en tentant l'accouchement par voie vaginale avaient aussi à répondre à des questions sur ces tentatives.

- Si le poids du bébé était déclaré en grammes, la valeur minimale fixée par l'application était de 1 000. Si le poids était plus bas encore, on devait entrer cette valeur minimale avec une note précisant le poids réel.

- De même, le maximum était de 995 milles ou kilomètres pour la distance à parcourir vers l'hôpital, ce qui ne suffisait pas dans certains cas. On demandait aux intervieweuses d'entrer 995 et de joindre une note précisant la distance réelle.

- Les intervieweuses au Québec ont dû modifier la question sur les prestations de maternité-paternité, les prestations en question n'étant plus versées par le Régime d'assurance-emploi, mais par un régime provincial.

Au début des interviews de l'EEM, on a éprouvé le besoin d'obtenir des consignes supplémentaires sur la façon de composer avec certaines limites de l'application ITAO (interviews téléphoniques assistées par ordinateur) et d'interpréter quelques-unes des questions. De même, les observatrices, c'est-à-dire les membres de l'équipe de l'EEM qui se sont rendus dans tous les bureaux régionaux et qui ont écouté plusieurs interviews, ont fait des suggestions qui devaient être transmises à toutes les intervieweuses. Le personnel des opérations d'enquête au bureau central était chargé de communiquer les consignes à tous les bureaux régionaux.

À l'étape de la collecte des données, ceux-ci ont demandé un certain nombre de choses :

- comment coder à la question PC_Q07B *(Avez-vous suivi ces cours...?)* la réponse « formation prénatale ou obstétricale dans les CLSC (cliniques de santé au Québec) - (Dans une clinique médicale);

- comment interpréter à la question PT_Q05 *(Dans quelle mesure étiez-vous impliquée dans la décision de subir ou non un test de dépistage du VIH?)* la réponse « examen de routine VIH »– (poser des questions supplémentaires afin de savoir si la répondante était impliquée, avait son mot à dire);

- à la question HP_Q01 (*Avant votre grossesse, aviez-vous une maladie quelconque ou des problèmes de santé qui ont nécessité la prise de médicaments pendant plus de 2 semaines, des soins spéciaux ou des tests*

supplémentaires pendant votre grossesse?) l'infécondité doit-elle être considérée comme un problème médical? – (Non);

- à la question PC_Q04 *(Combien de consultations prénatales avez-vous eues?)* les visites qui se font uniquement pour une échographie ou un test sanguin devraient-elles être considérées comme des consultations prénatales? – (Non);

- la question sur l'identité SD_Q08 *(Quelle est votre identité ethnique ou culturelle?)* porte-t-elle sur l'identité par l'ascendance ou une perception quelconque de l'intéressée; comme on demande comment la répondante se perçoit elle-même et le groupe auquel elle a le sentiment d'appartenir, ce pourrait être l'identification par ascendance ou non.

Dans la plupart des provinces, la période de collecte a duré de la semaine du 23 octobre au 8 décembre 2006.

8.2.3 Traitement des données

Pendant le traitement, on a écarté un certain nombre d'enregistrements soit parce que les répondantes n'avaient pas répondu à l'essentiel du questionnaire, c'est-à-dire les modules qui portent sur les expériences depuis la conception jusqu'au post-partum, soit parce qu'elles n'avaient pas donné à Statistique Canada la permission de partager l'information avec l'Agence de la santé publique du Canada. Il y avait 282 enregistrements de ce genre et ceux-ci ont été codés comme étant des cas de non-réponse.

Le traitement des données de l'EEM a été fait par étapes, y compris l'épuration des dossiers, la prévérification, le codage, le contrôle, l'estimation, et la pondération. À chaque étape, une photo des fichiers de sortie est prise et il est facile de faire une vérification en comparant les fichiers de l'étape en cours avec ceux de l'étape antérieure. Cette façon de procéder a beaucoup amélioré l'étape de traitement des données.

À l'étape du traitement, on a relevé un certain nombre d'incohérences et certaines valeurs déclarées ont été jugées impossibles. On a pris des mesures dans le cas des trois variables suivantes :

PC_Q08Y Quelle était la date de naissance prévue de (^nom du bébé)? Année

On a relevé et corrigé trois cas où l'année était inexacte.

