



*Plan commun de
surveillance de
l'état de santé de
la population et de
ses déterminants*

Guide spécifique des aspects méthodologiques
des données d'enquêtes sociosanitaires

– Enquête québécoise sur la santé de la population 2008 –

Institut de la statistique du Québec

en collaboration avec
Institut national de santé publique du Québec
Ministère de la Santé et des Services sociaux

Novembre 2010

Ce document a été réalisé dans le cadre des travaux du Comité tripartite MSSS-INSPQ-ISQ sur les aspects méthodologiques des données d'enquêtes sociosanitaires. Ce comité est sous la responsabilité de la Direction de la surveillance de l'état de santé du ministère de la Santé et des Services sociaux.

Coordination et rédaction

Jimmy Baulne, Direction de la méthodologie et de la qualité, ISQ

Collaboration à la rédaction

Maxime Boucher, Direction de la méthodologie et de la qualité, ISQ
Robert Courtemanche, Direction de la méthodologie et de la qualité, ISQ
Suzanne Gingras, Infocentre de santé publique, INSPQ
Monique Godin, Direction de la surveillance de l'état de santé, DGSP–MSSS
Denis Hamel, Unité Connaissance-surveillance, INSPQ

Secrétariat

Diane Brazeau, Direction de la surveillance de l'état de santé, DGSP–MSSS

Relecture

Membres du Comité tripartite MSSS-INSPQ-ISQ sur les aspects méthodologiques des données d'enquêtes sociosanitaires

Carole Beaudoin, Direction de la surveillance de l'état de santé, DGSP–MSSS
Mikaël Berthelot, Direction des statistiques de santé, ISQ
Robert Courtemanche, Direction de la méthodologie et de la qualité, ISQ
Régis Couture, Direction de santé publique du Saguenay–Lac-St-Jean (depuis juin 2008)
Suzanne Gingras, Infocentre de santé publique, INSPQ
Monique Godin, Direction de la surveillance de l'état de santé, DGSP–MSSS
Denis Hamel, Unité Connaissance-surveillance, INSPQ
Lyne Jobin, Direction de la surveillance de l'état de santé, DGSP–MSSS
Issouf Traoré, Direction des statistiques de santé, ISQ

Citation suggérée

INSTITUT DE LA STATISTIQUE DU QUÉBEC en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC et le MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX DU QUÉBEC. *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé de la population 2008*, Québec, Gouvernement du Québec, 2010, 117 p.

Avant-propos

Les données issues des enquêtes sociosanitaires constituent une source d'information cruciale pour une prise de décision éclairée dans le processus de planification sociosanitaire. Elles permettent de mesurer l'état de santé de la population et de ses déterminants et de contribuer à une meilleure compréhension de la performance et des impacts des programmes et des services sur l'état de santé. L'information provenant des enquêtes est complémentaire à celle fournie par les banques de données administratives. Souvent, seules les enquêtes sociosanitaires permettent de produire un portrait précis de l'état de santé de la population. Elles sont également une source d'information unique pour suivre l'évolution de l'état de santé dans le temps ou encore pour mesurer la qualité et l'appréciation des services rendus aux usagers.

Le ministère de la Santé et des Services sociaux (MSSS), en étroite collaboration avec l'Institut de la statistique du Québec (ISQ), a structuré une programmation québécoise d'enquêtes afin d'assurer une alimentation en données axée sur les besoins, notamment ceux de la planification sociosanitaire. Ainsi, le *Plan ministériel d'enquêtes sociales et de santé* vient préciser les orientations ministérielles en matière d'enquêtes sociosanitaires en lien avec la transformation du réseau et les obligations découlant du contexte législatif prescrit par la Loi sur la santé publique (L.R.Q., c. S-2.2) et la Loi sur les services de santé et les services sociaux (L.R.Q., c. S-4.2). Ce plan ministériel mise sur la valeur ajoutée des données d'enquêtes à des fins de gestion et de décision. En conséquence, il accorde une grande importance à l'alimentation en données et à l'accès, en temps opportun, à de l'information de qualité, utile à la prise de décision.

L'Enquête québécoise sur la santé de la population 2008 (EQSP) s'inscrit parmi cette nouvelle génération d'enquêtes planifiées à des fins de surveillance au *Plan ministériel d'enquêtes sociales et de santé*. Ces enquêtes ciblent, en premier lieu, la couverture des besoins liés aux plans de surveillance prescrits par la Loi sur la santé publique.

Plusieurs acteurs tant sur le plan national que régional et divers experts de contenu ont été associés de près aux différentes étapes de réalisation de l'EQSP. Dans la foulée des travaux visant à rehausser la capacité de surveillance au Québec, des efforts considérables de développement ont été consentis pour rendre disponible sur le Portail de l'Infocentre de santé publique l'information nécessaire à l'analyse de chaque indicateur mesuré à des fins de surveillance par l'EQSP.

Ce document vise à identifier les balises méthodologiques communes afin d'optimiser l'utilisation de l'information qui émane de l'EQSP, et ce, afin de mieux soutenir la prise de décision et d'informer la population sur son état de santé et ses déterminants.

*Le Comité tripartite MSSS-INSPQ-ISQ sur les aspects
méthodologiques des données d'enquêtes sociosanitaires*

Table des matières

AVANT-PROPOS	II
LISTE DES ANNEXES	VI
INTRODUCTION	1
I DESCRIPTION GÉNÉRALE DE L'ENQUÊTE QUÉBÉCOISE SUR LA SANTÉ DE LA POPULATION 2008 (EQSP)	3
1. Contexte général	3
2. Objectifs généraux et spécifiques de l'EQSP	3
3. L'accès aux données de l'EQSP	4
II ASPECTS MÉTHODOLOGIQUES ET STATISTIQUES	6
1. Plan d'échantillonnage et stratégie de collecte	6
1.1. Population cible	6
1.2. Base de sondage et méthode de sélection	6
1.3. Taille d'échantillon	8
1.4. Méthode de collecte	10
1.4.1. Stratégie de collecte	10
1.4.2. Instrument de collecte	12
1.4.3. Validation, saisie et codification	13
1.5. Taux de réponse et taux de déclaration par procuration	14
1.5.1. Taux de réponse	14
1.5.2. Taux de déclaration par procuration	17
2. Estimation des proportions et des effectifs	18
2.1. Définition	18
2.2. Pondération	19
2.2.1. Description générale des étapes	19
2.3. Non-réponse partielle	24
2.4. Imputation	25
2.4.1. Âge des membres des ménages échantillonnés	25
2.4.2. Imputation de la question 54 – Troubles musculo-squelettiques	25
2.4.3. Imputation de la question 120 – Revenu du ménage	26
2.5. Estimation des proportions	27
2.6. Estimation des effectifs	29
3. Mesures de précision pour les proportions	33
3.1. Estimation de la variance	33
3.1.1. Définition et méthodes	33
3.1.2. Impact de l'imputation sur la variance	34
3.2. Coefficient de variation	34
3.2.1. Définition et méthode de calcul	34
3.2.2. Recommandations sur les seuils de diffusion	35
3.3. Intervalle de confiance	35
3.3.1. Définition	35
3.3.2. Méthode de calcul pour les intervalles de confiance standards	36
3.3.3. Méthode de calcul pour les intervalles de confiance bootstrap	36
3.3.4. Recommandations relatives au choix du niveau de confiance	37
4. Ajustement des proportions selon l'âge	38
4.1. Définition	38
4.2. Méthode de calcul	38
4.3. Variance de proportions ajustées	39
5. Comparaison de proportions	39
5.1. Tests statistiques	39
5.1.1. Test global d'association	40
5.1.2. Test de la différence de deux proportions	40
6. Exploitation des données de l'EQSP	41
6.1. Comparabilité des résultats de l'EQSP avec d'autres sources de données	42

6.2.	Production de tableaux et comparaison des prévalences sur le portail de l'Infocentre	43
6.2.1.	Paramètres d'une requête paramétrable.....	43
6.2.2.	Algorithmes d'une requête paramétrable sur le portail pour les données de l'EQSP.....	47
6.2.3.	Exemples de résultats de requêtes paramétrables et recommandations.....	52
6.3.	Particularités des analyses produites à l'Infocentre.....	57
6.3.1.	Indicateur en répartition	57
6.3.2.	Nombre trop important de tests demandés par une requête.....	58
6.3.3.	Non-réponse partielle élevée et sous-estimation des effectifs.....	60
6.3.4.	Indice du niveau de revenu.....	63
6.3.5.	Diffusion par Réseaux locaux de services (RLS).....	64
6.3.6.	Point de césure de l'indice de détresse psychologique.....	66
6.3.7.	Indicateurs portant sur les enfants de 3 à 14 ans	67
6.3.8.	Indicateurs portant sur la période « au cours de la vie »	68
6.3.9.	Indicateur construit à partir de la médiane des scores	68
7.	Protection des renseignements personnels.....	69
7.1.	Confidentialité des fichiers de microdonnées.....	69
7.2.	Confidentialité des tableaux pour diffusion	70
III	RECOMMANDATIONS POUR L'INTERPRÉTATION DES RÉSULTATS	73
1.	Lien de causalité	73
2.	Inférence à la population visée	73
3.	Référence à un nombre	73
4.	Estimations et non des valeurs exactes	74
5.	Différence statistiquement non significative.....	74
6.	Présentation de résultats sous forme de ratio	75
7.	Présentation de résultats selon l'ordre des estimations	75
8.	Tendance.....	76
	RÉFÉRENCES BIBLIOGRAPHIQUES	77
	ANNEXE.....	79

Liste des annexes

- Annexe 1 Lettres d'information expédiées aux ménages correspondant aux numéros de téléphone échantillonnés pour lesquels une adresse a pu être obtenue
- Annexe 2 Information supplémentaire sur les taux de réponse
- Annexe 3 Comment utiliser la pondération lors de la production d'estimations à partir des données de l'EQSP
- Annexe 4 Étape de calage aux marges dans la pondération de l'EQSP
- Annexe 5 Diagnostic et analyse de la non-réponse partielle pour les données de l'EQSP
- Annexe 6 Information supplémentaire – imputation des questions 54 et 120 de l'EQSP
- Annexe 7 Exemple de calcul pour l'estimation d'un effectif
- Annexe 8 Information supplémentaire concernant la méthode du bootstrap
- Annexe 9 Intervalle de confiance binomial
- Annexe 10 Niveau de confiance
- Annexe 11 Exemple de calcul d'une proportion ajustée selon l'âge, et dans le cas où l'utilisation des poids bootstrap n'est pas possible, de sa variance
- Annexe 12 Tests statistiques – un bref survol
- Annexe 13 Validité du test du khi-deux
- Annexe 14 Tests de la différence de proportions – deux méthodes
- Annexe 15 Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre

Introduction

Le Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé de la population 2008 a été élaboré dans le cadre des travaux du Comité tripartite MSSS-INSPQ-ISQ sur les aspects méthodologiques des données d'enquêtes sociosanitaires. Mis en place en janvier 2006, ce comité a pour mandat d'identifier des façons de faire et des balises communes sur les aspects méthodologiques liés à l'utilisation des données d'enquêtes sociosanitaires aux fins des activités de surveillance.

La rédaction de ce guide s'inscrit dans la foulée des travaux du cadre méthodologique du projet du *Plan commun de surveillance à l'Infocentre de santé publique*. Il vise à soutenir les utilisateurs dans l'analyse et l'interprétation des résultats des indicateurs d'enquêtes sociosanitaires du plan commun accessibles sur le portail de l'Infocentre de santé publique (Infocentre) en fournissant l'information nécessaire pour :

- 1) expliquer les balises méthodologiques utilisées dans la production des résultats;
- 2) interpréter les mesures de précision et les résultats des tests statistiques offerts sur le portail;
- 3) choisir l'option méthodologique répondant le mieux à leurs besoins.

Ce guide porte de façon particulière sur les indicateurs du *Plan commun de surveillance de l'état de santé de la population et de ses déterminants* (PCS) dont la source est l'*Enquête québécoise sur la santé de la population 2008* (EQSP). L'EQSP est la seconde série d'enquêtes¹ à être traitée à l'Infocentre dans le cadre de la mise en œuvre du PCS.

Ce document a été rédigé en s'inspirant du « *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête sur la santé dans les collectivités canadiennes cycles 1.1, 2.1 et 3.1* » de l'Institut de la statistique du Québec, duquel certains éléments ont été tirés. La première partie du document donne une description générale de l'EQSP. La seconde, qui porte sur les aspects méthodologiques de l'enquête, est divisée de la façon suivante : la première section décrit la méthodologie générale et la stratégie de collecte utilisées dans l'EQSP. La section suivante présente les éléments méthodologiques utilisés pour produire les estimations des proportions et des effectifs. Les mesures de précision des proportions ainsi que l'ajustement de celles-ci selon l'âge sont respectivement abordés aux sections 3 et 4, alors que la section 5 traite de la comparaison des proportions. Les sections 6 et 7 portent respectivement sur l'exploitation des données de l'EQSP à l'Infocentre et sur la protection des renseignements personnels des répondants à l'EQSP. Finalement, des recommandations pour l'interprétation des résultats sont proposées dans la troisième et dernière partie du document.

Le *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires* du Plan commun de surveillance – *Enquête québécoise sur la santé de la population 2008* se veut un outil **dynamique** de soutien à l'utilisateur. Il est ainsi appelé à s'enrichir et à évoluer en fonction, notamment, des commentaires des utilisateurs, des prochaines éditions de l'EQSP et du développement de nouvelles méthodologies.

¹ La première étant l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC) de Statistique Canada.

I Description générale de l'Enquête québécoise sur la santé de la population 2008 (EQSP)

1. Contexte général

Au milieu des années 1980, le ministère de la Santé et des Services sociaux du Québec (MSSS) a confié à Santé Québec la réalisation d'une série d'enquêtes générales de santé à portée régionale. L'Enquête Santé Québec de 1987 fut la première à être réalisée, suivie des deux Enquêtes sociales et de santé de 1992-1993 et de 1998. Ainsi, durant plus d'une décennie, des données de qualité ont été disponibles au niveau régional, lesquelles ont fait l'objet de plusieurs analyses et comparaisons, dans l'espace et dans le temps.

Après trois éditions, cette série d'enquêtes quinquennales a été interrompue, Statistique Canada implantant au début des années 2000 l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC). Cette enquête transversale est constituée de cycles généraux (2000-2001, 2003, 2005, 2007-2008, 2009-2010) représentatifs à l'échelon régional dans chacune des provinces et dans les territoires. À partir de ce moment, l'ESCC a comblé une partie des besoins d'information du ministre de la Santé et des Services sociaux sur la santé de la population du Québec.

Les besoins en information du MSSS ont évolué au milieu des années 2000 dans la foulée de l'adoption de la Loi sur la santé publique (LSP) en 2001 qui est venue baliser l'exercice de la surveillance continue de l'état de santé de la population du Québec. La LSP requiert que les objets de surveillance et leurs sources, notamment, soient inscrits dans un plan de surveillance. Le premier plan de surveillance adopté par le ministre et les directeurs de santé publique est le *Plan commun de surveillance de l'état de santé de la population et de ses déterminants 2004-2007*².

Plusieurs objets de surveillance du PCS ont pour source de données l'ESCC, mais pas tous. Par ailleurs, la présence de certains des indicateurs de l'ESCC inscrits au PCS n'est pas assurée dans le contenu commun de cette enquête, ou encore, la précision à l'échelon régional n'est pas satisfaisante. Afin de répondre, notamment, à ces lacunes ainsi qu'à d'autres besoins en surveillance et à d'autres besoins ministériels, le MSSS s'est doté d'un nouveau programme d'enquêtes, le *Plan ministériel d'enquêtes sociales et de santé*. L'EQSP est une enquête transversale à portée régionale qui s'inscrit dans ce programme d'enquêtes. Complémentaire à l'ESCC, elle devrait être répétée tous les cinq ans. Le MSSS a confié sa réalisation par mandat à l'Institut de la statistique du Québec (ISQ), comme le prévoit l'article 42 de la LSP.

2. Objectifs généraux et spécifiques de l'EQSP

L'objectif général de l'enquête est d'alimenter en données les objets de surveillance du PCS pour contribuer à l'atteinte des objectifs de la surveillance.

Les objectifs spécifiques de l'EQSP sont de :

- Recueillir l'information nécessaire à la mesure des indicateurs suivants du PCS³ :
 - rhinite allergique
 - prise de tension artérielle

² Directions de santé publique, Ministère de la Santé et des Services sociaux, en collaboration avec l'Institut national de santé publique du Québec. *Plan commun de surveillance de l'état de santé de la population et de ses déterminants 2004-2007*, Québec, Gouvernement du Québec, février 2005, 599 p.

³ Ces indicateurs sont non couverts ou mal couverts par les sources de données existantes et ont été retenus parmi environ 80 indicateurs pour lesquels l'enquête constitue la meilleure source. Ils ont été considérés comme prioritaires par un Comité d'orientation de projet (COP) composé de représentants des directions de santé publique régionales, des tables de concertation nationales en santé publique, de l'ISQ, de l'INSPQ et du MSSS.

- test Pap
 - poids et image corporelle
 - santé et hygiène buccodentaire
 - blessures dues et non dues à des mouvements répétitifs
 - stress au travail et exposition à certains agresseurs ou situations à risque
 - troubles musculo-squelettiques
 - état de santé psychosociale et mentale y compris les tendances suicidaires
 - consommation de drogues
 - contraception et ITSS (infections transmissibles sexuellement et par le sang)
 - conditions socioéconomiques
- Obtenir des données précises à l'échelle régionale;
 - Documenter certains indicateurs sur la santé des enfants, le thème retenu comme prioritaire étant le développement psychosocial.

3. L'accès aux données de l'EQSP

L'accès aux résultats de l'EQSP peut se faire par l'exploitation de trois ensembles de données, soit le fichier maître, le fichier masqué contre l'identification involontaire (FMII) et le fichier aux fins d'analyse et de recherche externe (FARE).

Le **fichier maître** contient l'ensemble de l'information recueillie sur les répondants pour chacune des seize régions sociosanitaires couvertes par l'enquête⁴. Lors de la collecte des données, le répondant était informé que les réponses qu'il fournirait à l'enquête seraient partagées avec le MSSS et l'Institut national de santé publique (INSPQ). Le fichier maître est donc disponible à trois endroits, soit à l'ISQ, au MSSS et à l'INSPQ. L'Infocentre à l'INSPQ l'exploite pour le déploiement des indicateurs de surveillance sur son portail informationnel. Le MSSS l'exploite pour répondre aux besoins ministériels particuliers d'analyse aux fins de surveillance et de planification sociosanitaire. Alors que l'ISQ, l'organisme responsable du fichier, l'exploite pour ses besoins (rapports, publications diverses) tout en favorisant l'accès aux données à des chercheurs par le biais de son centre d'accès aux données de recherche de l'ISQ (CADRISQ). Il s'agit en fait du seul accès possible au fichier maître pour les chercheurs.

Le **FMII** est un fichier masqué qui peut être exploité par le biais d'un accès à distance octroyé par l'ISQ. Le FMII contient l'ensemble des répondants de l'EQSP. Pour quelques répondants, toutefois, certaines données ont dû être masquées dans le but de protéger leur identité. En effet, le type de masquage appliqué au FMII a pour but de protéger les répondants contre l'identification involontaire des individus (voir la section 7.1 pour des explications supplémentaires).

Le **FARE** est également un fichier masqué, lequel peut être exploité directement dans les locaux d'un organisme public. Tout comme le FMII, le FARE contient l'ensemble des répondants de l'EQSP. Toutefois, la quantité d'information disponible dans le FARE peut être légèrement inférieure à celle disponible dans le FMII en raison d'un masquage plus sévère appliqué au FARE (voir la section 7.1 pour des explications supplémentaires).