GEST_AGE Âge gestationnel du bébé en semaines

Dans 158 cas, on a codé comme « non déclarés » GEST_AGE et la date de naissance prévue c.-à-d. PC_Q08D (jour), PC_Q08M (mois) et PC_Q08Y (année). On a identifié ces cas en combinant l'âge gestationnel et le poids à la naissance et en comparant les valeurs à la nouvelle norme canadienne (Kramer et coll.) de 22 à 43 semaines. Pour chaque âge de gestation, on a jugé impossibles les valeurs de poids qui s'écartaient du poids moyen à la naissance de plus de 2 * IIQ. Cet intervalle interquartile est la différence entre le quartile du haut (Q_3) et le quartile du bas (Q_1).

On a également codé comme « non déclarés » l'âge gestationnel et la date de naissance prévue dans 12 autres enregistrements où cet âge était de plus de 45 semaines.

WAD03 Nombre de semaines de gestation lorsque la répondante a arrêté de travailler

Dans 13 cas où l'âge gestationnel déclaré était de 43 semaines et plus, on a codé comme « non déclarés » WAD03 et les variables liées.

Dans la plupart des cas où on a relevé des incohérences entre variables, on n'a rien fait, ne sachant au juste lesquelles des variables en question étaient erronées. Nous examinons plus loin ces questions de qualité des données :

Introduction de liquides et de solides

BFD08 L'âge du bébé quand d'autres liquides ont été ajoutés à son alimentation
BFD09 L'âge du bébé quand les aliments solides ont été ajoutés à son alimentation, en jours
BFD11 L'âge du bébé lors du dernier allaitement, en jours

On compte 349 cas où la répondante a dit avoir mis fin à l'allaitement avant d'ajouter des liquides à l'alimentation du bébé. Il y a aussi des cas où l'introduction de solides a précédé l'introduction de liquides. Ces incohérences ont diverses explications : faute de mémoire de la répondante, mauvaise interprétation de la question ou erreur d'entrée de l'intervieweuse.

Séjour hospitalier de la mère et âge de réhospitalisation

PP_Q12A Combien de jours, semaines ou mois êtes-vous demeurée à l'hôpital ou clinique après la naissance de (^nom du bébé)?
MH_Q23A Quel âge avait (^nom du bébé) la première fois que vous avez dû passer la nuit à l'hôpital?

Il y a quatre cas où le séjour hospitalier de la mère dépassait l'âge du bébé à la réhospitalisation.

Séjour hospitalier du bébé et âge de réhospitalisation

PPD15 Combien de jours votre bébé est-il resté à l'hôpital ou clinique?
BHD05 L'âge du bébé la première fois qu'il a passé une nuit à l'hôpital, en jours

Il y a 23 cas où le séjour hospitalier du nouveau-né dépassait son âge à la réhospitalisation. Dans 18 de ces cas, les bébés étaient aux soins intensifs ou spéciaux (en néonatalogie), source possible de confusion lorsque la mère a répondu à ces questions.

Séjour hospitalier et date du premier contact à domicile

PPD12 Combien de temps êtes-vous demeurée à l'hôpital ou clinique après la naissance de votre bébé ?
MH_Q02 Quel âge, en jours, avait (^nom du bébé) quand un soignant a communiqué avec vous pour la première fois à la maison?

Il y a 169 cas où le séjour hospitalier de la mère après l'accouchement dépassait l'âge du bébé à la date du premier contact à domicile.

Gain pondéral pendant la grossesse

HWD03IMP Gain de poids par la mère durant sa grossesse, en livres

Il y a 30 cas où le gain pondéral était de plus de 100 livres. Dans certains cas, le gain correspondait au poids de la période antérieure à la grossesse et du post-partum, ce qui porte à croire que les répondantes avaient indiqué leur poids réel en période de grossesse plutôt que le gain pondéral.

Âge gestationnel du bébé au moment où la répondante a cessé de travailler
Il y a plus d'une centaine de cas où la répondante a cessé de travailler à un âge gestationnel très bas, c'est-à-dire à moins de six semaines; il y a aussi 55 cas où l'âge gestationnel à la date de cessation du travail dépassait l'âge gestationnel à la naissance.

Naissance prématurée

PREMATUR Nombre de grossesses qui se sont terminées par une naissance prématurée

Il y a 48 cas où la mère n'avait pas déclaré de naissance avant terme (PREMATUR = 00), mais où l'âge gestationnel GEST_AGE calculé par la date prévue de l'accouchement et la date de naissance était de moins de 37 semaines. Il y a 22 cas où l'âge gestationnel était de 37 semaines et plus et où la mère avait déclaré que toutes ses naissances étaient avant terme.

Nombre de grossesses

BIRTH Combien de fois avez-vous donné naissance à un bébé vivant?
STILLBTH Combien de vos grossesses se sont terminées par la naissance d'un enfant mort-né?
ECTOPIC Combien de vos grossesses se sont terminées par une grossesse tubaire ou extra-utérine?
MISCARRY Combien de vos grossesses se sont terminées par une fausse couche?
ABORTION Combien de vos grossesses se sont terminées par un avortement thérapeutique ou volontaire?

Il y a 21 cas où le nombre déclaré de naissances vivantes, de mortinaissances, de fausses couches, de grossesses extra-utérines et d'avortements thérapeutiques ou provoqués ne correspond pas au nombre de grossesses. Certains écarts pourraient s'expliquer par des naissances multiples, mais ces données ne sont pas recueillies par le questionnaire de l'EEM.

Âge au premier accouchement

RH_Q05 Quel âge aviez-vous lorsque vous avez donné naissance à un enfant vivant pour la première fois?
AGEATBTH Âge de la mère à la naissance du bébé sélectionné

La fréquence déclarée du même âge au premier accouchement et à l'accouchement visé par l'EEM (36 cas) semble bien supérieure à ce que laissent prévoir les estimations tirées des données hospitalières sur les naissances. On ignore au juste les raisons de cette plus grande fréquence dans le cadre de l'EEM.

8.2.4 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec la répondante, ou que la répondante a refusé de participer à l'enquête. Les cas de non-réponse totale ont été traités en ajustant le poids des personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu.

Dans la plupart des cas, il y a eu non-réponse partielle au questionnaire d'enquête lorsque la répondante n'a pas compris ou a mal interprété une question, a refusé d'y répondre ou ne pouvait se rappeler l'information demandée.

On a comparé les caractéristiques des 1 887 cas non résolus et sans réponse à celles des 6 421 répondantes à l'aide des données de recensement disponibles dans la base de sondage de l'EEM. Pour cette évaluation, les données de recensement ont été pondérées selon le plan d'échantillonnage de l'enquête. Les différences entre les deux ensembles d'estimations sont l'indice de différences possibles de caractéristiques entre les deux groupes. On a établi un troisième ensemble pour les répondantes en combinant les données de recensement à la pondération finale de l'EEM, laquelle corrige la pondération d'échantillonnage en fonction des cas non résolus et des cas de non-réponse (voir le chapitre 11.0), le but étant de réduire le biais possible de non-réponse. On peut juger du degré réel de diminution de ce biais en comparant le troisième ensemble à un ensemble final d'estimations qui est fondé sur toute la base de sondage de l'EEM. Les résultats figurent au tableau qui suit :

Variables		**Répartition (en pourcentage) pour chaque source d'estimations**			
		Non résolus / Non-réponse	**Répondantes (pondération d'échantillonnage)**	**Répondantes (pondération finale)**	**Base de sondage intégrale EEM**
Âge de la mère lors de la naissance	Moins de 20	5,3	2,6	3,2	3,3
	20 à 29	47,0	44,4	45,3	44,6
	30 à 39	41,7	49,4	48,0	47,4
	40 ans et plus	7,9	3,6	3,6	4,7
Première langue apprise de la mère	Anglais	58,3	61,4	61,6	61,0
	Français	14,7	23,6	21,5	21,7
	Autre	27,0	15,0	16,9	17,2
Lieu de résidence	Non-RMR	27,4	33,7	32,5	32,2
	RMR	47,3	50,9	50,1	50,0
	Toronto	25,3	15,4	17,5	17,8
État matrimonial de la mère	Jamais mariée	34,8	30,4	29,6	30,8
	Autre	65,2	69,6	70,4	69,2
Nombre d'adultes dans le ménage (18 ans et plus)	Moins de 2	11,8	4,2	4,2	5,7
	2 et plus	88,2	95,8	95,8	94,3
Autres enfants (0 à 15 ans)	Non	40,1	42,9	42,8	42,3
	Oui	59,9	57,1	57,2	57,7
Sexe du bébé	Masculin	52,4	51,8	51,7	51,6
	Féminin	47,6	48,2	48,3	48,4

Les résultats de cette étude indiquent que les caractéristiques des cas de non-réponse et des cas non résolus tendent à différer de celles des cas de réponse. On peut également voir que, après correction pour la non-réponse, les répondantes paraissent plutôt représentatives pour les variables disponibles de l'étude.