Le tableau 1-1 présente certaines caractéristiques des fichiers, dont une mesure du potentiel analytique des fichiers masqués par rapport au fichier maître. La mesure retenue pour ce tableau est appelée la « mesure d'association – p-value ». Elle renseigne sur le pourcentage de conclusions des tests ayant changées, au seuil de 5 %, par rapport aux conclusions obtenues à partir du fichier maître. Cette mesure est évaluée sur le sous-ensemble des variables affectées par le masquage des données. Or, la valeur de

⁴ Seules les régions du Nunavik (17) et des Terres-Cries-de-la-Baie-James (18) ne sont pas couvertes par l'enquête.

cette même mesure, évaluée sur l'ensemble des variables du fichier, serait évidemment inférieure à celle présentée dans ce tableau. En fait, celle-ci serait tout près de zéro étant donné le nombre très important de variables du fichier qui ne sont pas affectées par le masquage.

Le lecteur pourra obtenir plus d'information sur le potentiel analytique à la section 7.1, alors que les détails sur les variables du fichier maître sont disponibles dans le cahier technique de l'enquête (Traoré et autres, 2010).

Tableau 1-1 Caractéristiques de chacun des fichiers de l'EQSP 2008

Fichier	Nombre de répondants	Nombre de variables				Mesure du potentiel analytique ¹
		Affectées par le masquage	Supprimées du fichier	Recatégorisées	Pour lesquelles des valeurs ont été supprimées	
Fichier maître (Fichier non masqué)	38 154	0	0	0	0	0 %
FMII (Fichier masqué contre l'identification involontaire)	38 154	60	23	4	33	0,4 %
FARE (Fichier masqué aux fins d'analyse et de recherche)	38 154	67	40	8	19	1,8 %

1. Il existe plusieurs mesures du potentiel analytique. Celle présentée dans ce tableau, la « mesure d'association – p-value », représente le pourcentage de conclusions des tests ayant changées au seuil de 5 % par rapport aux conclusions obtenues à partir du fichier maître.

Sources : Fichier maître de l'*Enquête québécoise sur la santé de la population 2008*.

Fichier masqué contre l'identification involontaire de l'*Enquête québécoise sur la santé de la population 2008*.

Fichier aux fins d'analyse et de recherche externe de l'*Enquête québécoise sur la santé de la population 2008*.

II Aspects méthodologiques et statistiques

Cette partie du guide présente la méthodologie générale de l'Enquête québécoise sur la santé de la population 2008.

Il importe de bien connaître les éléments méthodologiques d'une enquête afin de faire une utilisation adéquate des données. En effet, la connaissance du plan de sondage de l'enquête, des sous-populations visées par les indicateurs, de la taille et la répartition de l'échantillon sélectionné, de la procédure de collecte en général, de la validation des données et de leur codification, de la stratégie de pondération et du traitement de la non-réponse partielle est essentielle, car ceux-ci ont une incidence directe sur les estimations produites. La connaissance de ces éléments permettra à l'utilisateur des données de l'EQSP de procéder à des analyses adéquates qui permettront aux décideurs, à terme, de mettre en place des politiques plus appropriées et mieux adaptées aux populations visées.

Précisons qu'en général, les tableaux exposés dans cette partie du guide sont issus du fichier maître et présentent les données, pour l'ensemble du Québec et pour les régions.

1. Plan d'échantillonnage et stratégie de collecte

Le plan d'échantillonnage est l'un des trois éléments du plan de sondage. Les deux autres éléments, les méthodes d'estimation et les méthodes de mesure de la précision sont traités respectivement aux sections 2 et 3.

Le plan d'échantillonnage englobe la description des unités qui constituent la population cible, la base de sondage dans laquelle sont sélectionnés les individus à enquêter, la méthode de sélection de ces individus, ainsi que la taille de l'échantillon à sélectionner. La partie collecte des données inclut les méthodes de collecte utilisées et certains résultats de collecte, notamment les taux de réponse et les taux de déclaration par procuration. Tous ces éléments sont discutés dans la présente section.

1.1. Population cible

La population visée par l'EQSP est composée de l'ensemble des personnes âgées de 15 ans et plus qui vivent dans un ménage privé au Québec. Ne font pas partie de l'enquête les personnes vivant dans un ménage collectif et les personnes vivant sur une réserve amérindienne. L'enquête couvre 16 des 18 régions sociosanitaires (RSS) formant le Québec. Seules les régions du Nunavik et des Terres-Criées-de-la-Baie-James (régions sociosanitaires 17 et 18) ne sont pas couvertes par l'enquête. Les modes de vie et de culture particuliers de ces populations commandent des études séparées. La population visée par l'EQSP représente 6 326 523 Québécois de 15 ans et plus.

Une seconde population d'intérêt était visée par l'EQSP, soit les enfants âgés de 3 à 14 ans vivants dans les ménages privés⁵, qui représente 974 579 enfants pour l'ensemble du Québec.

1.2. Base de sondage et méthode de sélection

Pour répondre à l'un des objectifs spécifiques de l'enquête, qui est d'obtenir des statistiques précises pour chacune des régions sociosanitaires, le territoire québécois a été divisé selon le découpage sociosanitaire officiel comme établi par le MSSS.

La base de sondage de l'enquête a été constituée à l'aide d'une génération aléatoire de numéros de téléphone (GANT), une technique qui permet d'atteindre des ménages privés admissibles à l'enquête. Il existe plusieurs méthodes de génération aléatoire de numéros de téléphone. Afin d'augmenter la

⁵ La définition d'un ménage privé est la même que celle de la population principale des 15 ans et plus.

productivité de l'échantillon – autrement dit le taux de contact –, la méthode de l'élimination des banques non valides a été retenue pour l'enquête.

Une banque est constituée des huit premiers chiffres du numéro de téléphone incluant l'indicatif régional. Il y a 100 numéros dans une banque. Une banque est dite valide si elle contient au moins un numéro résidentiel listé dans un répertoire téléphonique; les banques non valides sont éliminées. C'est l'ensemble des banques valides de 100 numéros de téléphone qui sert de base de sondage. La base de sondage de l'EQSP couvre environ 92 %⁶ de la population québécoise âgée de 15 ans et plus.

La sélection⁷ de l'échantillon de l'EQSP est faite selon un plan de sondage stratifié, les divisions de recensement (DR)⁸ formant les strates; l'échantillon est ensuite regroupé pour former les régions sociosanitaires. La répartition de l'échantillon de base de l'EQSP n'est pas proportionnelle au nombre de banques valides dans les régions, et ce, afin de fournir des estimations fiables par région sociosanitaire. Les unités de l'échantillon sont donc réparties de façon égale entre les régions. Les unités de l'échantillon supplémentaire (voir section 1.3) ont quant à elles été sélectionnées dans les régions qui en ont fait la demande.

Pour chaque banque valide échantillonnée, un nombre entre 00 et 99 a été généré aléatoirement afin de constituer un numéro de téléphone potentiel. Pour chacun des numéros de l'échantillon, l'intervieweur devait vérifier, à l'aide de questions préalablement définies, si le numéro correspondait à un ménage privé admissible. Si ce n'était pas le cas, il rejetait le numéro de téléphone. Celui-ci était alors classé comme inadmissible (ex. : numéro associé à une résidence secondaire) ou non valide (ex. : numéro non attribué). Si le numéro correspondait à un ménage privé, l'intervieweur dressait la liste des personnes admissibles dans le ménage, soit celles de 15 ans et plus, à partir de l'information donnée par la personne rejointe au téléphone. Le processus était complété lorsqu'une seule personne par ménage était sélectionnée aux fins de l'enquête. Dans l'EQSP, toutes les personnes admissibles dans le ménage avaient la même probabilité d'être sélectionnées.

Sélection des enfants

Comme mentionné à la section 1.1, en plus de la population principale des personnes de 15 ans et plus, une seconde population d'intérêt était visée par l'EQSP, soit les enfants de 3 à 14 ans vivants dans un ménage privé au Québec. Ainsi, dans le cas où le répondant est le parent⁹ d'au moins un enfant de 3 à 14 ans habitant dans le ménage, une sélection aléatoire d'un de ses enfants est effectuée afin que le parent réponde à des questions sur le développement psychosocial de son enfant. Pour ce faire, vers la fin du questionnaire, tout juste avant les questions sociodémographiques, une série de questions prédéfinies permettaient d'établir les caractéristiques des enfants du ménage dans le but de faire cette sélection. Bien entendu, pour un enfant donné, la présence de ses deux parents dans le ménage augmentait ses chances d'être sélectionné. Cet élément a été pris en compte *a posteriori* lors d'une des étapes de la pondération de l'enquête (section 2.2.1). Afin de ne pas surreprésenter dans l'échantillon les enfants vivants dans les ménages où il y a un seul enfant, un ménage sur deux seulement était retenu dans ce cas pour la sélection d'un enfant. Dans les ménages où il y a au moins deux enfants, un seul enfant par ménage était sélectionné aléatoirement.

⁶ Selon l'*Enquête sur le service téléphonique résidentiel* (ESTR) 2008 de Statistique Canada, 0,9 % des ménages québécois n'ont aucun service téléphonique à la maison et près de 6,7 % des ménages ont uniquement un cellulaire. À cela pourrait s'ajouter l'absence possible des nouvelles banques valides en raison des délais pour la mise à jour des banques, mais cette proportion est considérée négligeable.

⁷ Ce travail a été confié à la firme ASDE par l'ISQ.

⁸ La division de recensement est une unité géographique plus fine que la région définie par Statistique Canada aux fins du recensement canadien. Les divisions de recensement utilisées pour l'EQSP 2008 sont celles du recensement canadien 2006.

⁹ On entend par « parent », les parents biologiques, les beaux-parents, les parents adoptifs et les tuteurs légaux.

1.3. Taille d'échantillon

La taille de l'échantillon visé et sa répartition sont fixées de façon à produire des estimations suffisamment précises pour les besoins de l'enquête, tout en tenant compte des ressources financières disponibles pour sa réalisation et de son échéancier. Le nombre de répondants attendus pour l'EQSP a été fixé à 2 011 pour chacune des régions sociosanitaires couvertes par l'enquête, ce qui permet d'estimer des facteurs assez peu répandus (proportions égales ou supérieures à 2,8 %¹⁰) avec une bonne précision relative, c'est-à-dire un coefficient de variation inférieur ou égal à 15 %. Le tableau 1-2 présente la prévalence minimale régionale que l'on peut s'attendre à estimer avec une bonne précision (C.V. ≤ 15 %), pour un échantillon régional de 2011 répondants visés, selon le groupe d'âge et le sexe. L'objectif de précision pour les estimations portant sur la population des enfants de 3 à 14 ans se limite à un objectif régional sans croisement. Celui-ci est également présenté au tableau 1-2.

Tableau 1-2 Prévalence minimale que l'on peut s'attendre à estimer avec une bonne précision (C.V. ≤ 15 %), pour un échantillon régional de 2 011 répondants visés, selon le groupe d'âge et le sexe

Domaine d'étude	Prévalence minimale
Âge et sexe confondus (15 ans et plus)	2,8 %
Homme ou femme (15 ans et plus)	5,4 %
3-14 ans	17,5 %
15-44 ans	5 %
15-24 ans	20 %
25-44 ans	7 %
45 ans et plus	6 %
45-64 ans	9 %
65 ans et plus	15 %
Femme de 15-44 ans	10 %
Homme de 40 ans et plus	9 %
Homme ou femme de 60 ans et plus	19 %
Homme ou femme de 65 ans et plus	25 %

Source : Institut de la statistique du Québec, EQSP 2008.

Compte tenu de certains besoins régionaux, l'ISQ et le MSSS ont offert aux directions de santé publique (DSP) régionales la possibilité d'acheter des unités d'échantillon supplémentaires afin de permettre l'obtention d'estimations fiables à l'échelle des réseaux locaux de services de santé et de services sociaux (RLS)¹¹. Six DSP se sont prévaluées de cette offre. Il s'agit des DSP des Agences de la santé et des services sociaux (ASSS) suivantes :

- ASSS du Saguenay–Lac-Saint-Jean (6 RLS);
- ASSS de la Mauricie et Centre-du-Québec (8 RLS);
- ASSS de l'Abitibi-Témiscamingue (6 RLS);

¹⁰ Cette proportion a été établie après une évaluation des différentes proportions des principaux indicateurs du PCS.

¹¹ « En décembre 2003, le gouvernement du Québec adoptait la Loi sur les agences de développement de réseaux locaux de services de santé et de services sociaux. Cette loi confiait aux agences la responsabilité de mettre en place un nouveau mode d'organisation des services dans chaque région basé sur des réseaux locaux de services. La création, en juin 2004, de 95 réseaux locaux de services à l'échelle du Québec a pour objectif de rapprocher les services de la population et de les rendre plus accessibles, mieux coordonnés et continus. Au cœur de chacun de ces réseaux locaux de services, on trouve un nouvel établissement appelé centre de santé et de services sociaux (CSSS) né de la fusion de centres locaux de services communautaires (CLSC), de centres d'hébergement et de soins de longue durée (CHSLD) et dans la majorité des cas, d'un centre hospitalier. » Extrait pris sur le site Internet du MSSS (<http://www.msss.gouv.qc.ca/reseau/rls/>).

- ASSS de la Gaspésie–Îles-de-la-Madeleine (5 RLS);
- ASSS de la Chaudière-Appalaches (5 RLS);
- ASSS de Lanaudière (2 RLS).

Outre la région de Lanaudière, dont la répartition prévue de l'échantillon régional permettait d'emblée l'atteinte des objectifs de précision infrarégionaux¹², la taille de l'échantillon des autres régions a été augmentée afin d'obtenir un nombre suffisant de répondants par RLS. L'ISQ a fixé ce nombre à 450 répondants par RLS afin d'obtenir des estimations fiables (prévalence $\geq 11,5\%$, C.V. $\leq 15\%$) pour chacun des RLS de ces régions. La DSP du Saguenay–Lac-Saint-Jean, souhaitant augmenter la précision de ses estimations, a plutôt opté pour une taille d'échantillon visant 650 répondants par RLS. Cette augmentation leur permettrait d'estimer une prévalence plus faible, de l'ordre de 8,5 %, avec la même précision.

Lorsque l'on considère l'échantillon supplémentaire nécessaire à la production d'estimations locales pour les six régions qui s'en sont prévalues, le nombre de répondants total attendu pour l'ensemble du Québec atteint 38 130. La répartition régionale du nombre de répondants visé, avec et sans l'échantillon supplémentaire, est présentée au tableau 1-3.

Afin d'atteindre la cible de 38 130 répondants visés, un échantillon de plus de 141 000 numéros de téléphone a été généré aléatoirement. Ce nombre a été calculé en faisant différentes hypothèses sur le taux de productivité¹³, le taux d'admissibilité¹⁴ et le taux de réponse¹⁵ anticipés de l'enquête.

Au moment d'établir le plan de sondage de l'EQSP, l'ISQ a évalué que l'effet de plan¹⁶ serait d'environ 1,3¹⁷ pour le niveau régional, ce qui mène à un effet de plan pour l'ensemble du Québec d'environ 2,7. Par conséquent, bien que la taille de l'échantillon de base attendue soit de 2 011 répondants par région (excluant l'échantillon supplémentaire), la taille de l'échantillon régional efficace est d'au plus 1 547 répondants.

¹² Bien que leur région n'ait pas nécessité l'ajout de répondants, les représentants de l'ASSS de Lanaudière ont demandé que la méthodologie développée pour leur région permette la production d'estimations à l'échelle locale.

¹³ Ce taux est défini par la proportion de numéros de téléphone correspondant à des ménages parmi les numéros générés.

¹⁴ Ce taux renseigne sur la proportion de ménages privés admissibles à l'enquête parmi l'ensemble des ménages. Un ménage est admissible à l'enquête s'il comprend au moins une personne de 15 ans ou plus, à l'exclusion des ménages dans les réserves amérindiennes et des ménages collectifs.

¹⁵ Ce taux correspond à la proportion de questionnaires complétés parmi les cas admissibles.

¹⁶ L'effet de plan est une mesure relative de la précision des estimations produites avec le plan de sondage choisi comparativement à la précision qu'on aurait eu avec un plan aléatoire simple contenant le même nombre d'unités. Un effet de plan supérieur à 1 est associé à une moins bonne précision que celle obtenue avec un échantillon aléatoire simple de même taille.

¹⁷ L'évaluation a été faite sur l'échantillon de base seulement.

Tableau 1-3 Nombre de répondants visé, avec et sans l'échantillon supplémentaire, selon la région sociosanitaire, Québec

Région sociosanitaire	Nombre de répondants visé		
	De base	Supplémentaire	Total
Bas-Saint-Laurent	2 011	-	2 011
Saguenay – Lac-Saint-Jean	2 011	1 902	3 913
Capitale Nationale	2 011	-	2 011
Mauricie – Centre-du-Québec	2 011	1 676	3 687
Estrie	2 011	-	2 011
Montréal	2 011	-	2 011
Outaouais	2 011	-	2 011
Abitibi – Témiscamingue	2 011	950	2 961
Côte-Nord	2 011	-	2 011
Nord-du-Québec	2 011	-	2 011
Gaspésie – Îles-de-la-Madeleine	2 011	501	2 512
Chaudière – Appalaches	2 011	925	2 936
Laval	2 011	-	2 011
Lanaudière	2 011	0	2 011
Laurentides	2 011	-	2 011
Montérégie	2 011	-	2 011
Ensemble du Québec	32 176	5 954	38 130

Source : Institut de la statistique du Québec, EQSP 2008.

1.4. Méthode de collecte

1.4.1. Stratégie de collecte

La collecte des données de l'EQSP, qui s'est déroulée du 13 février 2008 au 15 mars 2009, a été réalisée au moyen d'interviews téléphoniques assistées par ordinateur (ITAO). L'ITAO procure un certain nombre d'avantages quant à la qualité des données par rapport aux autres méthodes de collecte. Premièrement, le libellé des questions, comprenant les périodes de référence et les pronoms, est personnalisé automatiquement en fonction de facteurs comme l'âge et le sexe du répondant, de la date de l'entrevue et des réponses aux questions précédentes. En second lieu, des mesures de contrôle isolent les réponses incohérentes ou hors-normes, et des instructions apparaissent à l'écran lorsqu'une telle situation se présente. Le répondant reçoit une rétroaction immédiate et l'intervieweur peut corriger toute incohérence. Troisièmement, le processus permet au répondant de sauter automatiquement les questions qui ne le concernent pas.

Pour faciliter la gestion et le suivi sur le terrain, l'échantillon a été scindé en cinq vagues de collecte. La répartition visée de l'échantillon, ainsi que la répartition réalisée de celui-ci, sont présentées au tableau 1-4. Les deux premières vagues sont moins volumineuses pour deux raisons : premièrement, le lancement de plus petites vagues en début de collecte a permis de confirmer certaines hypothèses émises au moment de la planification de l'enquête, et deuxièmement, les deux premières vagues ont servi à la réalisation d'une étude sur l'impact de l'envoi d'une lettre d'information au répondant (voir plus loin dans la section).

À chaque vague, avant d'amorcer la collecte téléphonique, les numéros de téléphone ont été précomposés¹⁸, ce qui a permis d'exclure des opérations de collecte subséquentes certains numéros jugés

¹⁸ Lors de la précomposition, chacun des numéros de téléphone de l'échantillon est validé au moyen du signal émis lors de l'appel, ce qui permet de reconnaître les numéros qui ne sont pas en service. Ce travail a été confié à l'Échantillonneur ASDE par l'ISQ.

non valides. Ainsi, 45 624 numéros de téléphone de l'échantillon initial ont été exclus et c'est donc 95 654 numéros de téléphone qui ont été retenus pour la collecte de données. À cette étape, pour chacun des numéros de téléphone de l'échantillon, une recherche des noms et adresses dans des versions électroniques d'annuaires téléphoniques a été effectuée. Cette opération a permis d'associer un nom et une adresse complète à 50 102 numéros de téléphone, soit 52,3 % de la partie de l'échantillon (95 654) envoyée à la collecte de données après la précomposition.

Tableau 1-4 Répartition visée et réalisée de l'échantillon selon la vague de collecte

Vague	Répartition visée	Échantillon réalisé
Vague 1	1/8	14 746
Vague 2	1/8	15 061
Vague 3	1/4	33 986
Vague 4	1/4	40 677
Vague 5	1/4	36 808
Total		141 278

Source : Institut de la statistique du Québec, EQSP 2008.