En guise d'illustration de cette conclusion, considérons la proportion de cas où la langue première de la mère n'était ni le français ni l'anglais. Les proportions étaient de 27,0 % pour les cas de non-réponse et les cas non résolus et de 15,0 % pour les cas de réponse pondérés selon le plan d'échantillonnage. Après la correction de la pondération, les

proportions étaient de 16,9 % pour les répondantes et de 17,2 % pour la base de sondage de l'EEM. On peut donc voir que, bien que les proportions soient fort différentes pour les cas de non-réponse et non résolus, la correction aide à compenser et à rendre l'échantillon plus représentatif.

8.2.5 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. Cette 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 26,3 % des accouchements au Canada sont par césarienne et l'on constate que l'erreur-type de cette estimation est de 0,0058. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0{,}0058}{0{,}263}\right) X\ 100\,\% = 2{,}2\,\%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent 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, tout en étant en mesure d'obtenir 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 qui sont 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 de 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 les 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 technique 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 ne le sont en réalité.

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

Le plan d'échantillonnage utilisé pour l'Enquête sur les expériences de la maternité (EEM) n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête 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 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'EEM, 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'EEM.

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 d'accouchements par voie vaginale constitue un exemple de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : Avez-vous eu un accouchement par voie vaginale ou par césarienne pour (^nom du bébé)?
R : Par voie vaginale / Par césarienne

Q : Immédiatement après sa naissance, est-ce que (^nom du bébé) a été admis à une unité de soins intensifs ou de soins spéciaux?
R : Oui / Non

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et 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 consultations prénatales.

Le numérateur est une estimation du nombre total de consultations prénatales et son dénominateur est le nombre de mères ayant eu des consultations prénatales.

Exemples de questions quantitatives :

Q : Combien de consultations prénatales avez-vous eues?
R : |_|_| consultations

Q : Combien d'années de scolarité avez-vous complétées en commençant avec la première année du primaire et excluant les années redoublées?
R : |_|_| années

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 des proportions et des 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 consultations prénatales en Ontario multipliez la valeur déclarée à la question PC_Q04 (combien de consultations prénatales) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable PC_Q06 = 35 (reçu le plus de soins prénataux en Ontario).

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 consultations prénatales en Ontario,

a) estimez le nombre total de consultations prénatales en Ontario $\left(\hat{X}\right)$ tel qu'il est décrit ci-dessus,
b) estimez le nombre de mères $\left(\hat{Y}\right)$ incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable PC_Q06 = 35, 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'EEM repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondantes. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. 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 d'utiliser des poids, la signification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte 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. Les 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 toutes les mères de moins de 20 ans à la naissance du bébé. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

1) sélectionner tous les enregistrements du fichier où la mère avait moins de 20 ans à la naissance du bébé (AGEGRBTH = 01);

2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre d'enregistrements où la mère avait moins de 20 ans à la naissance du bébé;

3) pour chacune de ces répondantes, calculer un poids REMIS À L'ÉCHELLE égal au poids original divisé par le poids MOYEN;

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

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

On recommande d'employer la méthode bootstrap pour calculer les estimations de variance de l'EEM. C'est une méthode de rééchantillonnage pour des estimations valides de variance dans le cas de plans d'échantillonnage et d'estimateurs complexes. Une pondération bootstrap est produite pour l'EEM et présentée avec le fichier de données. Statistique Canada a élaboré un programme appelé Bootvar qui estime la variance en pondération bootstrap pour les totaux, les rapports, les différences entre rapports et les régressions linéaires et logistiques. On peut télécharger le programme et la documentation qui s'y rapporte du site Web de Statistique Canada à l'adresse http://www.statcan.ca/francais/rdc/whatdata_f.htm#tools.

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'EEM, 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, dont il a été question au chapitre 8.0, influencent la qualité des données. 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 répondantes retenues 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'échantillon 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 pondérées arrondies.

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.

Il convient de noter qu'on peut oublier l'exigence d'une taille d'échantillon de 30 lorsqu'on estime les coefficients de variation (CV) à l'aide de techniques d'estimation plus précises comme les méthodes de répétition, par opposition au calcul de variances approximatives d'après les tables de CV dont parle le chapitre 10.0.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	Les estimations proviennent d'une taille d'échantillon de 30 ou plus (si on utilise les tables de CV), et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %. Aucune mise en garde n'est requise.
2) Médiocre	Les estimations proviennent d'une taille d'échantillon de 30 ou plus (si on utilise les tables de CV), et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %. Ces estimations devraient être signalées par la lettre E (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	Les estimations proviennent d'une taille d'échantillon inférieure à 30 (si on utilise les tables de CV), ou présentent des coefficients de variation très élevés, supérieurs 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 F (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante : « Nous informons l'utilisateur que ces estimations (désignées avec la lettre F) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront pas fiables et seront très probablement invalides. »

9.6 Seuils pour la diffusion des estimations pour l'Enquête sur les expériences de la maternité

Le tableau ci-dessous fournit une indication de la précision des estimations démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 50 personnes possédant une caractéristique donnée à Terre-Neuve-et-Labrador est médiocre.

Veuillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Provinces et territoires	**CV acceptable 0,0 à 16,5 %**	**CV médiocre 16,6 à 33,3 %**	**CV inacceptable > 33,3 %**
Terre-Neuve-et-Labrador	90 et plus	30 à < 90	moins de 30
Île-du-Prince-Édouard	30 et plus		moins de 30
Nouvelle-Écosse	136 et plus	35 à < 136	moins de 35
Nouveau-Brunswick	136 et plus	36 à < 136	moins de 36
Québec	500 et plus	125 à < 500	moins de 125
Ontario	556 et plus	138 à < 556	moins de 138
Manitoba	231 et plus	61 à < 231	moins de 61
Saskatchewan	211 et plus	55 à < 211	moins de 55
Alberta	475 et plus	121 à < 475	moins de 121
Colombie-Britannique	471 et plus	120 à < 471	moins de 120
Yukon	30 et plus		moins de 30
Territoires du Nord-Ouest	30 et plus		moins de 30
Nunavut	30 et plus		moins de 30
Canada	**478 et plus**	**118 à < 478**	**moins de 118**

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 de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'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 ceux-ci une valeur modérée (habituellement le 75^{e} percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des 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 sur les expériences de la maternité (EEM).

Provinces et territoires	Effet du plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,01	279	1 020
Île-du-Prince-Édouard	0,92	184	288
Nouvelle-Écosse	1,00	344	1 725
Nouveau-Brunswick	1,04	303	1 488
Québec	1,03	1 256	18 333
Ontario	1,03	1 858	29 688
Manitoba	1,00	341	2 692
Saskatchewan	1,02	341	2 443
Alberta	1,01	651	9 435
Colombie-Britannique	1,02	631	8 997
Yukon	0,88	61	82
Territoires du Nord-Ouest	0,78	89	146
Nunavut	0,92	83	170
Canada	**1,20**	**6 421**	**76 508**

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 ou en utilisant la pondération bootstrap présentée avec le fichier de données. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation 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 est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

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, de femmes dont le soignant principal à l'accouchement était une sage-femme et qui jugeaient très positives leurs expériences du travail et de l'accouchement, est plus fiable que le nombre estimé de femmes dont le soignant principal à l'accouchement était une sage-femme et qui jugeaient très positives leurs expériences du travail et de l'accouchement. (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 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 femmes dont le soignant principal à l'accouchement était une sage-femme et le numérateur, le nombre de femmes dont le soignant principal à l'accouchement était une sage-femme et qui jugeaient très positives leurs expériences du travail et de l'accouchement.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de bébés de sexe masculin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement comparativement au nombre de bébés de sexe féminin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement, l'erreur-type du rapport des estimations est approximativement égale 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 de $\hat{X}_2$ 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 de l'EEM 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 les sages-femmes étaient le soignant principal à l'accouchement pour 3 207 femmes. 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.

2) L'agrégat estimé 3 207 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 3 000.

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 6,4 %.

4) Le coefficient de variation approximatif de l'estimation est donc 6,4 %. Le résultat selon lequel les sages-femmes étaient le soignant principal à l'accouchement pour 3 207 femmes (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1), peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que 2 268 / 3 207 = 70,7 % des femmes, dont le soignant principal à l'accouchement était une sage-femme, jugeaient très positives leurs expériences du travail et de l'accouchement. 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.

2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les femmes dont le soignant principal à l'accouchement était une sage-femme), il faut utiliser à la fois le pourcentage (70,7 %) et la portion numérateur du pourcentage (2 268) pour déterminer le coefficient de variation.

3) Le numérateur, 2 268, 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 2 500. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 70,0 %.

4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 4,0 %, est le coefficient de variation à employer.