Bien qu'il s'agisse d'une enquête ayant recours à une technique de GANT, la stratégie de collecte retenue comportait l'expédition d'une lettre présentant l'enquête aux ménages correspondant aux numéros de téléphone échantillonnés pour lesquels une adresse de résidence avait pu être obtenue. Cette lettre, expédiée avant chacune des vagues, avait pour but de les informer qu'un membre du personnel de l'ISQ allait communiquer avec eux prochainement afin de réaliser une entrevue téléphonique. Les lettres utilisées pour l'EQSP sont jointes à l'annexe 1.

En cas d'absence ou de refus¹⁹, la stratégie de collecte prévoyait aussi plusieurs tentatives afin de joindre ou de convaincre les ménages de participer à l'EQSP. Par ailleurs, au moment de lancer la cinquième vague de collecte, une version modifiée de la lettre d'information a été envoyée à certains ménages non-répondants des trois premières vagues, pour des régions ciblées. Cette lettre visait à leur faire prendre conscience de la nécessité d'atteindre les objectifs spécifiques de leur région quant au nombre de répondants. La même lettre, adaptée pour certaines régions, a été utilisée pour l'échantillon de la dernière vague afin de favoriser l'obtention du meilleur taux de réponse possible, d'autant plus que la période de temps pour convaincre les non-répondants à participer à l'enquête est plus courte pour cette vague²⁰.

Plusieurs facteurs peuvent influencer la propension à répondre des ménages et des individus échantillonnés. Afin d'évaluer l'efficacité de la stratégie consistant à envoyer une lettre d'information aux ménages échantillonnés, les deux premières vagues de collecte de l'EQSP ont été planifiées de façon à permettre la réalisation d'une étude, dans un environnement contrôlé, ayant pour but de mesurer l'impact de cette lettre sur le taux de réponse à l'enquête. Le choix de cette enquête pour cette étude s'imposait du fait qu'il s'agit d'une enquête réalisée auprès d'une population générale et que son envergure permette l'utilisation d'une partie de son échantillon, afin de contrôler les résultats, sans avoir trop d'impact sur le taux de réponse global de l'enquête. Cette étude était d'autant plus nécessaire puisque une méthode GANT était utilisée pour l'enquête, et que par définition, la proportion d'unités pour lesquelles il n'est pas possible d'obtenir une adresse complète est non négligeable pour ce type de méthode.

¹⁹ Pour des raisons éthiques, les cas de refus ferme n'ont pas fait l'objet de relance.

²⁰ Une des stratégies employées par l'ISQ est d'espacer dans le temps les rappels faits aux individus ayant refusé de participer à l'enquête (à l'exception des refus fermes qui ne sont pas recontactés). Cet espacement est plus limité pour la dernière vague de collecte, comparativement aux autres vagues.

L'utilisation de lettres d'information s'est avérée une stratégie efficace en vue d'améliorer la participation des personnes sollicitées. Les principaux constats de l'étude sur l'incidence des lettres d'information aux répondants sont les suivants :

- L'envoi d'une lettre d'information permet d'obtenir un taux de réponse jusqu'à 7,8 % plus élevé;
- L'incidence positive de la lettre s'observe tant chez les hommes que chez les femmes;
- Les plus influencés par la lettre sont les personnes de 50 ans et plus. Les jeunes sont peu influençables par des incitatifs légers tels que la lettre;
- L'incidence positive s'étend également au refus, avec une baisse significative de 5,0 % du taux de refus lorsqu'on expédie une lettre d'information.

Pour plus de détails sur cette étude, consulter Baulne et Courtemanche (2008).

1.4.2. Instrument de collecte

Le choix des questions et des thèmes abordés dans l'EQSP est issu des travaux du Comité d'orientation de projet (COP) créé au printemps 2007 aux fins de l'EQSP. Le COP, sous la responsabilité partagée du MSSS et de l'ISQ, est composé de représentants des directions de santé publique régionales, des tables de concertation nationales en santé publique et de l'INSPQ²¹. Une consultation sur les indicateurs du PCS à couvrir en priorité par l'EQSP avait été préalablement conduite par le MSSS auprès de ces partenaires du réseau. Cet exercice a alimenté, dès le départ, les travaux du COP.

Les critères retenus pour identifier les indicateurs à mesurer se résument principalement à :

- la nécessité d'obtenir une mesure pour un indicateur du PCS non couvert par une autre source. Cette nécessité était établie en fonction de priorités ministérielles ou dans le cadre de la mise en œuvre de certains programmes de santé et de services sociaux, notamment le Programme national de santé publique (PNSP);
- la possibilité d'obtenir ou de rehausser la précision d'un indicateur du PCS à l'échelle des régions;
- la complémentarité avec d'autres sources de données;
- la périodicité souhaitée des mesures;
- la cohérence de l'ensemble du questionnaire;
- les besoins d'analyse.

Les thématiques retenues à la fin de ce processus, donc couvertes par l'EQSP, sont présentées au tableau 1-5.

Le questionnaire final compte treize sections. Celles-ci sont présentées au tableau 1-6, ainsi que le nombre de questions posées pour chacune d'elles. Enfin, mentionnons qu'un questionnaire français et un anglais ont été préparés pour l'enquête²².

²¹ Le mandat du comité est de conseiller l'ISQ dans la réalisation de l'EQSP afin que celle-ci réponde adéquatement aux besoins d'information du ministre, du réseau de la santé et des services sociaux et de ses partenaires.

²² Ceux-ci sont disponibles sur le site internet de l'ISQ et sur le portail de l'Infocentre.

Tableau 1-5 Thématiques couvertes par l'enquête

Rhinite allergique
Prise de la tension artérielle
Cytologie (test Pap)
Poids et image corporelle
Santé buccodentaire
Blessures dues à des mouvements répétitifs et non dues à des mouvements répétitifs
Stress au travail et exposition à certains agresseurs ou situations à risque
Troubles musculo-squelettiques
État de santé psychosociale et mentale
Consommation de drogue
Contraception et comportements à risque pour les itss (infections transmises sexuellement et par le sang)
Conditions socioéconomiques

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau 1-6 Nombre de questions posées dans le questionnaire selon la section

Section	Nombre de questions posées
Santé générale et allergies	16
Poids et perception de son apparence physique	33
Hygiène et santé buccodentaire	5
Blessures, mouvements répétitifs	3
Blessures, autres	13
Admissibilité à l'enquête limitations d'activités*	8
Santé au travail	29
Troubles musculo-squelettiques (TMS)	13
Bien-être et santé mentale	7
Tendances suicidaires	6
Habitudes de vie (tabac, drogues, sexualité)	36
Santé des enfants (développement psychosocial)	34
Profil sociodémographique	15
Ensemble des sections	218

* Cette section avait pour but de déterminer l'admissibilité des répondants à une seconde enquête portant sur les limitations d'activités. Or, cette enquête a été reportée au tout début de la collecte de l'EQSP. Toutefois, les questions ont tout de même été posées aux répondants de l'EQSP.

Source : Institut de la statistique du Québec, EQSP 2008.

1.4.3. Validation, saisie et codification

Le recours à l'ITAO (interview téléphonique assistée par ordinateur) pour mener les entrevues de l'EQSP permet une saisie automatique par l'intervieweur des réponses des individus échantillonnés. La très grande majorité des questions sont en outre autocodées, c'est-à-dire que l'intervieweur n'a qu'à cocher le choix du répondant : la saisie s'en trouve donc réduite au minimum. Toutefois, certains renseignements nécessitent davantage de travail de la part de l'intervieweur. En effet, les réponses aux questions ouvertes – par exemple celles qui portent sur la description du travail effectué (Q47) ou la description du genre d'entreprise au sein de laquelle la personne est employée (Q45) – doivent être entièrement saisies par l'intervieweur. En plus de ces deux exemples, plusieurs questions portant sur différents sujets – problèmes de nez (Q6D), méthode pour perdre ou maintenir son poids (Q18M), genre de blessure subie (Q31) et enfin quelques caractéristiques sociodémographiques (Q110, Q112, Q113, Q114 et Q115) – comportaient une catégorie « Autre » permettant au répondant de décrire une situation particulière. Au

terme de la collecte, une attention spéciale a été accordée à la codification des renseignements recueillis par les intervieweurs.

Par ailleurs, la collecte téléphonique étant entièrement informatisée, plusieurs validations de base ont été réalisées pendant l'entrevue, comme la vérification du respect des intervalles de réponse pour les questions fermées ou celle de l'adéquation des sauts de section pour les blocs de questions ne s'adressant qu'à un sous-groupe de répondants (par exemple la section portant sur la santé au travail).

De plus, une validation effectuée *a posteriori* afin de vérifier principalement la cohérence des réponses d'un même répondant a servi à relever quelques erreurs. À cette étape, la codification des réponses aux questions permettant de distinguer les non-répondants partiels des répondants qui n'étaient pas visés par une question a également été vérifiée. Enfin, un examen approfondi de la séquence des codes de résultats de la collecte a été fait, ce qui a permis de corriger certains codes finaux de résultats.

Les questions sur le genre d'emploi occupé (Q47) et sur le genre d'entreprise pour laquelle le répondant travaille (Q45) ont servi à déterminer le code à quatre chiffres de la profession occupée selon la Classification nationale des professions (CNP) de 2006 et le code à quatre chiffres du secteur d'activité économique de l'entreprise selon le Système de classification des industries de l'Amérique du Nord (SCIAN) de 2002. Ce travail a été effectué par la Direction des statistiques du travail et de la rémunération (DSTR) de l'ISQ. Quant aux autres questions ouvertes, de concert avec les intervenants associés à l'enquête, la plupart des libellés ont été recodés et reclassés dans les catégories existantes ou ont été regroupés afin de former de nouvelles catégories.

1.5. Taux de réponse et taux de déclaration par procuration

1.5.1. Taux de réponse

Pour être sélectionné pour participer à l'enquête, un individu doit avoir été choisi parmi les membres de son ménage, mais il faut aussi, au préalable, que son ménage ait été lui-même choisi parmi les ménages admissibles, comme décrit à la section 1.2. En d'autres termes, la probabilité de sélection globale d'un individu doit tenir compte des deux niveaux de sélection : la sélection du ménage et celle de l'individu. De même, le calcul du taux de réponse global de l'enquête doit intégrer chacun de ces niveaux, tant au chapitre des probabilités de sélection qu'à celui des ajustements pour tenir compte de la non-réponse.

Les taux de réponse produits et diffusés par l'ISQ pour les enquêtes auprès des individus et des ménages sont, à moins d'avis contraire, des taux pondérés²³. Ce qui veut dire, entre autres, qu'ils tiennent compte de la non-proportionnalité régionale de l'échantillon de l'enquête. Pour avoir plus de détails sur les taux de réponse pondérés *individu*, *ménage* et *global* de l'EQSP, et pour connaître le détail des formules qui servent à calculer ces taux, consultez l'annexe 2.

Des 62 069 ménages privés admissibles de l'échantillon, 46 621 ménages ont répondu aux questions de sélection, soit un taux de réponse ménage pondéré de 73,8 %. De ces 46 621 ménages répondants, autant d'individus ont été sélectionnés du fait qu'une seule personne par ménage était retenue pour participer à l'enquête. De ceux-ci, 79,1 % ont répondu à l'enquête (taux de réponse individu pondéré). Ce qui veut dire qu'au total, 38 154 personnes ont répondu à l'enquête, pour un taux de réponse global pondéré de 58,4 %. Le tableau 1-7 présente les taux de réponse globaux pondérés de l'EQSP et le nombre de répondants pour chacune des régions sociosanitaires et pour l'ensemble du Québec. Le tableau 1-8 présente ceux de chaque RLS des régions ayant fait l'achat d'un échantillon supplémentaire.

²³ Les taux de réponse diffusés par Statistique Canada, par exemple ceux de l'ESCC, sont des taux de réponse non pondérés contrairement à ceux de l'ISQ. La comparaison du taux de réponse de l'EQSP avec celui de cette enquête ne peut donc être faite directement à partir des résultats fournis dans la documentation de Statistique Canada.

Bien que le taux de réponse obtenu pour l'ensemble du Québec soit quelque peu inférieur à la cible initiale, le taux visé par région (un minimum de 60 % tout en espérant approcher les 65 %) a toutefois été atteint dans la majorité des cas. En fait, les taux de réponse régionaux oscillent entre 60 % et 67 %, à l'exception des régions de Montréal et de Laval qui ont obtenu des taux de 47 % et 54 % respectivement. Notons qu'il est habituel d'obtenir des taux de réponse inférieurs dans ces deux régions.

Au niveau local, les taux de réponse sont plutôt stables pour les RLS d'une même région. Et, outre le RLS 801 de l'Abitibi-Témiscamingue et trois RLS de la Gaspésie-Îles-de-la-Madeleine (1101, 1102 et 1103), qui obtiennent respectivement des taux de réponse de 55 %, 59 %, 59 % et 58 %, ceux des autres RLS oscillent tous entre 61 % et 68 %.

**Tableau 1-7 Taux de réponse global pondéré et nombre de répondants selon la région
socio-sanitaire, Québec**

Région socio-sanitaire	Taux de réponse pondéré	Nombre de répondants
	(%)	(n)
Bas-Saint-Laurent	64,7	2 028
Saguenay-Lac-Saint-Jean	66,6	4 008
Québec	64,8	2 025
Mauricie-Centre-du-Québec	64,2	3 788
Estrie	64,8	2 024
Montréal	46,6	1 986
Outaouais	62,8	1 972
Abitibi-Témiscamingue	64,0	2 787
Côte-Nord	62,8	2 012
Nord-du-Québec	63,5	1 971
Gaspésie-Îles-de-la-Madeleine	59,9	2 538
Chaudière-Appalaches	63,1	2 958
Laval	53,8	1 996
Lanaudière	62,8	2 070
Laurentides	59,9	1 945
Montérégie	61,7	2 046
Ensemble du Québec	58,4	38 154

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau 1-8 Taux de réponse globale pondéré et nombre de répondants selon le RLS, régions qui ont fait l'achat d'un échantillon supplémentaire

Région sociosanitaire / RLS	Taux de réponse pondéré	Nombre de répondants
	(%)	(n)
Saguenay–Lac-Saint-Jean		
201 RLS de La Baie	66,6	721
202 RLS de Chicoutimi	66,8	654
203 RLS de Jonquière	66,9	669
204 RLS du Domaine-du-Roy	67,6	656
205 ¹ RLS de Maria-Chapdelaine	62,7	649
206 ² RLS de Lac-Saint-Jean-Est	67,3	659
Mauricie–Centre-du-Québec		
401 RLS du Haut-Saint-Maurice	63,1	452
402 RLS de la Vallée de la Batiscan	64,4	458
403 RLS du Centre-de-la-Mauricie	66,6	473
404 RLS de Maskinongé	65,1	470
405 RLS de Trois-Rivières	62,5	555
406 RLS de Bécancour - Nicolet-Yamaska	64,4	483
407 RLS de Drummond	62,8	423
408 RLS d'Arthabaska - de l'Érable	65,9	474
Abitibi–Témiscamingue		
801 RLS du Témiscaming	55,3	362
802 RLS de Ville-Marie	63,9	490
803 RLS de Rouyn-Noranda	62,2	443
804 RLS de l'Abitibi-Ouest	68,4	510
805 RLS de l'Abitibi	67,1	479
806 RLS de la Vallée-de-l'Or	62,8	503
Gaspésie–Îles-de-la-Madeleine		
1101 RLS de la Baie-des-Chaleurs	58,5	669
1102 RLS du Rocher-Percé	58,8	483
1103 RLS de La Côte-de-Gaspé	58,0	456
1104 RLS des Îles-de-la-Madeleine	66,3	480
1105 RLS de la Haute-Gaspésie	62,0	450
Chaudière-Appalaches		
1201 RLS de Lac-Etchemin	67,8	511
1202A RLS du Littoral - Urbain	62,6	517
1202B RLS du Littoral - Rural	65,9	500
1203 RLS de Beauce	65,1	447
1204 RLS de l'Amiante	61,2	495
1205 RLS de Montmagny-L'Islet	62,7	488
Lanaudière		
1401 RLS de Lanaudière-Nord	63,6	1 094
1402 RLS de Lanaudière-Sud	62,1	976
Total		18 149

1. Incluant Saint-Ludger-de-Milot (c.f. à la section 6.3.5)

2. Excluant Saint-Ludger-de-Milot (c.f. à la section 6.3.5)

Source : Institut de la statistique du Québec, EQSP 2008.

Comme mentionné à la section 1.2, le répondant était invité à répondre à un certain nombre de questions portant sur les difficultés socioémotionnelles de son enfant ou l'un de ses enfants, de 3 à 14 ans. Les questions utilisées proviennent du *Strengths and Difficulties Questionnaire (SDQ)*²⁴, de son supplément sur l'impact de ces difficultés ainsi que du questionnaire du *U.S. National Health Interview Survey*. Or, étant donné qu'aucun « changement » d'interlocuteur n'était nécessaire, seulement trois répondants ont refusé de répondre à cette section. La proportion pondérée de répondants à la section enfant est donc, une fois arrondie, de 100 %. Pour connaître le nombre de répondants à la section enfant pour l'ensemble du Québec et pour chacune des régions sociosanitaires, consultez le tableau 1-9.

Tableau 1-9 Nombre de répondants à la section enfant selon la région sociosanitaire, Québec

Région sociosanitaire	Nombre de répondants - Section enfant -
	(n)
Bas-Saint-Laurent	240
Saguenay-Lac-Saint-Jean	524
Québec	245
Mauricie-Centre-du-Québec	436
Estrie	263
Montréal	217
Outaouais	312
Abitibi-Témiscamingue	388
Côte-Nord	294
Nord-du-Québec	317
Gaspésie-Îles-de-la-Madeleine	296
Chaudière-Appalaches	421
Laval	319
Lanaudière	319
Laurentides	340
Montérégie	317
Ensemble du Québec	5 248

Source : Institut de la statistique du Québec, EQSP 2008.

1.5.2. Taux de déclaration par procuration

Dans l'EQSP, les entrevues devaient s'effectuer directement auprès de la personne sélectionnée dans le ménage. Or, si la personne était incapable de répondre pour des raisons de santé physique ou mentale, ou en cas d'absence prolongée, les renseignements à son sujet pouvaient alors être fournis par un autre membre bien informé du ménage. Cette façon de faire est qualifiée de déclaration par procuration.

Quoique les tiers répondants aient été en mesure de donner des réponses exactes à la plupart des questions de l'enquête, les questions plus délicates ou personnelles pouvaient dépasser leurs connaissances. Par conséquent, certaines questions ou sections complètes, comme celles sur les tendances suicidaires et les habitudes sexuelles, sont demeurées sans réponse lorsqu'elles ont été soumises à des tiers répondants.

Le tableau 1-10 présente les taux pondérés de déclaration par procuration pour chacune des régions sociosanitaires. Les taux selon le découpage local ne sont pas présentés vu le taux global très faible et l'absence de variabilité à ce niveau. Sur les 38 154 entrevues complétées, 37 500 ont été réalisées avec la

²⁴ Élaboré et diffusé par Robert Goodman qui en détient les droits. Voir le site internet du SDQ : <http://www.sdqinfo.com/b1.html>.

personne sélectionnée et 654 à l'aide d'une tierce personne, ce qui correspond à un taux de déclaration par procuration pondéré de 1,8 % pour l'ensemble du Québec. Ce taux varie très peu selon la région sociosanitaire (entre 1,4 % et 2,9 %). Étant donné que ces taux sont faibles²⁵, le biais engendré par la non-réponse découlant des répondants par procuration peut être considéré négligeable lors de la production d'estimations tirées de l'EQSP.

Tableau 1-10 Taux pondéré de déclaration par procuration selon la région sociosanitaire, Québec

Région sociosanitaire	Taux pondéré de déclaration par procuration
	(%)
Bas-Saint-Laurent	2,7
Saguenay–Lac-Saint-Jean	1,7
Québec	1,6
Mauricie–Centre-du-Québec	1,4
Estrie	2,0
Montréal	1,8
Outaouais	1,9
Abitibi–Témiscamingue	2,4
Côte-Nord	2,3
Nord-du-Québec	1,6
Gaspésie–Îles-de-la-Madeleine	2,9
Chaudière-Appalaches	2,0
Laval	1,4
Lanaudière	1,6
Laurentides	2,3
Montérégie	1,5
Ensemble du Québec	1,8

Source : Institut de la statistique du Québec, EQSP 2008.