5) Le coefficient de variation approximatif de l'estimation est donc 4,0 %. Le résultat selon lequel 70,7 % des femmes, dont le soignant principal à l'accouchement était une sage-femme, jugeaient très positives leurs expériences du travail et de l'accouchement, peut être publié sans réserve.

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

Supposons qu'un utilisateur estime que le soignant principal à l'accouchement pour 2 268 / 3 207 = 70,7 % des femmes était une sage-femme et pour 36 688 / 69 376 = 52,9 % des femmes était un obstétricien, un gynécologue, un médecin de famille ou une infirmière/infirmière praticienne et qu'elles jugeaient très positives leurs expériences du travail et de l'accouchement. 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 du CANADA de la même façon que celle décrite dans l'exemple 2, donne un CV de l'estimation pour les femmes dont le soignant principal à l'accouchement était une sage-femme de 4,0 % et un CV de l'estimation pour les femmes dont le soignant principal à l'accouchement était un obstétricien, un gynécologue, un médecin de famille ou une infirmière/infirmière

praticienne de 1,5 %.

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 (les femmes dont le soignant principal à l'accouchement était une sage-femme), $\hat{X}_2$ est l'estimation 2 (les femmes dont le soignant principal à l'accouchement était un obstétricien, un gynécologue, un médecin de famille ou une infirmière/infirmière praticienne) 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 $\hat{d} =$ 0,707 – 0,529 = 0,178 est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0{,}707)(0{,}040)]^2 + [(0{,}529)(0{,}015)]^2} \\ &= \sqrt{(0{,}0007997) + (0{,}0000629)} \\ &= 0{,}02937\end{aligned}$$

3) Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}} / \hat{d} =$ 0,02937 / 0,178 = 0,165.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 16,5 %. La différence entre les estimations est considérée acceptable et cette estimation peut être publié sans réserve.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime qu'en Colombie-Britannique 646 bébés de sexe masculin et 295 bébés de sexe féminin ont été admis aux soins intensifs ou spéciaux immédiatement après l'accouchement. L'utilisateur est intéressé à comparer l'estimation des bébés de sexe masculin à celle des bébés de sexe féminin en Colombie-Britannique 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 d'un rapport, où le numérateur de l'estimation ($\hat{X}_1$) est le nombre de bébés de sexe masculin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement en Colombie-Britannique. Le dénominateur de l'estimation ($\hat{X}_2$) est le nombre de bébés de sexe féminin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement en Colombie-Britannique.

2) Reportez-vous à la table des coefficients de variation pour la Colombie-Britannique.

3) Le numérateur de cette estimation de rapport est 646. Le chiffre qui s'en rapproche le plus est 650. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 13,7 %.

4) Le dénominateur de cette estimation de rapport est 295. Le chiffre qui s'en rapproche le plus est 300. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 20,7 %.

5) Le coefficient de variation approximatif de l'estimation du rapport est donc 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 :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0{,}137)^2 + (0{,}207)^2} \\ &= \sqrt{0{,}0188 + 0{,}0428} \\ &= 0{,}248\end{aligned}$$

6) Le rapport obtenu entre les bébés de sexe masculin et ceux de sexe féminin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement en Colombie-Britannique est de 646 / 295, c'est-à-dire 2,19 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 24,8 %. La qualité de cette estimation est considérée marginale. Si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre E (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les bébés de sexe masculin et ceux de sexe féminin admis aux soins intensifs ou spéciaux immédiatement après l'accouchement s'établit à 1,30 pour l'Ontario tandis qu'il est de 2,19 pour la Colombie-Britannique. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport de l'Ontario ($\hat{R}_1$) et le rapport de la Colombie-Britannique ($\hat{R}_2$) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport de l'Ontario est 12,7 % et 24,8 % pour celui de la Colombie-Britannique.

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

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

où α_1 et α_2 sont les coefficients de variation de $\hat{R}_1$ et $\hat{R}_2$ respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d}$ = 1,30 – 2,19 = -0,89 est :

$$\sigma_{\hat{d}} = \sqrt{[(1{,}30)(0{,}127)]^2 + [(2{,}19)(0{,}248)]^2}$$
$$= \sqrt{(0{,}0273) + (0{,}2950)}$$
$$= 0{,}568$$

3) Le coefficient de variation de $\hat{d}$ est donné par $\sigma_{\hat{d}} / \hat{d}$ = 0,568 / (-0,89) = -0,638.

4) Le coefficient de variation approximatif de la différence entre les estimations est donc 63,8 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

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

Bien que les coefficients de variation soient 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 suit :

> 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.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues 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. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

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

Des intervalles de confiance pour une estimation peuvent être calculés directement 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 pour 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 de femmes dont le soignant principal à l'accouchement était une sage-femme et qui jugeaient très positives leurs expériences du travail et de l'accouchement (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$\hat{X}$ = 70,7 % (ou exprimé sous forme de proportion 0,707)

t = 2

$\alpha_{\hat{x}}$ = 4,0 % (0,040 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.

$IC_{\hat{x}}$ = {0,707 – (2) (0,707) (0,040), 0,707 + (2) (0,707) (0,040)}

$IC_{\hat{x}}$ = {0,707 – 0,057, 0,707 + 0,057}

$IC_{\hat{x}}$ = {0,650, 0,764}

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 65,0 % et 76,4 % des femmes dont le soignant principal à l'accouchement était une sage-femme jugeaient très positives leurs expériences du travail et de l'accouchement.

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

Des erreurs-types peuvent aussi être utilisés 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, elle 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 $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ 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 de femmes dont le soignant principal à l'accouchement était une sage-femme et la proportion de femmes dont le soignant principal à l'accouchement était un obstétricien, un gynécologue, un médecin de famille ou une infirmière/infirmière praticienne, et qui jugeaient très positives leurs expériences du travail et de l'accouchement. 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 0,02937. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0{,}707 - 0{,}529}{0{,}02937} = \frac{0{,}178}{0{,}02937} = 6{,}06$$

Puisque t = 6,06 est supérieur à 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'EEM sont principalement de nature catégorique.

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 consultations prénatales serait supérieur au coefficient de variation de la proportion correspondante de mères ayant reçu des soins prénataux. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés en utilisant la pondération bootstrap présentée avec le fichier de données (voir la section 9.4).

10.5 Tables des coefficients de variation

Consulter le fichier EEM2006_CVTabsF.pdf pour les tables de coefficients de variation.

11.0 Pondération

Ce chapitre décrit les étapes pour établir la pondération finale de l'Enquête sur les expériences de la maternité (EEM). Les enregistrements ont d'abord été répartis dans les catégories suivantes :

Catégorie		Description	Nombre
Non résolus (UNR)	UNR1	Enregistrement non transmis à la collecte de données sur le terrain (p. ex. manque de coordonnées)	70
	UNR2	Enregistrement envoyé sur le terrain, mais aucune donnée reçue pour déterminer s'il fait partie du champ de l'enquête	1 285
	UNR3	Données partielles reçues, mais ne suffisant pas à déterminer si l'enregistrement fait partie du champ de l'enquête (dans le champ selon le module SRC mais le module VSB est incomplet)	250
Hors du champ de l'enquête (OOS)	OOS1	Hors du champ de l'enquête du type 1 (selon le module SRC)	99
	OOS2	Hors du champ de l'enquête du type 2 (selon le module VSB)	135
Admissibles	RESP	Répondantes (RESP) : données suffisantes et permission de partager l'information	6 421
	NR	Non-répondantes (NR) : données insuffisantes ou absence de permission de partager l'information	282
		Taille totale de l'échantillon initial	**8 542**

Note : Le module SRC comporte des questions sur la présence d'un bébé dans le ménage (au moins une nuit par mois), le nom de la mère et sa propre présence dans le ménage. S'il y a erreur d'identification de la mère dans le fichier-échantillon, l'intervieweuse demande à parler à la mère.

Le module VSB comporte des questions sur la mère et la date de naissance du bébé; on confirme qu'il s'agit de la mère biologique, que le bébé est né au Canada et qu'il ne s'agit pas d'une naissance multiple.

Toutes les fois que les effectifs sont suffisants, on calcule des valeurs de pondération d'échantillonnage dans des classes qui, pour la plupart, correspondent aux strates initiales. On a défini des classes de pondération supplémentaires au Nouveau-Brunswick, au Québec et en Colombie-Britannique pour les mères plus âgées en fonction de la première langue apprise et dans les territoires pour ces mêmes mères en fonction de l'état autochtone. On réalise les étapes suivantes :

1. Calcul de la pondération d'échantillonnage

Chacun des 8 542 enregistrements de l'échantillon initial reçoit une valeur de pondération déterminée par le plan d'échantillonnage W_1 qui correspond au poids censitaire multiplié par l'inverse des probabilités de sélection dans chaque strate :

$$W_1 = \text{poids censitaire} \times \left(\frac{\text{nombre d'unités dans la base de sondage du recensement}}{\text{nombre d'unités dans l'échantillon de l'EEM}} \right)$$

Le poids censitaire des enregistrements dans les provinces est le nombre de formules 2A, 2B et 2D divisé par le nombre de formules 2A (5/4 environ) dans l'unité de collecte (UC); dans le cas des enregistrements du Nord, le poids censitaire est l'unité.