2. Estimation des proportions et des effectifs

2.1. Définition

Deux types d'estimations peuvent être calculés à l'Infocentre à partir des données de l'EQSP : des effectifs et des proportions. L'estimation d'effectifs est utilisée lorsqu'on désire savoir combien de personnes dans une population possèdent une caractéristique donnée. Celle-ci est obtenue en faisant la somme du poids populationnel de ces individus²⁶ (voir section 2.6). Si l'on désire plutôt savoir quel est le pourcentage de la population que ces personnes représentent, alors on utilise l'estimation de proportions. Pour ce faire, il suffit de diviser la valeur obtenue pour l'effectif par l'ensemble des personnes de la population, qu'elles possèdent ou non cette caractéristique (voir section 2.5).

Parce qu'elles sont calculées à partir de données d'enquête, chacune de ces estimations utilise la pondération afin de produire des résultats adéquats, c'est-à-dire pouvant être inférés à la population. La pondération permet de tenir compte, entre autres, de la probabilité de sélection et de la non-réponse totale, mais n'inclut pas d'ajustement pour la non-réponse partielle. Or, la présence de non-réponse à la question

²⁵ D'ailleurs, les taux de déclaration par procuration obtenus dans l'EQSP sont très semblables à ceux d'autres enquêtes. Les taux de procuration de l'ESCC sont : 2,2 % pour le cycle 2003, 1,6 % pour le cycle 2005 et 2,3 % pour le cycle 2007-2008.

²⁶ En supposant qu'il y a absence de non-réponse partielle.

analysée, appelée non-réponse partielle, a pour effet de fausser les estimations. L'ampleur de cet effet et la nécessité de procéder à des ajustements dépendent de l'importance de la non-réponse partielle. Les prochaines sections fournissent des précisions à propos de la pondération (section 2.2) et de la non-réponse partielle (section 2.3).

2.2. Pondération

Afin que les estimations produites à partir des données de l'EQSP soient représentatives de la population, et non seulement de l'échantillon, une pondération a été calculée. L'objectif de la pondération est d'attribuer une valeur, un poids, à un répondant à l'enquête qui correspond au nombre de personnes qu'il « représente » dans la population. Ce poids est appelé « poids populationnel ».

La pondération est essentielle pour l'obtention des diverses estimations tirées de l'EQSP. C'est ce qui permet de rapporter les données des répondants à la population visée et, ainsi, de faire des inférences adéquates à cette population, bien que celle-ci n'ait pas été observée dans sa totalité. C'est également un des éléments à considérer pour estimer correctement la précision des données. À titre d'exemple, la répartition de l'échantillon selon les régions sociosanitaires ne représente pas bien celle de la population et l'achat d'un échantillon supplémentaire par certaines régions a accentué ce déséquilibre (voir la section 1.3). La pondération permet de corriger cette distorsion créée par le plan de sondage (sélection de l'échantillon). De plus, il est connu que, dans ce type d'enquête, la probabilité de répondre de certains groupes d'âge et sexe n'est pas la même. Il est donc nécessaire de tenir compte de ces éléments, qui doivent dès lors être intégrés à la pondération lors de la production de statistiques tirées de l'EQSP. Pour savoir comment utiliser la pondération, consultez l'annexe 3.

La stratégie de pondération développée par l'ISQ pour l'EQSP est relativement complexe et tient compte, entre autres, de la probabilité de sélectionner une unité dans la base de sondage, des unités hors champ, de la non-réponse totale²⁷ (ménage et individu), de même que d'un possible effet saisonnier et d'un ajustement aux comptes de la population du recensement. Deux pondérations distinctes ont été construites pour l'EQSP : une pour le répondant de 15 ans et plus et une pour l'enfant de 3 à 14 ans du répondant. Le poids enfant doit être utilisé lorsque des résultats en provenance de la section enfant sont produits (Q103 à Q108, auxquels s'ajoutent les indicateurs 18.1 à 18.10 construits à partir de ces questions), afin de permettre une inférence adéquate pour la population des enfants de 3 à 14 ans vivant dans les ménages privés au Québec. Le poids du répondant de 15 ans et plus doit être utilisé dans toutes les autres situations, c'est-à-dire celles portant sur les personnes de 15 ans et plus.

Le tableau 2-1 présente un condensé des étapes ayant mené à la création de ces deux poids; celui du répondant de 15 ans et plus et celui de l'enfant de 3 à 14 ans. La prochaine section offre une description de ces étapes.

2.2.1. Description générale des étapes

Cette section décrit, de façon générale, les différentes étapes de la stratégie de pondération de l'EQSP présentées au tableau 2-1. L'objectif est de présenter les grandes étapes de la pondération et non de fournir des détails précis sur chacune d'entre elles.

²⁷ Le terme « non-réponse totale » fait référence à l'absence de réponse pour l'ensemble des questions d'un questionnaire. Le taux de non-réponse totale s'obtient en soustrayant le taux de réponse de la valeur unitaire. Pour l'EQSP, on peut calculer trois taux différents de non-réponse totale : celui au niveau de l'individu, celui au niveau du ménage et celui au niveau global. Ce dernier correspond au complément du taux de réponse global de l'enquête décrit à la section 1.5.1.

Comme décrit dans les premières sections de ce document, l'EQSP a eu recours à une base de sondage GANT pour la sélection de son échantillon. Outre la sélection du répondant de 15 ans et plus, le plan d'échantillonnage de l'EQSP prévoyait aussi la sélection d'un enfant de 3 à 14 ans lorsque le répondant est le parent d'un enfant de cet âge. La stratégie de pondération a été développée séparément pour ces deux unités de sélections. Cependant, les premières étapes de la création des poids sont communes aux deux. Ce n'est qu'à partir de la cinquième étape que les ajustements diffèrent.

La première étape de la pondération concerne la probabilité de sélectionner une unité de la base de sondage dans l'échantillon. Pour l'EQSP, cette probabilité possède deux niveaux : la sélection d'un ménage et la sélection d'un individu du ménage. La première sera abordée dans la partie commune, alors que la seconde sera discutée dans les parties qui distinguent les deux pondérations.

Partie commune

Pour la partie commune des deux pondérations, il y a quatre étapes : la première étape (C1) est, comme nous venons de le mentionner, la composante ménage de la probabilité de sélection. Le poids créé à partir de cette probabilité est lié aux chances qu'a un ménage d'être choisi dans l'échantillon. En d'autres termes, le poids correspond à l'inverse de la probabilité de sélectionner ce ménage et représente des ménages de la population visée qui n'ont pas été sélectionnés pour participer à l'enquête.

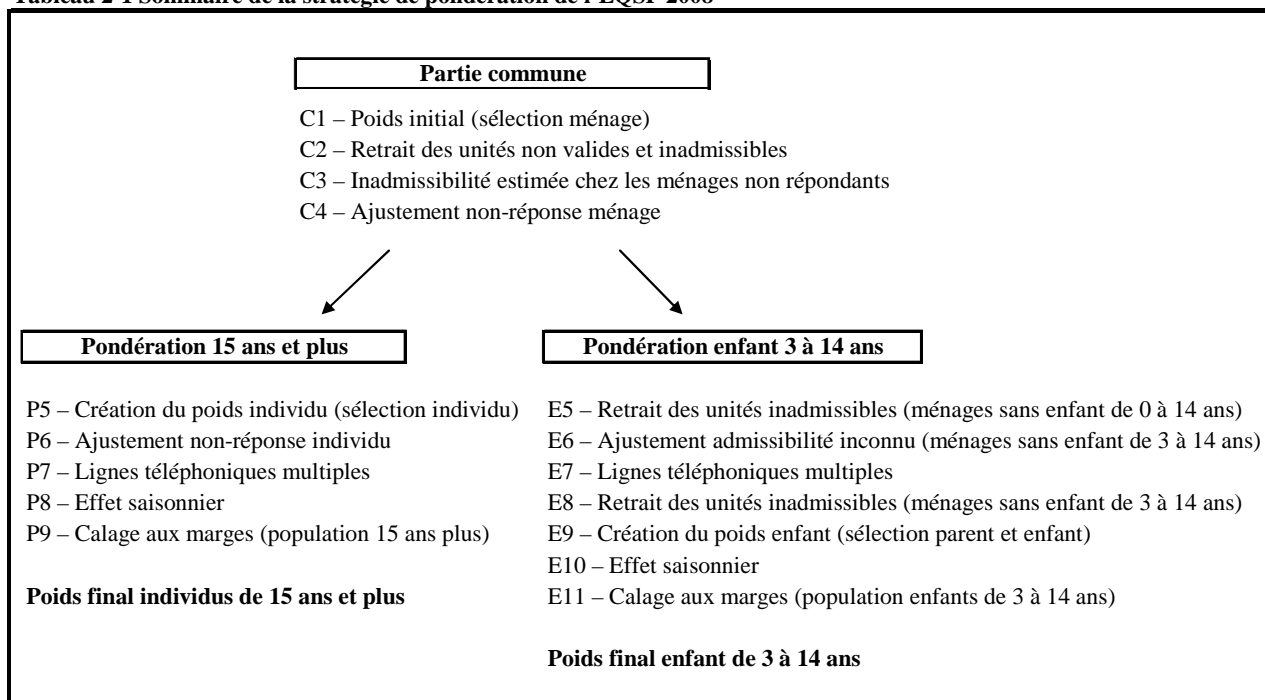
L'étape C2 correspond à l'étape d'ajustement due aux résultats de collecte. En effet, lors de la collecte des données, certaines unités sont associées à des numéros hors service, à des entreprises, à des résidences secondaires, à des ménages collectifs ou des ménages vivant sur des réserves amérindiennes, etc. Ces unités sont soit non valides, soit inadmissibles à l'enquête et sont simplement retirées du fichier. En termes statistiques, ce geste a pour effet de poser l'hypothèse que les unités représentées par celles qui ont été retirées sont également des unités hors champ.

Malgré les nombreux contacts ou tentatives de contact fait pour chacune des unités, à la fin de la collecte certains ménages échantillonnés n'ont pas fourni de renseignements permettant d'établir leur statut d'admissibilité. L'étape C3 a pour objectif d'estimer le nombre d'unités admissibles parmi les ménages non-répondants. L'hypothèse sous-jacente à cette étape est qu'un certain nombre d'unités dont l'admissibilité n'a pu être déterminée sont en réalité inadmissibles.

La pondération doit également inclure un ajustement afin de tenir compte de la non-réponse. Cette non-réponse peut survenir au niveau du ménage (C4) ou au niveau de l'individu (P6 et E6). Que ce soit parce qu'un ménage refuse de répondre à l'enquête ou que celui-ci n'a pu être joint, ces ménages non-répondants doivent être retirés du fichier de données. Par conséquent, les poids des ménages répondants doivent être ajustés afin de représenter ceux exclus. Ainsi, les poids ajustés, qui sont associés aux ménages répondants à l'enquête, représentent les ménages qui n'ont pas été sélectionnés pour participer à l'enquête, ainsi que les ménages qui l'ont été, mais qui n'ont pas répondu. Les ajustements pour la non-réponse sont appliqués de façon indépendante pour chaque RSS. Le fait que le numéro de téléphone soit répertorié dans un bottin a également été considéré lors de cet ajustement.

Le poids résultant de l'étape C4 constitue le poids de départ des deux pondérations de l'EQSP. Les étapes subséquentes de la pondération traitent séparément la pondération du répondant de 15 ans et plus et la pondération des enfants de 3 à 14 ans. Elles sont donc présentées de façon distincte.

Tableau 2-1 Sommaire de la stratégie de pondération de l'EQSP 2008



Source : Institut de la statistique du Québec, EQSP 2008.

Pondération individu de 15 ans et plus

La pondération des individus de 15 ans et plus nécessite cinq autres étapes.

La première est associée à la probabilité de sélection de l'individu dans le ménage et correspond au passage du poids représentant les ménages au poids représentant les individus dans les ménages (P5). Il s'agit tout simplement d'ajuster le poids afin de tenir compte de la probabilité qu'a un individu d'être sélectionné parmi les membres admissibles de son ménage.

Au même titre que l'étape C4 de la partie commune, certains individus sélectionnés pour répondre à l'enquête n'ont finalement pas répondu. Pour cette raison, il est nécessaire d'exclure ces unités du fichier de données et d'ajuster les poids des répondants afin que ces derniers représentent également ceux exclus (P6). Ces ajustements ont été faits indépendamment par RSS. De plus, des classes d'ajustements ont été formées afin de regrouper les unités qui ont une propension à répondre aux questions similaires. Ces classes ont été définies à partir du sexe et de l'âge de l'individu, de la langue de l'entrevue, du nombre de personnes de 14 ans et moins dans le ménage et du fait que le numéro de téléphone soit répertorié ou non dans les bottins de téléphone.

L'étape P7 vise à prendre en considération le fait que certains ménages possèdent plus d'une ligne téléphonique résidentielle, ce qui leur donne plus de chance d'être sélectionnés à l'enquête. Pour ces ménages, le poids de l'individu sélectionné a été divisé par le nombre de lignes téléphoniques.

L'avant-dernière étape de la pondération des 15 ans et plus est un ajustement pour contrôler l'effet saisonnier (P8). Un effet saisonnier peut être induit aux estimations en raison du léger débalancement de la répartition des répondants obtenue selon les mois de l'année où ceux-ci sont contactés. Les trimestres utilisés pour l'ajustement sont les suivants : janvier à mars, avril à juin, juillet à septembre et octobre à décembre. Cet ajustement a été fait indépendamment par RSS.

La dernière étape nécessaire afin d'obtenir le poids final des 15 ans et plus de l'EQSP consiste à l'ajustement aux données de la population, aussi appelée calage aux marges (P9). Elle fait en sorte que la somme des poids finaux correspond aux estimations de la population de 2008²⁸ définies à l'échelle des RSS, pour chacun des quatorze groupes d'âge-sexe d'intérêt, c'est-à-dire les sept groupes d'âge 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65-74 ans et 75 ans et plus, pour chacun des deux sexes. Il est à noter que pour les six régions qui ont fait l'achat d'un échantillon supplémentaire, le calage aux marges a été appliqué à l'échelle locale (RLS) sous la contrainte de respecter les tailles de population des régions pour chacun de ces groupes. L'annexe 4 offre de l'information supplémentaire sur le calage aux marges effectué dans l'EQSP.

Les groupes d'âge utilisés à l'échelle locale sont les mêmes que ceux utilisés à l'échelle régionale, à l'exception des groupes 65-74 ans et 75 ans et plus qui ont été regroupés pour chacun des RLS²⁹ et des groupes 15-24 ans et 25-34 ans chez les femmes de la région 12 qui ont été regroupés³⁰, tous les deux en raison d'un nombre insuffisant d'unités.

Mentionnons que le niveau géographique utilisé pour le calage aux marges correspond à celui des objectifs de l'enquête; l'échelle locale pour les régions qui ont fait l'achat d'un échantillon supplémentaire, et l'échelle régionale pour les autres RSS. Un ajustement plus fin pour les régions sans suréchantillon aurait eu pour impact de diminuer la qualité de l'ajustement de la pondération pour celles-ci, en plus de diminuer la qualité de la donnée pour l'ensemble du Québec. Il a donc été convenu de se limiter aux objectifs de l'enquête. L'utilisateur qui souhaiterait produire des résultats géographiques plus fins devrait prendre conscience de cette limite de la pondération de l'EQSP.

Après ces derniers ajustements, le poids de calage devient le poids final du répondant de 15 ans et plus de l'EQSP.

Pondération enfant de 3 à 14 ans

La pondération de l'enfant de 3 à 14 ans du répondant s'effectue à partir du poids résultant de l'étape C4 de la partie commune aux deux pondérations de l'EQSP. Rappelons que ce poids est un poids ménage, c'est-à-dire que sa valeur correspond au nombre de ménages que l'unité représente.

Pour compléter cette pondération, sept autres étapes sont nécessaires. La première correspond au retrait du fichier des ménages sans enfants de 0 à 14 ans (E5), car ces ménages ne sont pas admissibles à répondre à la section enfant. Une question du questionnaire permettait d'obtenir cette information.

L'étape suivante (E6) a pour objectif d'ajuster les poids des ménages répondants dont l'individu sélectionné n'a pas répondu, car l'admissibilité du ménage ne peut être confirmée que par une question posée à l'individu répondant juste avant la section enfant. En l'absence de réponse à cette question, le statut d'admissibilité du ménage ne peut être confirmé. Les ajustements pour cette étape ont été faits de façon indépendante pour chaque RSS. Des classes d'ajustements ont été formées afin de regrouper les unités qui ont une propension à répondre aux questions similaires. Ces classes ont été définies à partir du fait d'avoir un numéro de téléphone qui soit répertorié ou non dans les bottins de téléphone, de la langue de l'entrevue, de l'âge et du sexe du répondant et du nombre total d'individus dans le ménage.

²⁸ Les estimations de la population en 2008 vivant en ménage privé sont basées sur les comptes du Recensement de 2006 et d'une estimation du sous-dénombrement net, de même que sur les comptes de naissance, décès, immigration et émigration depuis ce temps. Ces estimations ont été produites par la Direction des statistiques sociodémographiques de l'ISQ.

²⁹ Ces ajustements devaient toutefois conserver l'ajustement au compte de la population des 65-74 ans et des 75 ans et plus de chaque région.

³⁰ Ces ajustements devaient toutefois conserver l'ajustement au compte de la population des femmes de 15-24 ans et des femmes de 25-34 ans de la région 12.

Tout comme l'étape P7, l'étape E7 de la pondération enfant sert à tenir compte du fait que certains ménages possèdent plus d'une ligne téléphonique résidentielle, ce qui leur donne plus de chance d'être sélectionnés à l'enquête. Pour ces ménages, le poids de l'individu sélectionné a été ajusté à la baisse.

Étant donné que la section enfant portait sur les enfants âgés de 3 à 14 ans, les ménages ne contenant que des enfants de 0 à 2 ans ont été exclus du fichier en raison de leur inadmissibilité (E8). Or, la question permettant de déterminer la présence d'enfants de 3 à 14 ans, chez les ménages ayant déjà déclaré contenir des enfants de 0 à 14 ans, était posée juste avant la section enfant. Le retrait des unités inadmissibles s'effectue à cette étape de la pondération pour cette raison.

Le passage du poids représentant les ménages avec enfants de 3 à 14 ans au poids représentant les enfants eux-mêmes de ces ménages correspond à l'étape E9. Il s'agit tout simplement d'ajuster le poids afin de tenir compte de la probabilité de sélectionner un parent parmi les personnes de 15 ans et plus du ménage, ainsi que de la probabilité qu'a un enfant d'être sélectionné parmi les enfants admissibles de son ménage. Cet ajustement dépend donc du nombre d'enfants de 3 à 14 ans dans le ménage, mais aussi du nombre de parents et du nombre total de personnes de 15 ans et plus dans le ménage.

Par ailleurs, le plan d'échantillonnage de l'EQSP prévoyait que pour les ménages avec un seul enfant admissible, un ménage sur deux seulement était retenu pour répondre à la section enfant. Par conséquent, l'étape E9 de la pondération enfant fait en sorte que les ménages sélectionnés au hasard représentent ceux non sélectionnés. Ces ajustements ont été faits de façon indépendante par RSS.

Les deux dernières étapes de la pondération enfant sont identiques à celles de la pondération des individus de 15 ans et plus. À savoir, l'étape E10 qui concerne l'ajustement fait pour contrôler l'effet saisonnier³¹ et l'étape E11 qui concerne l'ajustement aux données de la population (calage aux marges). Cette dernière, de la même façon que l'étape P9, fait en sorte que la somme des poids finaux corresponde aux estimations de la population des enfants de 3 à 14 ans vivants en ménage privé en 2008³². Ces estimations de la population sont définies à l'échelle des régions sociosanitaires, pour chacun des huit groupes d'âge-sexe d'intérêt, c'est-à-dire les quatre groupes d'âge 3-5 ans, 6-8 ans, 9-11 ans et 12-14 ans, pour chacun des deux sexes.