2. Correction en fonction des unités non mises en collecte (UNR1)

On corrige la pondération pour tenir compte des 70 enregistrements non envoyés à la collecte sur le terrain. Les enregistrements sont regroupés en classes de pondération et on calcule les corrections suivantes pour chaque classe :

$$W_2 = \left(\frac{\sum W_1 \text{ pour les enregistrements mis en collecte} + \sum W_1 \text{ pour les enregistrements non mis en collecte}}{\sum W_1 \text{ pour les enregistrements mis en collecte}} \right) \times W_1$$

3. Correction en fonction des unités non résolues pour lesquelles il y a absence de données (UNR2)

On corrige la pondération pour tenir compte des 1 285 enregistrements mis en collecte, mais pour lesquels on n'a pas reçu de données pour déterminer si les unités font partie du champ de l'enquête. Les enregistrements sont regroupés en classes de pondération et on calcule les corrections suivantes pour chaque classe :

$$W_3 = \left(\frac{\sum W_2 \text{ pour les enregistrements caractérisés comme RESP, NR, OOS1, OOS2 ou UNR3} + \sum W_2 \text{ pour les enregistrements caractérisés comme UNR2}}{\sum W_2 \text{ pour les enregistrements caractérisés comme RESP, NR, OOS1, OOS2 ou UNR3}} \right) \times W_2$$

4. Correction en fonction des unités non résolues pour lesquelles les données sont partielles (UNR3)

On corrige la pondération pour tenir compte des 250 enregistrements pour lesquels les données sont partielles, mais ne suffisent pas à déterminer si les unités font partie du champ de l'enquête. Les enregistrements sont regroupés en classes de pondération et on calcule les corrections suivantes pour chaque classe :

$$W_4 = \left(\frac{\sum W_3 \text{ pour les enregistrements caractérisés comme RESP, NR ou OOS2} + \sum W_3 \text{ pour les enregistrements caractérisés comme UNR3}}{\sum W_3 \text{ pour les enregistrements caractérisés comme RESP, NR ou OOS2}} \right) \times W_3$$

5. Correction en fonction des cas de non-réponse (NR)

On corrige la pondération pour tenir compte des 282 cas de non-répondantes faisant partie du champ de l'enquête. Il s'agit d'enregistrements dont les données sont insuffisantes ou d'enregistrements où la répondante n'a pas donné la permission de partager l'information avec l'Agence de la santé publique du Canada. Les enregistrements sont regroupés en classes de pondération et on calcule les corrections suivantes pour chaque classe :

$$W_5 = \left(\frac{\sum W_4 \text{ pour les enregistrements caractérisés comme RESP} + \sum W_4 \text{ pour les enregistrements caractérisés comme NR}}{\sum W_4 \text{ pour les enregistrements caractérisés comme RESP}} \right) \times W_4$$

6. Calage

La base de sondage initiale de l'EEM faisait appel à des données de recensement non vérifiées et ne tenait pas compte de toute la population visée. Les poids de l'EEM ont fait l'objet d'un calage d'après les estimations provinciales-territoriales de la population enquêtée selon les données définitives de recensement :

$$W_6 = \left(\frac{\text{chiffres définitifs du recensement}}{\sum W_5 \text{ pour les enregistrements caractérisés comme RESP} + \sum W_5 \text{ pour un certain nombre d'enregistrements OOS}} \right) \times W_5$$

Les enregistrements hors du champ de l'enquête inclus dans le dénominateur sont ceux des mères ayant accouché à l'étranger, des mères non biologiques, des mères ou des bébés décédés ou institutionnalisés et des mères qui ne vivent plus avec leur bébé. On prend alors en considération le surdénombrement à prévoir de la population enquêtée par les chiffres définitifs de recensement (quantité inconnue) en fonction de la représentation pondérée de ces unités dans notre échantillon (quantité connue)

12.0 Questionnaire

Consulter le fichier EEM2006_QuestF.pdf pour la version française du questionnaire utilisé pour la collecte des données.

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier EEM2006_Partagé_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.