Pour le volet enfant, les objectifs de l'enquête, en matière de précision des estimations, se limitent à des objectifs régionaux. Pour cette raison, aucun ajustement de la stratégie de pondération des enfants de 3 à 14 ans n'a été effectué à l'échelle locale.

À la suite de ces derniers ajustements, le poids de calage devient le poids final de l'enfant de 3 à 14 ans.

³¹ Les trimestres utilisés sont les mêmes que pour l'étape P8, soit janvier à mars, avril à juin, juillet à septembre et octobre à décembre. L'ajustement est fait de façon séparée pour chacune des RSS.

³² Les estimations de la population en 2008 vivant en ménage privé sont basées sur les comptes du Recensement de 2006 et d'une estimation du sous-dénombrement net, de même que sur les comptes de naissance, décès, immigration et émigration depuis ce temps. Ces estimations ont été produites par la Direction des statistiques sociodémographiques de l'ISQ.

2.3. Non-réponse partielle

Outre la non-réponse totale³³, ménage et individu, qui est habituellement compensée par la pondération (section 2.2), il existe également un autre type de non-réponse, soit celle associée à une question précise, appelée non-réponse partielle. Cette non-réponse se définit comme le rapport entre le nombre pondéré d'individus n'ayant pas répondu à la question et le nombre pondéré d'individus devant y répondre.

$$\text{Taux pondéré de non-réponse partielle} = \frac{\text{nombre pondéré d'individus n'ayant pas répondu à la question}}{\text{nombre pondéré d'individus devant répondre à la question}}$$

La non-réponse partielle peut entraîner des biais dans les estimations si les non-répondants présentent des caractéristiques différentes de celles des répondants. Plus la proportion de non-réponse partielle est élevée, plus le risque de biais occasionnés par un déséquilibre des caractéristiques des non-répondants par rapport à celles des répondants est grand.

En général, lorsqu'on analyse des données d'enquête, telle l'EQSP, il est raisonnable de faire l'hypothèse que pour une non-réponse partielle inférieure à 5 % au niveau de l'ensemble du Québec et à 10 % au niveau régional ou local, les risques de biais sont plutôt faibles.

Ce qui veut dire que l'opération qui sera présentée à la section 2.5, consistant à redistribuer la non-réponse partielle dans les autres catégories de réponse, sans toutefois tenir compte des caractéristiques des non-répondants, est une démarche acceptable. Elle est acceptable du fait que le biais des estimations est en général peu important comparativement à l'erreur d'échantillonnage (c'est-à-dire la précision de l'estimation) et peut être considéré comme négligeable. Par exemple, un biais de 1 % pour une estimation dont l'erreur-type est de 10 % peut être considéré comme négligeable. Par contre, par rapport à une erreur-type de 2 %, ce biais de 1 % ne peut pas être considéré comme négligeable.

Les questions pour lesquelles le taux de non-réponse partielle dépasse le seuil critique doivent en général être analysées plus à fond au regard des biais, les non-répondants pouvant présenter des caractéristiques différentes des répondants, ce qui à son tour peut entraîner un impact non négligeable sur les estimations présentées (sous ou surestimation). Pour savoir comment procéder au diagnostic et à l'analyse de la non-réponse partielle, consultez l'annexe 5. Notons que pour les données de l'EQSP, les risques de biais importants associés à la non-réponse partielle relativement à la précision des estimations sont assez minimes. Il peut donc s'avérer plus pertinent de faire une analyse qualitative, plutôt que quantitative, de la non-réponse partielle. Plus de détails sont fournis à la section 2.5.

Le fait que les répondants aient rempli le questionnaire avec l'aide d'un intervieweur a permis de minimiser la non-réponse partielle, comparativement à un questionnaire autoadministré où le répondant ne bénéficie pas du support de l'intervieweur. Par contre, l'utilisation de la déclaration par procuration a eu l'effet contraire, c'est-à-dire qu'il a contribué à augmenter la non-réponse partielle, du fait que certaines sections du questionnaire n'ont pas pu être complétées en raison de leur caractère intime ou personnel. Toutefois, comme mentionné à la section 1.5.2, le taux de déclaration par procuration varie très peu d'une région à l'autre (voir tableau 1-10). Il est d'ailleurs comparable à celui d'autres enquêtes³⁴. Pour ces raisons, l'ISQ a jugé qu'il n'était pas nécessaire d'avoir recours à l'imputation de valeur pour compenser la non-réponse partielle due à la déclaration par procuration. Pour ces mêmes raisons, et en

³³ Le terme « non-réponse totale » fait référence à l'absence de réponse pour l'ensemble des questions d'un questionnaire. Le taux de non-réponse totale s'obtient en soustrayant le taux de réponse de la valeur unitaire. Pour l'EQSP, on peut calculer trois taux différents de non-réponse totale : celui au niveau de l'individu, celui au niveau du ménage et celui au niveau global. Ce dernier correspond au complément du taux de réponse global de l'enquête décrit à la section 1.5.1.

³⁴ Taux de procuration de l'ESCC : 2,2 % pour le cycle 2003, 1,6 % pour le cycle 2005 et 2,3 % pour le cycle 2007-2008.

raison de la faiblesse globale de la non-réponse partielle, aucune analyse particulière n'a été effectuée par l'ISQ pour caractériser les non-répondants partiels résultant ou non de la déclaration par procuration.

D'autres variables ont toutefois dû être imputées en raison d'une trop grande proportion de non-réponse partielle; il s'agit de l'âge des membres des ménages échantillonnés, de la question 54 (A, B, C et D) concernant les douleurs reliées au travail et de la question 120 qui porte sur le revenu total du ménage. La prochaine section traite de ce sujet.

2.4. Imputation

Comme mentionné au précédent paragraphe, certains renseignements recueillis dans le questionnaire de l'EQSP ont dû être imputés en raison d'une trop grande proportion de non-réponse partielle; l'âge des membres des ménages échantillonnés, la Q54 (A, B, C et D)³⁵ et la Q120³⁶. La présente section décrit les raisons du recours à l'imputation, les caractéristiques des répondants nécessitant une imputation et les méthodes utilisées pour l'imputation. Certains éléments plus techniques se retrouvent à l'annexe 6.

2.4.1. Âge des membres des ménages échantillonnés

Dans le but de sélectionner au hasard une personne du ménage, l'intervieweur devait demander l'âge de chacun des membres du ménage. La personne contactée avait la possibilité de fournir l'âge des individus sous forme de catégorie d'âge (15-24 ans, 25-39 ans, 40-49 ans, 50-64 ans et 65 ans et plus). Toutefois, pour les besoins statistiques, une valeur précise de l'âge était nécessaire pour chacun des répondants. L'ISQ a donc procédé à l'imputation d'un âge précis à l'intérieur de la catégorie mentionnée. L'âge d'environ 3 % des répondants a été imputé. En plus du groupe d'âge indiqué par le répondant, l'information sur le nombre d'enfants de 14 ans et moins et d'individus de 15 ans et plus dans le ménage a, entre autres, été considérée lors de l'imputation.

2.4.2. Imputation de la question 54 – Troubles musculo-squelettiques

Lors de l'administration du questionnaire informatif de l'EQSP, une erreur s'est glissée dans les conditions menant au passage à la question 54. Cette question visait à savoir si les douleurs au cou, au dos et aux membres inférieurs ou supérieurs ressenties par les répondants étaient liées à leur travail principal actuel. Cette section présente l'explication du problème, son impact potentiel et les solutions mises de l'avant pour minimiser son impact réel sur les estimations et leurs variances.

Comme dans tout questionnaire informatif, certains sauts de questions ou de sections sont programmés de sorte que le répondant se voit poser uniquement les questions qui doivent lui être posées. Dans le cas de la question 54, certains répondants ne se sont tout simplement pas fait poser la question, malgré le fait qu'ils auraient dû y répondre. Ceci fait en sorte que plusieurs données sont restées sans réponse, occasionnant une non-réponse partielle élevée à cette question. En tout, 5 060 réponses sont manquantes en raison de ce mauvais saut de question. Le pourcentage de non-réponse partielle varie entre 22,0 % et 24,4 % (tableau 2-2). Les individus qui ont subi ce saut de question injustifié ont le profil suivant : ils ont déclaré une douleur à au moins deux des quatre parties du corps couvertes par la question 54 et ont déclaré qu'au moins l'une d'entre elles n'était pas reliée au travail. En résumé, pour les individus correspondants à ce profil, nous avons l'information selon laquelle leur douleur est reliée à leur travail actuel ou antérieur (Q53), mais nous ne savons pas s'il s'agit, en particulier, de leur emploi principal actuel (Q54). Les répondants de tous les autres profils ont été administrés adéquatement dans le questionnaire.

³⁵ La question 54 porte sur les troubles musculo-squelettiques reliés à l'emploi principal actuel. En particulier, les douleurs reliées au cou, au dos, aux membres inférieurs et aux membres supérieurs.

³⁶ La question 120 porte sur le revenu total de tous les membres du ménage provenant de toutes les sources, avant impôt et autres retenues.

Tableau 2-2 Nombre de réponses manquantes dues au mauvais saut de question et taux pondéré de non-réponse partielle à la question 54 (A, B, C et D) avant et après imputation, Québec

Question	Nombre de réponses manquantes	Pourcentage de NRP	
		avant imputation	après imputation
Q54A - cou	1 171	22,6 %	3,6 %
Q54B - dos	1 497	23,3 %	2,8 %
Q54C - membres inférieurs	1 506	24,4 %	2,7 %
Q54D - membres supérieurs	886	22,0 %	4,1 %

Source : Institut de la statistique du Québec, EQSP 2008.

Afin de minimiser le risque de biais engendré par la non-réponse partielle générée par cette erreur de programmation, l'ISQ a décidé de procéder à l'imputation de valeur pour ces quatre questions. La quantité et la qualité de l'information disponible pour effectuer une imputation ont motivé cette décision. Il faut préciser que seulement la non-réponse résultant de l'erreur de passage a été imputée, de sorte qu'après l'opération d'imputation, le taux de non-réponse partielle de la question 54 n'est pas nul.

L'imputation de données a pour effet de gonfler artificiellement le nombre de répondants ayant fourni des réponses sans toutefois gonfler la variance de l'estimation. Il faut donc être très prudent lors de l'analyse de données imputées. À ce sujet, le lecteur est invité à consulter la section 3.1.2 afin de prendre connaissance de l'impact de l'imputation sur la variance de la Q54 et des indicateurs associés à cette question. Entre autres, des recommandations quant à l'utilisation de la donnée imputée y sont formulées.

Les méthodes utilisées pour l'imputation des questions Q54A, Q54B, Q54C et Q54D, ainsi que les résultats de certaines analyses effectuées pour mesurer l'impact de cette imputation sur la variance des estimations sont présentés à l'annexe 6.

2.4.3. Imputation de la question 120 – Revenu du ménage

Il est connu que le revenu personnel des individus est une information plutôt difficile à obtenir des répondants d'une enquête auprès de la population. L'EQSP ne fait pas exception.

La question 120 du questionnaire de l'enquête demandait d'estimer, pour les douze derniers mois, le revenu total de tous les membres du ménage provenant de toutes les sources, avant impôt et autres retenues. Le répondant avait la possibilité de fournir cette estimation soit par une valeur exacte, bien qu'approximative, soit sous forme de catégorie en indiquant la tranche de revenu³⁷ dans laquelle elle se trouve. Sur les 38 154 répondants de l'EQSP, 4 411 ont refusé ou n'étaient pas en mesure de fournir le revenu du ménage pour la période de référence de douze mois, soit une proportion pondérée de 14,4 %. Par ailleurs, 12 190 ont déclaré le revenu du ménage à l'aide d'une catégorie, ce qui correspond à une proportion de 32,3 %.

En présence d'une non-réponse partielle élevée, il est souhaitable de procéder à l'imputation de données de manière à minimiser les risques de biais occasionné par une non-réponse caractérisée. L'ISQ a donc décidé de procéder à l'imputation du revenu total du ménage pour les non-répondants partiels de l'EQSP. Il a également été décidé d'imputer une valeur exacte de revenu aux répondants qui ont fourni un revenu sous forme de catégorie, car cette valeur était nécessaire afin de créer un indice de faible revenu. Dans les

³⁷ Les tranches de revenus proposées sont les suivantes : moins de 10 000 \$, 10 000 \$ à 19 999 \$, 20 000 \$ à 29 999 \$, 30 000 \$ à 39 999 \$, 40 000 \$ à 49 999 \$, 50 000 \$ à 59 999 \$, 60 000 \$ à 79 999 \$, 80 000 \$ à 99 999 \$, 100 000 \$ et plus.

paragraphes suivants, l'absence totale de renseignements et la déclaration sous forme de catégorie seront considérées comme une non-réponse partielle. De cette façon, le taux global pondéré de non-réponse partielle atteint 46,7 % pour l'ensemble du Québec (tableau 2-3).

Il est ressorti des analyses effectuées sur les non-répondants partiels au revenu, que ceux-ci sont principalement des jeunes de moins de 25 ans (74,2 %³⁸) et des femmes de 65 ans et plus (62,2 %). En effet, nous avons constaté que très peu de jeunes sont en mesure de fournir le revenu total du ménage, car la proportion de moins de 18 ans qui a fourni une valeur exacte du revenu est très faible (12,2 % - donnée non présentée).

Tableau 2-3 Taux pondéré de non-réponse partielle pour la question 120 et à l'indice du niveau de revenu du ménage avant imputation, pour deux sous-population plus affectées par la non-réponse

Sous-population	Pourcentage de NRP	
	Q120	Indice du niveau de revenu*
15 - 24 ans	74,2 %	57,3 %
Femmes de 65 ans et plus	62,2 %	45,4 %
Ensemble du Québec	46,7 %	32,1 %

* Il s'agit uniquement de la non-réponse attribuable à l'absence d'information sur le revenu. La non-réponse attribuable à l'absence d'autres informations nécessaires à la création de l'indice (comme le code postal) n'est pas comptabilisée ici.

Source : Institut de la statistique du Québec, EQSP 2008.

Pour les analyses des données de l'EQSP, la principale variable de revenu utilisée n'est pas tirée directement de la question 120; il s'agit plutôt d'un indice construit à partir de cette dernière. Ainsi, l'indice du niveau de revenu du ménage se définit comme le niveau de revenu ajusté pour le seuil de faible revenu selon la taille du ménage et de la communauté, présenté en quintiles de 1 (faible) à 5 (élevé). Or, certaines des catégories de revenu déclarées par les répondants suffisent pour attribuer une valeur à cet indice, ce qui veut dire que 14,6 % des réponses fournies sous forme de catégorie n'ont pas à être considérées comme de la non-réponse partielle. Par conséquent, le taux de non-réponse partielle de l'indice du niveau de revenu attribuable à l'absence du revenu est en réalité de 32,1 % pour l'ensemble du Québec (tableau 2-3); les taux sont respectivement de 57,3 % et de 45,4 % chez les moins de 25 ans et les femmes de 65 ans et plus.

Dans le cas des analyses faites à partir du revenu, l'utilisateur est fortement invité à consulter la section 3.1.2 du guide concernant l'impact de l'imputation sur la variance des estimations. Des recommandations essentielles à une bonne analyse des données sont également présentées à la section 6.3.4.

Plus de détails sur les méthodes utilisées pour l'imputation du revenu du ménage dans l'EQSP sont présentés à l'annexe 6.

2.5. Estimation des proportions

Lorsqu'un chercheur analyse un indicateur, ou simplement une question de l'enquête, il lui appartient de décider s'il considère ou non la non-réponse partielle (identifiée par « NRP » dans les tableaux du guide³⁹) comme une réponse possible ou comme une absence de réponse. S'il la considère comme une réponse en soit, il pourra estimer adéquatement la proportion de personnes qui « ne savent pas » ou qui

³⁸ Pourcentage pondéré des répondants de ce groupe qui n'ont pas déclaré de revenu exact à l'enquête.

³⁹ Pour les besoins du guide, nous avons regroupé sous l'appellation « NRP » les catégories de non-réponse suivantes : « ne sait pas », « ne répond pas », « refus global » et « non-réponse en raison de la déclaration par un tiers répondant ».

« ne veulent pas dire » s'ils possèdent la caractéristique étudiée, mais il ne pourra pas estimer correctement la proportion de la population qui possède cette caractéristique. On parle alors de l'étude de la « déclaration » d'un phénomène et non de l'étude du phénomène directement. Si le chercheur souhaite plutôt obtenir une estimation sans biais d'un phénomène précis, il devrait considérer la non-réponse comme une absence de réponse et ne pas estimer la proportion d'individus de cette catégorie. En faisant cela, la non-réponse partielle sera répartie dans les autres catégories de réponse permettant par conséquent d'estimer correctement la proportion du phénomène étudié.

En général, ceci a peu d'impact sur les estimations des proportions d'intérêt puisque la non-réponse partielle est plutôt faible dans l'EQSP. Malgré tout, pour les estimations de proportions tirées de l'EQSP, il est recommandé de répartir systématiquement les non-répondants partiels de la façon dont les répondants se distribuent au sein des différentes catégories de la variable. L'hypothèse sous-jacente est que les répondants sont représentatifs des non-répondants partiels. En pratique, il suffit de remplacer la valeur identifiant la catégorie « NRP » par une valeur manquante de sorte que la distribution des réponses se fasse sans considérer cette catégorie. Les tableaux 2-4 et 2-5 montrent un exemple d'application. On peut y constater que le dénominateur servant à calculer la proportion diminue, alors que le numérateur (pour chaque catégorie de réponse) demeure inchangé. Ceci a pour effet de gonfler les proportions en redistribuant, en quelque sorte, la représentation des non-répondants partiels de façon proportionnelle à la distribution des répondants. Ne pas répartir les non-répondants partiels dans les autres catégories de réponse reviendrait à émettre l'hypothèse que tous les individus qui n'ont pas fourni de réponse à la question n'ont pas la caractéristique étudiée.

Étant donné que toutes les estimations produites sur le portail, à partir des données de l'EQSP, sont calculées en appliquant la méthode décrite ci-dessus, les pourcentages du tableau 2-5 (mais pas les effectifs – voir la note au bas du tableau) correspondent aux valeurs que l'on retrouve sur le portail de l'Infocentre, comme en témoigne l'image de la sortie produite à l'Infocentre présenté au tableau 2-6.

Tableau 2-4 Effectif et proportions pondérées de la population ayant déjà pris ou essayé du cannabis (marijuana ou haschich) au cours de la vie, Québec

MARI (A déjà pris ou essayé de la marijuana, du cannabis ou du haschich)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP	152 088	2,4	152 088	2,4
Oui	2 711 013	42,9	2 863 101	45,3
Non	3 463 422	54,7	6 326 523	100,0

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau 2-5 Effectif et proportions pondérées de la population ayant déjà pris ou essayé du cannabis (marijuana ou haschich) au cours de la vie, Québec

MARI (A déjà pris ou essayé de la marijuana, du cannabis ou du haschich)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP
Oui	2 711 013	43,9	2 711 013	43,9
Non	3 463 422	56,1	6 174 435	100,0

* Les valeurs présentées dans cette colonne ne sont pas des estimations sans biais de l'effectif de population, car la non-réponse partielle n'a pas été intégrée adéquatement aux calculs. Ces valeurs ne sont présentées que pour illustrer le traitement recommandé pour l'estimation des proportions. Pour une estimation sans biais de l'effectif, consulter la section 2.6 (tableau 2-6).

Source : Institut de la statistique du Québec, EQSP 2008.

Bien qu'aucune étude n'ait été réalisée, on estime, à l'image des résultats préliminaires tirés de l'EQSP qui ont montré que pour la majorité des variables le taux de non-réponse partielle au niveau de l'ensemble du Québec est inférieur à 5 %, que les risques de biais importants associés à la non-réponse partielle relativement à la précision des estimations sont assez minimes. Ainsi, aucune attention particulière n'est nécessaire au moment d'analyser les données de l'édition de 2008 de l'EQSP. Il y a certes quelques variables avec un taux de non-réponse partielle supérieur à 5 %, mais il s'agit, sauf exception, de petites populations pour lesquelles la précision est déjà faible. Si de tels cas se présentent, l'utilisateur pourra alors se questionner sur la signification d'un tel résultat en procédant plutôt à une analyse qualitative, notamment en interrogeant les causes d'un aussi haut taux de non-réponse partielle.

Pour les estimations régionales ou locales tirées de l'EQSP, le taux de non-réponse partielle n'a pas fait l'objet d'un examen détaillé. Cette tâche revient donc à l'utilisateur. Ce dernier devra déterminer lui-même si le biais des estimations de proportions peut être considéré comme négligeable par rapport à l'erreur d'échantillonnage (c'est-à-dire la précision de l'estimation).

2.6. Estimation des effectifs

Dans le cas de l'estimation des effectifs de population, la non-réponse partielle a pour effet direct de sous-estimer les effectifs. Et plus la non-réponse partielle est importante, plus les effectifs de population sont sous-estimés. Ainsi, des corrections devraient être apportées pour l'estimation des effectifs de façon à produire des estimations pour lesquelles on minimise les biais quel que soit le taux de non-réponse partielle.

L'estimation des effectifs provient de l'addition du poids populationnel de chacun des individus qui possède la caractéristique étudiée. Ce poids a préalablement été ajusté pour tenir compte des personnes qui n'ont pas participé à l'enquête – non-réponse totale – (voir section 2.2), mais n'a pas été ajusté pour tenir compte du fait que certains individus n'ont pas répondu à une question donnée – non-réponse partielle – (voir section 2.3). Ainsi, lorsqu'on produit un tableau sans la catégorie « NRP », comme recommandé à la section 2.5, la somme des poids s'en trouve diminuée proportionnellement à la représentativité des poids des « NRP ». Ce qui mène à une sous-estimation des effectifs de population pour ce phénomène.

Les tableaux 2-4 et 2-5 peuvent à nouveau être consultés afin de prendre conscience de l'effet de la non-réponse partielle sur l'estimation des effectifs de population. Cet exemple montre qu'en présence de non-réponse partielle non redistribuée, la valeur de l'effectif, qui correspond à la somme des poids populationnels, sous-estime le nombre d'individus de la population qui possède la caractéristique étudiée. L'explication vient du fait qu'un certain nombre de personnes n'ayant pas fourni de réponse possède peut-être cette caractéristique. Ainsi, il n'est pas approprié d'utiliser uniquement la somme des poids populationnels pour estimer les effectifs de population lorsqu'on est en présence de non-réponse partielle. Il faut aussi tenir compte de cette dernière afin d'estimer adéquatement les effectifs.

Une solution simple pour contrer le problème de sous-estimation est de multiplier l'effectif estimé en présence de non-réponse partielle par l'inverse du taux de réponse de la question. On redistribue ainsi l'effectif de la non-réponse dans les autres catégories de réponse en tenant compte de la distribution des effectifs des répondants. Les résultats offerts sur le portail de l'Infocentre ont été produits en utilisant cette méthode, comme le démontre, pour notre exemple, les valeurs du tableau 2.6 (voir le détail des calculs à l'annexe 7). Prenez note également que les estimations d'effectif produites à l'Infocentre sont toutes arrondies à la centaine.

Tableau 2-6 Proportion de la population ayant consommé du cannabis (marijuana ou haschich) au cours de la vie pour l'ensemble du Québec, EQSP 2008

Statistiques					
Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
2 777 800	43,9	1,0	0,43	(43,1 - 44,8)	2,4

Source : Infocentre de santé publique, octobre 2010.

Non-réponse en cascade

Par ailleurs, lorsqu'on analyse les réponses à une question pour laquelle une sous-population est définie par la réponse à une autre question, il faut alors considérer l'effet « cascade » émanant de cette question filtre⁴⁰. Par exemple, lorsqu'on souhaite estimer sans biais le nombre de personnes qui a sérieusement songé à se suicider ou à s'enlever la vie au cours des 12 derniers mois, on doit tenir compte de la non-réponse à cette question (tableau 2-8), mais également de la non-réponse à la question filtre qui identifie la sous-population qui y a déjà songé au cours de leur vie (tableau 2-7)⁴¹.

Tableau 2-7 Effectif et proportions pondérées de la question 63 portant sur les idées suicidaires et les tentatives de suicide à vie, Québec

Q63 (Avez-vous déjà sérieusement songé à vous suicider ou à vous enlever la vie?)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP	149 630	.	149 630	.
Oui	877 165	14,2	1 026 795	14,2
Non	5 299 728	85,8	6 326 523	100,0

* La fréquence de la catégorie « NRP » n'a pas été redistribuée afin de présenter les valeurs à redistribuer.

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau 2-8 Effectif et proportions pondérées de la question 64 portant sur les idées suicidaires et les tentatives de suicide au cours des 12 derniers mois, Québec

Q64 (Est-ce que cela s'est produit au cours des 12 derniers mois?)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP	2 132	.	2 132	.
Oui	171 209	19,6	173 341	19,6
Non	703 824	80,4	877 165	100,0

* La fréquence de la catégorie « NRP » n'a pas été redistribuée afin de présenter les valeurs à redistribuer.

Source : Institut de la statistique du Québec, EQSP 2008.

⁴⁰ Il existe d'autres formes de non-réponse en cascade pour les indicateurs de l'EQSP qui ne sont pas liés à une question filtre; comme celle pour les indicateurs construits à partir de plus d'une question, mais dont aucune n'est une question filtre. C'est le cas de l'indicateur *Proportion de la population ayant eu des troubles musculo-squelettiques liés au travail actuel ou antérieur à au moins une région corporelle au cours des 12 derniers mois*, qui est construit à partir des questions 53A à 53D et 52A à 52D. Dans ce cas, la répartition de la non-réponse est plus compliquée, mais s'appuie sur les mêmes principes de base. Dans le présent document, pour une raison de simplicité, seule la non-réponse en cascade émanant d'une question filtre est discutée.

⁴¹ La question 64, qui permet de mesurer ce phénomène au cours des 12 derniers mois, n'a été posée qu'aux personnes ayant d'abord déclaré y avoir songé au cours de leur vie (question 63).

Les schémas 2 et 3 présentent la façon utilisée pour traiter la non-réponse partielle pour les données de l'EQSP lorsque les répondants de la question d'analyse sont identifiés par la réponse à une question filtre. Le premier schéma montre les catégories de réponse selon la codification originale des données, alors que le second montre les catégories obtenues une fois la non-réponse de la question filtre reportée en partie à la question d'analyse. Une fois ce traitement effectué sur la non-réponse, l'effectif de chaque catégorie de réponse peut être estimé adéquatement, c'est-à-dire sans biais⁴². D'abord, quelques explications sur la codification illustrée par les deux schémas.

Le schéma 2 présente la distribution originale des catégories de la question d'analyse et de la question filtre, alors que le schéma 3 présente la distribution de ces mêmes catégories une fois le traitement de la non-réponse effectué. On peut remarquer, dans ce dernier, que la catégorie « NRP » de la question filtre est scindée en deux parties. L'une d'elles (en gris pâle) est la proportion estimée des personnes possédant la caractéristique A, qui auraient donc dû répondre à la question d'analyse, mais qui n'ont pas répondu à la question filtre. Pour estimer sans biais l'effectif de population des catégories de la question d'analyse, on doit tenir compte de cette partie de la non-réponse de la question filtre.

La méthode de calcul est identique à celle décrite précédemment, c'est-à-dire qu'on multiplie l'effectif de la catégorie étudiée par l'inverse du taux de réponse à la question analysée. Toutefois, ce taux de réponse sera calculé en considérant maintenant la partie estimée de non-réponse à la question filtre. L'annexe 7 présente les détails du calcul pour l'estimation « en cascade » d'un effectif dans le cas des données présentées aux tableaux 2-7 et 2-8. Encore une fois, les résultats offerts sur le portail de l'Infocentre sont conformes à cette approche, comme en témoigne le tableau 2-9. Cependant, cette solution reposant sur la même hypothèse que celle portant sur l'estimation de proportions, à savoir que les répondants sont représentatifs des non-répondants partiels, est valable uniquement lorsque la non-réponse partielle n'est pas trop élevée. Dans le cas contraire, une analyse de biais est recommandée afin de confirmer la validité de l'hypothèse selon laquelle les caractéristiques des non-répondants ne sont pas différentes de celles des répondants⁴³.

Tableau 2-9 Proportions de la population qui a déjà songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois pour l'ensemble du Québec, EQSP 2008

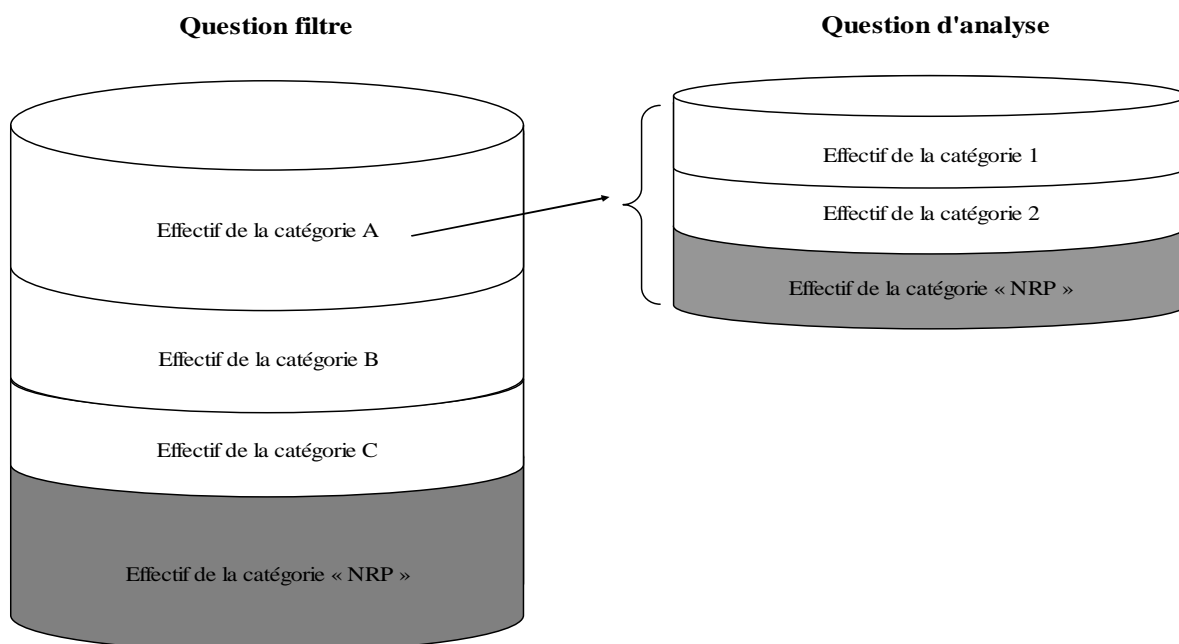
Statistiques					
Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
175 800	2,8	5,5	0,15	(2,5 - 3,1)	2,6

Source : Infocentre de santé publique, octobre 2010.

⁴² Sous l'hypothèse d'une non-réponse partielle distribuée aléatoirement.

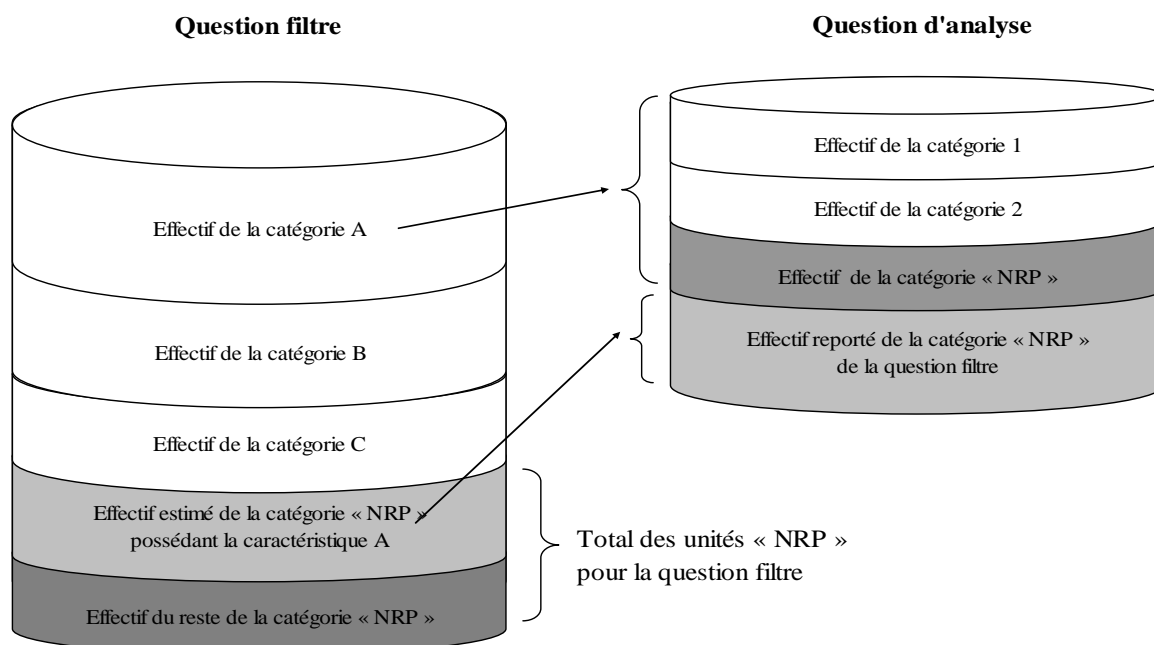
⁴³ Pour savoir comment procéder au diagnostic et à l'analyse de la non-réponse partielle, consultez l'annexe 5.

Schéma 2 Catégories de réponse selon la codification originale des données dans le cas d'une sous-population créée à partir d'une question filtre



Note: la catégorie « NRP » représente l'ensemble de la non-réponse partielle, ce qui veut dire qu'elle inclut les « ne sait pas », « ne répond pas », « refus global » et « non-réponse en raison de la déclaration par un tiers répondant ».

Schéma 3 Catégories de réponse après le traitement de la non-réponse partielle pour l'estimation d'effectif dans le cas d'une sous-population créée à partir d'une question filtre



3. Mesures de précision pour les proportions

3.1. Estimation de la variance

Il a été question à la section 2.2 du poids des répondants, qui sert à produire des estimations adéquates des prévalences d'intérêt de façon à ce qu'elles puissent être inférées à l'ensemble de la population visée par l'enquête. Dans cette section, nous aborderons une autre forme de poids, les poids bootstrap, qui vous seront expliqués à la section 3.1.1. Ceux-ci ne remplacent pas la variable principale de pondération. Ils constituent une façon d'estimer la précision des prévalences analysées. Nous verrons également que cette approche, des poids bootstrap, nous permettra de construire des intervalles de confiance selon une méthode empirique (section 3.3.3), qui nous permettra d'effectuer des tests en s'affranchissant de l'hypothèse habituelle de normalité des observations (section 5.1.2).

3.1.1. Définition et méthodes

Comme toutes les estimations d'enquête par échantillon, celles tirées de l'EQSP sont entachées d'erreurs puisque les données proviennent d'un échantillon plutôt que de l'ensemble de la population. Lorsqu'on infère les caractéristiques d'une population à partir d'un échantillon, on ne peut parler de valeurs exactes, mais bien d'une estimation de ces valeurs exactes. Il s'avère donc nécessaire d'estimer la précision de chaque proportion que l'on extrait de l'échantillon.

La complexité du plan de sondage de l'EQSP influence grandement la précision des estimateurs. En particulier, la répartition régionale non proportionnelle, dont l'impact est accentué par l'achat d'unités supplémentaires par des régions peu peuplées, fait diminuer la précision par rapport à la précision d'estimations faites à partir d'un échantillon aléatoire simple de même taille. L'impact n'est pas négligeable et doit être inclus dans l'estimation de la précision.

Les dernières versions de certains logiciels d'analyse statistique, tels SPSS ou SAS, offrent maintenant des procédures⁴⁴ qui permettent d'obtenir des estimations sans biais de la précision pour certains plans de sondage complexes comme celui utilisé dans l'EQSP. En effet, il est maintenant possible d'estimer la variance à l'aide d'une méthode de linéarisation, aussi appelée approximation par séries de Taylor.

D'autres logiciels spécialisés comme SUDAAN ou WESVAR peuvent aussi être utilisés pour calculer correctement la précision des estimations faites à partir d'un plan complexe. Ces logiciels ont l'avantage d'offrir plusieurs méthodes d'estimation de la variance qui tiennent compte de la complexité du plan de sondage : approche par linéarisation et approches dites par « réplification », dont fait partie le bootstrap⁴⁵.

Peu importe le logiciel utilisé, l'emploi de l'une ou l'autre des méthodes mentionnées ci-dessus nécessite d'avoir accès à l'information du plan de sondage. Pour utiliser la méthode de linéarisation, il faut fournir les indicateurs de strate d'échantillonnage avec les microdonnées. Pour utiliser la méthode bootstrap, il suffit d'avoir accès aux ensembles de poids bootstrap, étant donné que ce sont eux qui « reflètent » la façon dont l'échantillon a été sélectionné.

Pour l'EQSP, il a été convenu d'estimer la précision des estimations produites à partir des données de l'enquête à l'aide de la méthode bootstrap. Pour cela, l'ISQ a produit des ensembles de poids bootstrap qui accompagnent le fichier maître des microdonnées de l'enquête. Une copie de ce fichier, fourni par l'ISQ en raison du consentement obtenu des répondants, est également disponible à l'Infocentre. Il contient donc également l'information du plan de sondage nécessaire aux logiciels. Pour ce qui est des

⁴⁴ Dans la version 9 de SAS, ces procédures sont offertes dans le module de base, alors que pour SPSS (version 17), il faut se procurer le module complémentaire « Complex samples » pour y avoir accès.

⁴⁵ Le lecteur trouvera un traitement détaillé des méthodes de réplification dans les ouvrages de Lohr (1999), Rust et Rao (1996) et Wolter (1985).

fichiers masqués, FMII et FARE, ils ne contiennent pas tous les éléments du plan de sondage. En effet, les poids bootstrap ne sont pas fournis pour ce type de fichier. Par contre, une information équivalente aux strates d'échantillonnage est incluse dans les fichiers, de sorte que les utilisateurs de ces fichiers peuvent estimer adéquatement la précision des estimations en utilisant un logiciel approprié.

De façon très résumée, voici comment fonctionne la méthode bootstrap (plus de détails sont fournis à l'annexe 8). Pour obtenir une estimation de la variance d'une proportion (statistique calculée à partir du poids de sondage), il suffit de recalculer cette même proportion 500 fois en utilisant les 500 ensembles de poids bootstrap fournis avec les données de l'enquête. Cette opération nous donne une distribution de valeurs possibles pour notre estimateur. La variabilité observée entre les 500 résultats représente l'estimation de la variance. Cette distribution de valeurs nous permet également de calculer différentes mesures de précision (voir la section 3.3.3).

3.1.2. Impact de l'imputation sur la variance

L'imputation de valeurs pour compenser l'absence de réponse à certaines questions est un moyen fréquemment utilisé dans les enquêtes par sondage. Or, il se pose dans la pratique un problème d'estimation de la variance lorsqu'une partie des données sont attribuées par imputation. En effet, il est généralement reconnu que traiter des valeurs imputées comme s'il s'agissait de valeurs observées conduit à une sous-estimation de la variance. Cette sous-estimation a tendance à croître avec le taux de valeurs imputées et peut être considérable si ce taux est élevé.

Nous avons mentionné à la section 2.4 que l'imputation a été utilisée pour traiter la non-réponse dans deux situations précises; une erreur de passage de la question 54 et le refus ou l'incapacité de fournir le revenu total du ménage à la question 120. Dans le cas de la question 54, la non-réponse n'est pas caractérisée, c'est-à-dire qu'elle n'est pas attribuable à certaines sous-populations précises, parce qu'il s'agit d'une erreur de programmation et non d'une décision délibérée du répondant de ne pas répondre. L'impact de l'imputation sur la variance est dans ce cas relativement uniforme. Parce que la variance est également relativement faible pour l'indicateur créé à partir de cette question (de l'ordre de 5 %), elle peut être considérée comme étant négligeable.

Par contre, dans le cas de la question 120 sur le revenu (de même que l'indice du niveau de revenu), la non-réponse est très caractérisée (voir section 2.4.3) et son impact sur la variance ne peut être considéré comme négligeable, sauf dans une situation précise, qui est décrite à la section 6.3.4.

3.2. Coefficient de variation

3.2.1. Définition et méthode de calcul

La précision d'une estimation peut s'exprimer sous diverses formes : marge d'erreur, coefficient de variation, intervalle de confiance, cote, etc. Peu importe la forme qu'elle prend, la prise en compte du plan de sondage est essentielle dans le calcul de la mesure de précision. La mesure de précision recommandée par l'ISQ pour les données d'enquête de santé est le coefficient de variation.

Le coefficient de variation (C.V.) est une mesure de précision relative. Exprimé en pourcentage, le C.V. s'obtient très facilement en calculant le rapport entre l'erreur-type de l'estimation et l'estimation elle-même. L'erreur-type d'une estimation correspond à la racine carrée de la variance de l'estimation. Par exemple, pour une proportion estimée :

$$\text{C.V.} = \frac{\text{Erreur-type de la proportion}}{\text{Proportion}}$$

Plus le C.V. est petit, meilleure est la précision. À l'inverse, un C.V. élevé commande à la prudence.

3.2.2. Recommandations sur les seuils de diffusion

Bien entendu, certaines estimations ont une excellente précision alors que d'autres ont une précision moins bonne. Le potentiel d'analyse des premières est clair alors que celui de la deuxième catégorie soulève plus de questionnements. L'ISQ utilise un système de classification des estimations, en matière de précision, pour la diffusion des résultats produits à partir de ses enquêtes. Le tableau 3-1 présente ce système. Contrairement à d'autres organismes qui ne publient pas les estimations qui ont une précision au-delà d'un certain seuil, l'ISQ recommande plutôt, pour juger de la pertinence de la diffusion d'une statistique, d'examiner l'utilité de cette statistique par rapport aux autres sources disponibles. L'ISQ recommande également, dans le cas de statistiques peu précises, d'émettre des indications claires sur les limites à l'usage de ces estimations.

En dernière analyse, la responsabilité de la diffusion d'un résultat et du choix de la norme à utiliser incombe à l'utilisateur. Cependant, dans un souci d'uniformisation des normes utilisées pour toutes les enquêtes sociosanitaires au Québec, le Comité tripartite MSSS-INSPQ-ISQ a décidé que les normes de l'ISQ seront utilisées pour les résultats des indicateurs de l'EQSP déployés sur le portail de l'Infocentre.

Tableau 3-1 Système de classification des estimations, en termes de précision, en vigueur à l'Institut de la statistique du Québec

Précision de l'estimation (C.V.)	Signe distinctif	Réserve à mentionner lors de la diffusion
$C.V. \leq 15\%$ ¹	(aucun signe)	Estimation diffusée sans mention
$15\% < C.V. \leq 25\%$	*	Estimation à interpréter avec prudence
$C.V. > 25\%$	**	Estimation à utiliser avec circonspection, fournie à titre indicatif seulement

1. Au besoin, il est possible de définir des catégories plus fines pour les estimations dont les C.V. sont inférieurs à 15 %. Par exemple, les cotes A : ($0\% < C.V. \leq 5\%$); B : ($5\% < C.V. \leq 10\%$) et C : ($10\% < C.V. \leq 15\%$) peuvent être utilisées.

Source : Institut de la statistique du Québec.

3.3. Intervalle de confiance

3.3.1. Définition

Une autre mesure de la précision d'une estimation est l'étendue de l'intervalle de confiance (IC) : plus l'intervalle est court, mieux la valeur du paramètre est circonscrite. L'intervalle de confiance associé à une proportion représente la zone d'incertitude liée à cette estimation. Cette zone d'incertitude est étroitement liée au niveau de confiance choisi. Ainsi, un intervalle de confiance pour une proportion, à un niveau de confiance de 95 %, signifie que si on répétait l'enquête 100 fois et qu'à chaque reprise on estimait la proportion et l'on calculait son intervalle de confiance, alors 95 des 100 intervalles ainsi créés contiendraient la vraie valeur de la proportion dans la population.

Il y a plusieurs façons de construire un intervalle de confiance. Certains se distinguent par le choix de la loi servant à l'approximation de la loi de l'estimateur. Les méthodes standards utilisent habituellement la loi Normale ou la loi Binomiale comme approximation de la loi de l'estimateur pour construire un intervalle. Ces intervalles seront discutés à la section 3.3.2.

Il existe également d'autres méthodes qui permettent de s'affranchir de l'hypothèse que les données suivent une loi connue; la méthode basée sur le calcul des percentiles de la distribution empirique bootstrap est l'une d'elles. Ces intervalles de confiance, qu'on appellera intervalles bootstrap, ne doivent pas obligatoirement être ceux utilisés lorsqu'on emploie des ensembles de poids bootstrap pour estimer la

variance (voir section 3.1.1). Par contre, ils constituent une option supplémentaire lorsque ceux-ci sont utilisés.

En somme, le choix de la méthode pour construire des intervalles de confiance est indépendant du choix de la méthode pour estimer la variance. Les prochaines sections présentent deux méthodes différentes pour construire des intervalles de confiance pour les estimations produites à partir des données de l'EQSP. D'abord, il sera question des intervalles de confiance standards (section 3.3.2). Par la suite, les intervalles de confiance bootstrap seront discutés (section 3.3.3).

3.3.2. Méthode de calcul pour les intervalles de confiance standards

Dans le cas de la loi Normale, l'intervalle de confiance est déterminé à l'aide de la marge d'erreur, qui se calcule de la façon suivante :

$$\text{marge d'erreur} = z_{\alpha/2} * \text{erreur-type de l'estimation}$$

où $z_{\alpha/2}$ est la valeur d'une loi Normale centrée réduite associée à un niveau de confiance de $(1-\alpha)$ %. Par exemple, si α est fixé à 5 %, alors $z_{\alpha/2} = 1,96$. L'intervalle de confiance pour une proportion, lorsque l'approximation par la loi Normale est utilisée, se définit alors comme suit :

$$\text{IC} = \text{estimation} \pm \text{marge d'erreur}$$

L'intervalle de confiance basé sur la loi Normale est symétrique⁴⁶ et il a l'avantage de fournir, dans un seul terme, deux informations à la fois : une idée de la valeur de l'estimation (la valeur centrale) d'une part, et sa précision (la longueur de l'intervalle) d'autre part.

Dans le cas des proportions, la plupart du temps, la loi Normale est une bonne approximation de la loi de l'estimateur de la proportion. Dans certaines conditions, principalement lorsque la proportion étudiée est faible et que le nombre de répondants est petit, il s'avère plus approprié d'utiliser une autre loi. La loi Binomiale est alors le meilleur choix. La construction de l'intervalle de confiance Binomial est plus complexe et les détails de sa construction ne sont pas présentés dans ce guide. Cependant, les conditions d'application pour le choix de la loi Normale ou la loi Binomiale sont présentées à l'annexe 9. Dans le cas de l'intervalle de confiance Binomial, la marge d'erreur perd son sens en raison de l'asymétrie de cet intervalle. En effet, la demi-longueur à gauche de l'estimation ponctuelle n'est plus égale à la demi-longueur à droite.

À l'ISQ, plusieurs enquêtes utilisent et produisent des intervalles de confiance qui sont basés sur l'une ou l'autre de ces deux méthodes, Normale et Binomiale, que nous appellerons les méthodes standards.

3.3.3. Méthode de calcul pour les intervalles de confiance bootstrap

Pour l'EQSP, une autre approche a été utilisée en remplacement des méthodes standards. Il s'agit d'intervalles de confiance bootstrap. Cette approche utilise directement la loi de la distribution empirique des estimations, ce qui a pour avantage de s'affranchir de l'hypothèse selon laquelle les données suivent une loi connue. Cet avantage a sa principale répercussion au niveau des tests statistiques. Plus de détails sont fournis à la section 5.1.2.

⁴⁶ Ce qui veut dire que la longueur entre la borne inférieure de l'intervalle et l'estimation ponctuelle (au centre) est la même que celle entre l'estimation ponctuelle et la borne supérieure de l'intervalle.

Cette approche est valide pour les données de l'EQSP et dans un contexte d'exploitation spécifique, tel celui de l'ISQ et de l'Infocentre. Elle pourrait, pour diverses raisons, ne pas être appropriée pour d'autres situations ou d'autres enquêtes. Une évaluation des avantages et des désavantages devrait être faite, le cas échéant, avant d'utiliser cette méthode.

La littérature suggère plusieurs méthodes pour construire des intervalles de confiance bootstrap. Celle retenue pour l'EQSP est la méthode basée sur le calcul des percentiles de la distribution empirique bootstrap, à laquelle nous avons appliqué une correction pour la continuité. De façon très succincte, on peut dire qu'un intervalle de confiance basé sur les percentiles bootstrap s'obtient de la façon suivante : on se sert de chacun des ensembles de poids bootstrap discuté à la section 3.1.1, disons B ensembles de poids, pour calculer B estimations ponctuelles, par exemple une proportion. On ordonne ensuite ces B valeurs en ordre croissant, puis on attribue comme bornes de l'intervalle de confiance, de niveau de confiance 95 %, les 2,5^e et 97,5^e percentiles de la distribution empirique ordonnée⁴⁷.

Comme il a été mentionné à la section 3.1.1, l'estimation de la précision des estimations repose sur une approche dite par « réplication », soit celle de la méthode bootstrap. Encore une fois selon la littérature, un nombre de 500 répliques (ou ensembles de poids bootstrap) est considéré suffisant pour estimer adéquatement la précision des données d'enquêtes. Toutefois, dans le cas de l'estimation de percentiles (nécessaires à la construction des intervalles de confiance bootstrap), ce nombre peut s'avérer grandement insuffisant. En effet, un nombre trop faible de répliques engendre une variabilité des bornes de l'intervalle qui n'est pas souhaitée. C'est pourquoi une évaluation détaillée de l'utilisation des intervalles de confiance bootstrap a été réalisée par l'ISQ au début de l'année 2010⁴⁸. Cette étude avait deux objectifs principaux : 1) savoir si le nombre de 500 ensembles de poids bootstrap était suffisant pour obtenir des intervalles de confiance stables et comparables à ceux obtenus à l'aide des méthodes standards, et 2) dans la négative, déterminer le nombre d'ensembles de poids bootstrap nécessaires.

L'étude de l'ISQ conclut que 500 ensembles de poids bootstrap, bien que suffisant pour estimer la précision des estimations, ne sont pas suffisants pour estimer adéquatement les percentiles de la distribution empirique bootstrap. L'étude recommande d'ailleurs qu'au minimum 2 000 répliques doivent être utilisées pour répondre à cet objectif, qui est d'estimer adéquatement les percentiles d'une distribution bootstrap afin de construire des intervalles de confiance bootstrap de niveau de confiance de 95 %.

Par conséquent, il a été décidé par le Comité tripartite MSSS-INSPQ-ISQ que, pour des raisons d'efficacité et de qualité, 500 ensembles de poids bootstrap seraient utilisés pour estimer la variance des estimations analysées, dont pour le test du Khi-deux, mais que 2 000 ensembles de poids bootstrap⁴⁹ seraient utilisés pour la construction des intervalles de confiance bootstrap. Ces 2 000 ensembles de poids bootstrap ont donc été créés pour l'EQSP et sont disponibles avec les données du fichier maître.

3.3.4. Recommandations relatives au choix du niveau de confiance

En raison du choix de retenir l'intervalle de confiance bootstrap pour les données de l'EQSP et des contraintes qui en découlent quant au nombre d'ensembles de poids bootstrap à utiliser, il est recommandé d'utiliser uniquement des intervalles bootstrap de niveau de confiance de 95 %⁵⁰. Par contre, l'utilisateur qui souhaiterait produire des intervalles de confiance standards à partir d'un fichier masqué

⁴⁷ Auxquelles on applique une correction pour la continuité.

⁴⁸ Rapport à paraître.

⁴⁹ Les 500 premiers ensembles servent à estimer la variance des estimations, auxquels s'ajoutent 1 500 autres ensembles de poids pour la construction des intervalles de confiance.

⁵⁰ C'est d'ailleurs ce qui est produit sur le portail de l'Infocentre de santé publique. Les autres niveaux de confiance permis pour d'autres diffusions ne sont pas disponibles pour l'EQSP.

ou en prenant l'erreur-type fournie dans les tableaux produits à l'Infocentre est libre d'utiliser le seuil qui convient à ses analyses.

Pour plus d'information concernant le niveau de confiance d'un intervalle, consultez l'annexe 10.

4. Ajustement des proportions selon l'âge

4.1. Définition

Dans la plupart des situations, les proportions issues des données d'enquête sont estimées sans égard à la façon dont la population est constituée. Ces proportions sont dites brutes. Elles permettent ainsi d'estimer le fardeau réel d'un facteur de risque ou la prévalence d'une maladie ou de tout autre indicateur de santé dans la population. Mais qu'en est-il de leur utilisation dans la comparaison de deux populations différentes en fonction de structure d'âge?

Bien que l'ajustement des proportions selon l'âge soit une pratique peu courante pour des données d'enquête, nous suggérons de produire des proportions ajustées lorsque le phénomène qu'on veut étudier est fortement lié à l'âge et que l'on souhaite effectuer des comparaisons. C'est le cas, entre autres, des indicateurs de l'état de santé pour lesquels l'âge est un déterminant majeur. Il est en effet normal, par exemple, d'observer plus de problèmes de santé au sein d'une population dont la structure d'âge indique une plus grande proportion de gens âgés. Cet ajustement selon l'âge permet donc d'épurer les mesures de l'effet confondant de l'âge, afin de pouvoir comparer les mesures de deux populations ayant des structures d'âge différentes. Mentionnons que les proportions ajustées n'ont de signification que lorsqu'elles sont utilisées pour comparer l'état de santé de diverses populations; elles ne sont que des mesures synthétiques. Par conséquent, elles ne représentent aucunement une mesure du fardeau réel supporté par les populations, contrairement aux proportions brutes (Fleiss, 1981).

4.2. Méthode de calcul

Nous recommandons de calculer les proportions ajustées selon l'âge en utilisant la méthode d'ajustement directe qui consiste à appliquer les proportions par âge d'une population, pour une année donnée, à la structure d'âge d'une population de référence. Voir l'annexe 11 pour un exemple détaillé de calcul de proportions ajustées selon l'âge. Ces proportions synthétiques reflètent ainsi ce qu'auraient été les proportions brutes des populations étudiées si celles-ci avaient la même structure d'âge que la population de référence (Muecke *et coll.*, 2005). La population de référence est généralement celle de l'ensemble du Québec pour les comparaisons régionales ou du Canada pour les comparaisons provinciales au plus récent recensement, sexes réunis et corrigée pour le sous-dénombrement. Toute autre population de référence peut également être considérée dépendamment du type et de la période de comparaison. Mentionnons enfin qu'il est possible d'utiliser différentes structures d'âge de la population de référence. Il est cependant important de considérer des groupes d'âge pour lesquels nous pouvons obtenir des estimations de proportions avec une assez bonne précision. Pour l'EQSP, nous proposons la structure par âge suivante pour les indicateurs se rapportant à la population des 15 ans et plus : 15-24 ans, 25-44 ans, 45-64 ans, 65-74 ans et 75 ans et plus.

S'il n'y a aucun répondant pour au moins une des cinq catégories d'âge de la standardisation, il n'est pas possible de calculer la proportion ajustée, à moins de faire un regroupement de catégories d'âge. Sur le portail de l'Infocentre, la possibilité de regrouper les catégories d'âge de standardisation n'est pas offerte, par conséquent un message d'erreur s'affiche. Ce message est le suivant : « Pour au moins une des sous-populations, les proportions ajustées par la standardisation directe ne sont pas calculables ».

4.3. Variance de proportions ajustées

Au même titre que pour le calcul des proportions brutes, la variance des proportions ajustées est calculée en utilisant les ensembles de poids bootstrap. Cela revient tout simplement à calculer pour chaque ensemble de poids bootstrap la proportion ajustée selon l'âge et de prendre la variance de ces estimations de proportions ajustées. L'utilisation des poids bootstrap est également recommandée pour l'obtention des intervalles de confiance, comme expliqué à la section 3.3.3.

Toutefois, si les poids bootstrap ne sont pas disponibles dans les fichiers (par exemple les fichiers masqués FARE et FMII), il est tout de même possible d'obtenir une approximation grossière de cette variance. Il s'agit d'abord de calculer indépendamment la variance de la proportion brute de chacun des groupes d'âge de l'ajustement en considérant les variables du plan de sondage dans les logiciels statistiques (comme vu à la section 3.1.1). Ensuite, l'approximation de la variance est obtenue en prenant la somme pondérée de ces variances. Le poids considéré est celui de la proportion au carré que représente le groupe d'âge dans la population de référence (voir à cet effet un exemple à l'annexe 11). On parle ici d'une approximation, car cette façon de faire ne tient pas compte des covariances qui peuvent exister, attribuables au plan de sondage complexe, entre les proportions estimées de deux différents groupes d'âge, ce qui est pris en compte avec la méthode utilisant les poids bootstrap.

Pour le calcul des proportions ajustées, la population de référence qui a été retenue est celle de l'enquête proprement dite. Pour retrouver les chiffres de population de chaque groupe d'âge, il suffit de prendre la somme des poids de l'enquête ou autrement dit, les effectifs de l'enquête par groupe d'âge. Les variances et les intervalles de confiance ont été obtenus à l'aide des méthodes utilisant les poids bootstrap.

5. Comparaison de proportions

La section 5 de ce guide a pour objectif de présenter les tests statistiques qui doivent être utilisés pour analyser adéquatement les données de l'EQSP 2008. L'ensemble de l'information et des recommandations présentées dans les précédentes sections du guide converge vers une utilisation de qualité des données. Les tests statistiques, abordés dans la prochaine section, visent le même objectif. On y traite notamment de la théorie des tests statistiques ainsi que leurs avantages, leurs limites et les conditions d'application.

Par la suite, à la section 6, il sera question de la comparabilité des données de l'EQSP, des particularités de certaines variables, indicateurs ou indicateurs associés, et de ce qu'il faut savoir pour exploiter les données de l'EQSP, plus précisément dans le cas de l'utilisation des résultats diffusés sur le portail de l'Infocentre.

5.1. Tests statistiques

Dans la majorité des cas, lorsqu'on désire analyser un indicateur du PCS, on souhaite vérifier si celui-ci varie en fonction d'une variable de croisement, tels l'âge ou le sexe. Un test global d'association (section 5.1.1) permet de répondre à cette question. Une fois que l'on sait que l'indicateur prend des valeurs différentes selon les catégories de cette variable de croisement, il s'avère intéressant d'identifier exactement ces catégories. Pour ce faire, on compare ces catégories, par exemple chaque groupe d'âge, à l'aide d'un test de la différence de deux proportions (section 5.1.2). Le lecteur désirant obtenir de l'information générale sur les tests statistiques, leur utilité et ce qu'ils apportent aux analyses, est invité à consulter l'annexe 12.

Au même titre que les estimations de variance décrites à la section 3.1, les tests statistiques présentés dans les deux prochaines sections doivent, pour être adéquats, être corrigés pour tenir compte du plan de sondage de l'enquête. Les tests doivent notamment intégrer la pondération de l'enquête.

5.1.1. Test global d'association

Nous l'avons vu précédemment, lorsqu'on souhaite comparer des estimations de proportions pour des indicateurs à plus de deux catégories ou lorsqu'on analyse un indicateur selon une variable de croisement à plus de deux catégories⁵¹ tout en ayant comme objectif de contrôler le seuil global du test, la première démarche à entreprendre est la production d'un test global d'association.

Le test du khi-deux peut être utilisé pour comparer globalement un indicateur pour l'ensemble des catégories d'une variable de croisement, par exemple l'âge. En d'autres termes, ce test nous permet de confirmer ou d'infirmer l'hypothèse selon laquelle des valeurs différentes de l'indicateur peuvent être observées selon la catégorie de la variable de croisement⁵².

Au même titre que les estimations de variance produites à partir des données de l'EQSP, qui doivent intégrer la pondération et les poids bootstrap⁵³ pour tenir compte correctement du plan de sondage complexe de l'enquête, il est recommandé d'utiliser un test du khi-deux qui, lui aussi, intègre ces éléments de sorte qu'il soit, lui aussi, adéquat. Par ailleurs, il est important de s'assurer que les conditions de validité du test du khi-deux soient satisfaites. Les résultats produits sur le portail ont été vérifiés afin de s'assurer que ces conditions sont remplies. Dans le cas contraire, une note s'affiche au-dessous du tableau, à l'endroit où est normalement indiqué le résultat du test global. Pour connaître les conditions de validité du test ou pour savoir quoi faire lorsque le test n'est pas valide, consultez l'annexe 13.

Ce n'est que lorsque le test global est significatif au seuil voulu qu'il est justifié d'aller plus loin et d'examiner plus attentivement les proportions présentant les écarts les plus importants. Cette démarche est discutée à la prochaine section. Cependant, il peut arriver, dans de rares occasions, que le test global du khi-deux soit significatif sans qu'aucune des comparaisons deux à deux (voir section 5.1.2) ne présente de différence significative. La principale raison expliquant un tel résultat est que le test du khi-deux est plus puissant que les comparaisons deux à deux, car il évalue la relation globale entre deux variables. Plus de détails sont donnés à la fin de la prochaine section.

5.1.2. Test de la différence de deux proportions

Comme mentionné précédemment, une fois la relation significative identifiée entre un indicateur et une variable de croisement (par un test global d'association significatif), il est de mise d'examiner plus en détail les proportions qui définissent cette relation. Pour ce faire, on utilise un test qui permet de comparer directement les proportions entre elles. Il faut mentionner que dans le cas d'un indicateur à deux catégories seulement, et analysé selon une variable de croisement qui ne possède également que deux catégories (comme le sexe), le test décrit dans cette section est équivalent au test global du khi-deux présenté à la section précédente. Dans une telle situation, il est inutile d'effectuer au préalable un test du khi-deux.

Pour les besoins du portail de l'Infocentre, le test retenu par le Comité tripartite MSSS-INSPQ-ISQ pour évaluer si la différence entre deux proportions est significative est le test empirique de la différence de deux proportions (méthode 1). Cette méthode est basée sur l'utilisation des ensembles de poids bootstrap de l'enquête et a l'avantage de s'affranchir de l'hypothèse selon laquelle les données suivent approximativement une loi connue (Normale ou Binomiale). Cet avantage se traduit par le fait que ce test peut être utilisé, peu importe les circonstances, contrairement au test utilisé pour les données de l'ESCC à

⁵¹ Dans le cas d'un indicateur à deux catégories croisé selon une variable de croisement à deux catégories, se référer au test de la section 5.1.2, car ceux-ci deviennent équivalents.

⁵² Par exemple, que le niveau de détresse psychologique diffère selon les groupes d'âge.

⁵³ Comme mentionné à la section 3.3.3, seulement 500 ensembles de poids bootstrap sont utilisés pour calculer la statistique du test du khi-deux.

l'Infocentre⁵⁴. Bien que les détails de la construction de ce test sont présentés à l'annexe 14, mentionnons brièvement que le concept des poids bootstrap est encore ici mis à profit. En effet, le test de la méthode 1 repose sur la construction d'un intervalle de confiance bootstrap de la différence des proportions. Comme décrit à la section 3.3.3, il suffit de calculer la différence des proportions pour chaque ensemble de poids bootstrap et, à partir de la distribution ordonnée des valeurs obtenues, construire l'intervalle de la différence des proportions en prenant le 2,5^e percentile et le 97,5^e percentile de cette distribution. Le résultat du test s'obtient très simplement : on peut conclure que deux proportions sont différentes au seuil de 5 % lorsque l'intervalle de confiance de la différence ne contient pas la valeur zéro. La puissance du test de la méthode 1 est du même ordre de grandeur que celle d'un test de Student (test *t*).

Toutes les comparaisons 2 à 2 produites à l'Infocentre sont le résultat du test de la méthode 1. Cependant, dans certaines circonstances, les comparaisons 2 à 2 ne sont pas produites par la requête (voir la section 6.3). Dans ces cas, l'utilisateur pourra employer une autre méthode pour comparer les proportions. Il s'agit de la méthode basée sur la construction et la comparaison de deux intervalles de confiance, un pour chacune des proportions estimées (méthode 2). En bref, l'utilisateur peut conclure que deux proportions sont différentes lorsque les deux intervalles de confiances sont disjoints. Des précisions sur le test de la méthode 2 sont également disponibles à l'annexe 14.

Enfin, soulignons que le test basé sur la construction de deux intervalles de confiance (méthode 2) est plus conservateur que le test empirique de la différence de deux proportions (méthode 1), ce qui veut dire qu'il est moins puissant et qu'il rejette moins souvent qu'il devrait l'hypothèse d'égalité des proportions, au seuil désiré⁵⁵. C'est notamment pour cette raison que le Comité tripartite MSSS-INSPQ-ISQ a décidé de privilégier le test de la méthode 1 lorsque celui-ci peut être produit.

Cependant, il peut arriver, dans de rares occasions, que le test global du khi-deux soit significatif sans qu'aucune des comparaisons deux à deux ne présente de différence significative. Cette situation peut survenir lorsqu'une relation existe réellement entre deux variables (khi-deux significatif), mais qu'elle est si diffuse parmi l'ensemble des catégories qu'elle n'est pas détectable lorsqu'on l'étudie du point de vue des comparaisons deux à deux, en particulier si la méthode 2 est utilisée pour effectuer ces comparaisons. Pour l'EQSP, certains éléments méthodologiques peuvent également contribuer à obtenir un tel résultat. En effet, pour le test du khi-deux, la statistique du test est calculée en utilisant 500 ensembles de poids bootstrap et en supposant que les observations suivent une loi connue. Pour le test de la différence de deux proportions, la méthode utilisée permet de s'affranchir de l'hypothèse que les observations suivent une loi connue et un nombre supérieur d'ensembles de poids bootstrap est utilisé (2 000 ensembles de poids).

Lors de l'interprétation des résultats d'un test statistique, certaines précautions s'imposent. À ce titre, le chapitre III présente un certain nombre de recommandations pour l'interprétation de résultats provenant de données d'enquête.

6. Exploitation des données de l'EQSP

La présente section aborde des éléments à considérer pour faire une saine exploitation des données de l'EQSP 2008. Plus spécifiquement, la section 6.1 porte sur la comparabilité de l'enquête avec d'autres sources de données. La section 6.2 recense les possibilités, en matière de production statistique, qu'offre le portail de l'Infocentre pour les données de l'EQSP. Finalement, la section 6.3 rassemble les principales

⁵⁴ Pour les données de cette enquête à l'Infocentre, le test utilisé dépend de la méthode d'estimation de chacune des deux proportions comparées. Si la méthode d'estimation des deux proportions est la méthode normale, alors le résultat affiché est celui d'un test *t* (méthode 1). Si par contre, la méthode d'au moins une des deux proportions comparées est la méthode binomiale, alors la méthode 1 ne peut être utilisée, et le résultat d'un test plus conservateur est affiché, soit celui de la comparaison de deux intervalles de confiance (méthode 2). Pour plus de détails, consulter ISQ (2009).

⁵⁵ Voir à ce sujet l'article de Schenker et Gentleman (2001).

particularités propres aux indicateurs de l'EQSP et aux résultats diffusés par l'Infocentre. On y présente notamment les choix faits pour certains indicateurs et les impacts sur leur utilisation et leur interprétation. Notez que la section 6.3 n'a pas pour but de couvrir l'ensemble des choix faits pour l'EQSP, mais seulement de donner de l'information supplémentaire sur certains choix faits par le Comité tripartite MSSS-INSPQ-ISQ.

6.1. Comparabilité des résultats de l'EQSP avec d'autres sources de données

Avant de procéder à une comparaison des résultats de l'EQSP avec d'autres sources de données, il est important de se questionner sur la comparabilité de celles-ci. La présente section n'est pas le fruit d'une étude sur la comparabilité des données de l'EQSP. Elle vise seulement à mettre en évidence les éléments qui peuvent avoir un impact sur cette comparabilité : la base de sondage, la population visée, la méthode de collecte, la formulation des questions, l'administration des questions lors de l'entrevue et l'univers couvert par l'indicateur.

Pour l'EQSP, certaines questions posées dans le questionnaire proviennent intégralement de questionnaires d'autres enquêtes. C'est le cas par exemple des questions sur la perception générale de l'état de santé (Q1), de la consommation de tabac (Q68) et des questions permettant de mesurer la santé mentale – échelle K6 – (Q57 à Q62) qui proviennent de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) de Statistique Canada. Malgré la correspondance parfaite des questions des deux enquêtes, certains facteurs peuvent nuire à la comparabilité des données : des bases de sondage ou des années de référence différentes, ou des modes de collecte distincts. Le contexte de l'enquête en soi peut également avoir un impact.

La base de sondage de l'EQSP étant une base de numéros de téléphone générés aléatoirement (GANT), une comparaison avec une enquête utilisant une base aréolaire pourrait présenter des limites quant à la comparabilité des données. En effet, la couverture des deux bases n'est pas la même, celle d'une base aréolaire étant supérieure à la couverture d'une base téléphonique. Par ailleurs, la collecte des données de l'EQSP est réalisée au téléphone. Or, il a été démontré (St-Pierre et Béland, 2004) que la déclaration au téléphone peut être différente de la déclaration en face à face.

Il arrive également, dans certains cas, que les contextes des enquêtes ne soient pas tout à fait identiques. Par exemple, la consommation de tabac mesurée dans une enquête sur la santé générale versus la consommation mesurée dans une enquête sur le tabagisme peut donner des résultats très différents (Gilmore, 2002). Par ailleurs, extraire un sous-ensemble de questions d'un module plus étoffé d'un questionnaire d'une enquête afin de l'introduire dans le questionnaire d'une autre enquête peut mener à des résultats différents en raison de l'absence de certaines questions du module étoffé qui change le contexte général du module de questions. Aussi, énumérer ou ne pas énumérer les choix de réponse offerts au répondant peut également influencer la comparabilité des données.

Il est possible qu'une formulation différente d'une même question dans deux enquêtes distinctes puisse n'avoir aucun impact lorsqu'on peut procéder à la reconstitution des variables pour les rendre similaires. Dans le même ordre d'idée, lorsque la seule différence entre deux enquêtes concerne l'univers des répondants, qui n'est pas le même (15 ans et plus vs 12 ans et plus), alors il faut simplement définir un domaine d'étude qui est commun aux deux enquêtes (même population des 15 ans et plus) afin de rendre possible la comparaison. Voici d'autres exemples de différences qui peuvent être contournées pour permettre une comparaison avec d'autres sources :

- Scission des questions en deux questions distinctes.
- Ajout d'une ou plusieurs catégories de réponses à une question qui existe dans une autre enquête. Dans certains cas, il est possible de corriger la situation en regroupant des catégories.

- Changement sur l'univers des questions, ce qui concerne le plus souvent le groupe d'âge visé par la question.

Ces situations peuvent aussi mener à une impasse pour la comparabilité, il n'est pas toujours possible de regrouper les catégories, de recréer le même univers ou d'évaluer l'impact de l'ajout de précisions dans les catégories ou dans l'introduction. D'ailleurs, certaines modifications ont un impact majeur qui empêche la comparabilité des résultats. Parmi les modifications qui rendent difficile, voire impossible, toute comparaison avec des données d'autres sources, on cite :

- L'ajout d'exemples pour illustrer les catégories de réponses. Cela pourrait avoir pour conséquence de faire augmenter les prévalences observées des variables issues de ces catégories.
- Le changement du libellé de la question.
- Le changement des choix de réponses ou changement dans les libellés des choix de réponses.
- Le changement de l'univers des questions, sans retour possible.

Selon l'importance de l'impact, il peut y avoir une mise en garde quant à l'interprétation des résultats étant donné que la comparabilité est limitée ou encore un constat de l'impossibilité de la comparabilité des résultats dans le temps.

En somme, bien qu'aucune étude détaillée n'ait été effectuée sur la comparabilité des données de l'EQSP avec d'autres enquêtes, l'utilisateur qui souhaite comparer ses résultats d'analyse à d'autres sources de données devrait tenir compte des éléments soulevés précédemment, de façon à bien expliquer des écarts inattendus ou l'absence d'écarts attendus.

6.2. Production de tableaux et comparaison des prévalences sur le portail de l'Infocentre

Le portail de l'Infocentre propose plusieurs options d'analyse pour les indicateurs du PCS dont la source est l'EQSP 2008. La première partie de cette section 6.2 présente les différents paramètres d'une requête type ainsi que les valeurs que peuvent prendre ces paramètres. La seconde partie de la section (6.2.2) affiche l'arborescence des niveaux de production possibles pour une requête paramétrable : le niveau géographique, les variables de croisement autorisées, les tests effectués, etc. La dernière partie de la section (6.2.3) fournit quelques exemples de requêtes et de sorties pour chacun des niveaux géographiques permis (ensemble du Québec, régional et local). Des recommandations importantes à l'utilisateur y sont également formulées quant aux tests à utiliser pour la comparaison d'indicateurs du PCS et quant à la production de tableaux. Soulignons d'ailleurs que ces recommandations ont toutes été retenues et appliquées sur le portail.

Par ailleurs, soulignons que le début de chacune des sections 6.2.1 et 6.2.2 porte sur les paramètres d'analyse des indicateurs sur la santé ou les habitudes de vie des individus de 15 ans et plus de l'EQSP. Les considérations méthodologiques touchant les paramètres offerts pour les indicateurs de la santé mentale et psychosociale des enfants de 3 à 14 ans sont présentées à la fin de chacune de ces sections.

6.2.1. Paramètres d'une requête paramétrable

Lors d'une requête sur le portail, l'utilisateur qui souhaite analyser les données de l'EQSP doit d'abord définir la valeur des paramètres de cette requête. La possibilité de retenir un paramètre dépend souvent du paramètre choisi précédemment. La liste qui suit présente chacun de ces paramètres, leur rôle et les valeurs qu'ils peuvent prendre.

Choix de l'indicateur ou de la famille d'indicateurs

Bien entendu, la première étape vers l'exécution d'une requête paramétrable est le choix d'un indicateur ou d'une famille d'indicateurs parmi ceux disponibles sur le portail. En cliquant sur le nom de l'indicateur vous avez accès à la requête, et les paramètres suivants doivent être spécifiés.

Paramètre « Niveau géographique »

Ce paramètre permet de définir l'unité territoriale d'analyse pour la production du tableau. Les unités territoriales offertes sont : « Ensemble du Québec », « Régional » et « Local »⁵⁶.

Paramètre « Région sociosanitaire »

Ce paramètre permet de choisir la région sociosanitaire à analyser lorsque le niveau géographique « Régional » est sélectionné. Seize régions sont disponibles.

Paramètre « Région sociosanitaire avec suréchantillon »

Ce paramètre permet de choisir la région sociosanitaire à analyser lorsque le niveau géographique « Local » est sélectionné. Six régions seulement ont fait l'achat d'un échantillon supplémentaire en 2008. Seules ces régions sont disponibles pour ce paramètre.

Paramètre « Sous-population »

Ce paramètre permet de définir une seule ou plusieurs sous-populations d'analyse. Les choix offerts sont : « Sexe », « Âge », « Sexe-âge » et « Totale ». Dans le cas où, par exemple, le paramètre « Sexe » est sélectionné et que l'utilisateur spécifie « par sexe », les résultats seront fournis séparément pour les hommes et pour les femmes, en plus du résultat pour les sexes combinés. Lorsqu'aucune sous-population spécifique n'est souhaitée, il suffit de choisir la valeur « Totale » pour ce paramètre.

Paramètre « Première variable de croisement »

Ce paramètre permet de définir une variable de croisement selon laquelle l'indicateur sera ventilé. Par exemple, si le paramètre « Âge » est sélectionné, alors la répartition de l'indicateur sera produite selon l'âge des individus. Si un test statistique est demandé lors de la requête, le test exécuté permettra d'évaluer si l'indicateur analysé varie selon cette variable de croisement. Plusieurs variables sont disponibles pour ce paramètre, soient celles non sélectionnées comme variable de sous-population, et celles listées dans la fiche de l'indicateur.

Paramètre « Comparaison régionale »

Ce paramètre permet de définir quel type d'analyse régionale l'utilisateur souhaite faire. Les choix possibles sont : « Aucune », « Région comparée avec le reste du Québec » et « Région comparée avec une autre région ». Lors de la sélection de ce dernier, le paramètre « Choisissez la 2^e région » apparaît. Et comme son nom l'indique, il permet de choisir la région de comparaison. Quinze régions sociosanitaires sont alors disponibles. Ce paramètre est affiché seulement lorsque le niveau géographique spécifié est « Régional ».

⁵⁶ L'unité géographique la plus fine, soit le territoire d'un RLS, ne sera disponible qu'aux utilisateurs du portail détenant un profil d'accès 20 ou 30, et ce, pour une période de six mois à partir de la date du déploiement initial des données de l'EQSP 2008 à l'Infocentre.

Paramètre « Comparaison locale »

Ce paramètre permet de définir quel type d'analyse au niveau des RLS l'utilisateur souhaite effectuer. Les choix possibles sont : « Aucune » et « RLS comparés avec le reste de la région ».

Paramètre « Deuxième variable de croisement »

Ce paramètre permet de définir une seconde variable de croisement (pour l'ensemble du Québec seulement) selon laquelle l'indicateur sera ventilé. Le cas échéant, aucun test statistique ne peut être demandé. Les valeurs possibles sont : « Aucune », la région sociosanitaire et chacune des variables de croisement listées dans la fiche de l'indicateur.

Paramètre « Test statistique »

Ce paramètre permet de demander l'exécution de tests statistiques. Tout d'abord, un test global du khi-deux est produit⁵⁷. Ensuite, si le résultat de ce test est significatif, toutes les comparaisons 2 à 2 sont produites⁵⁸. Comme le nom du seul choix actif l'indique « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population », l'autre étant « Aucun », ce paramètre permet de produire séparément pour chacune des sous-populations définies par la variable de sous-population, les tests mentionnés précédemment.

L'utilisateur de l'Infocentre remarquera rapidement que les tests statistiques effectués sur les données de l'EQSP sur le portail de l'Infocentre prennent toujours la même forme. D'ailleurs, le diagramme 6-1 illustre le cheminement que doit suivre un utilisateur afin de bien comprendre les tests effectués lors d'une requête. En effet, un test global du Khi-deux est systématiquement produit lorsqu'une comparaison est demandée par l'utilisateur, afin d'évaluer s'il existe une relation significative entre l'indicateur et la variable de croisement, pour la sous-population étudiée⁵⁹. Dans le cas où le test global montre une relation significative (valeur p du test inférieure au seuil de signification α fixé à 5 %⁶⁰), alors des tests de comparaison de proportions deux à deux sont effectués et un tableau présentant les résultats est affiché à la suite du tableau des estimations.

Comme il a été mentionné à la section 5.1.2, dans le cas d'un indicateur à deux catégories seulement, analysé selon une variable de croisement qui ne possède également que deux catégories, le test global du khi-deux (section 5.1.1) est équivalent au test de la différence de deux proportions (section 5.1.2). Dans une telle situation, il est inutile d'effectuer au préalable un test du khi-deux et il a été décidé par le Comité tripartite MSSS-INSPQ-ISQ que seul le test de la différence de deux proportions serait effectué (comme illustré au diagramme 6-1). Le tableau présentant le résultat de ce test est donc affiché directement à la suite du tableau présentant les estimations. Le test effectué correspond à celui de la méthode 1 décrit à la section 5.1.2.

Dans certaines circonstances, qui sont décrites aux sections 6.2.2, 6.3.1 et 6.3.2, seuls les tests globaux du khi-deux seront produits par la requête. L'utilisateur est alors invité, pour un test du khi-deux significatif, à utiliser les intervalles de confiance de chacune des proportions qui sont présentés dans les tableaux afin d'identifier la source de la différence observée (méthode 2 – section 5.1.2). L'utilisateur doit donc lui-

⁵⁷ Dans le cas d'un indicateur à deux catégories et d'une variable de croisement aussi à deux catégories, seules les comparaisons 2 à 2 sont produites. Voir la section 5.1.2 pour plus de détails.

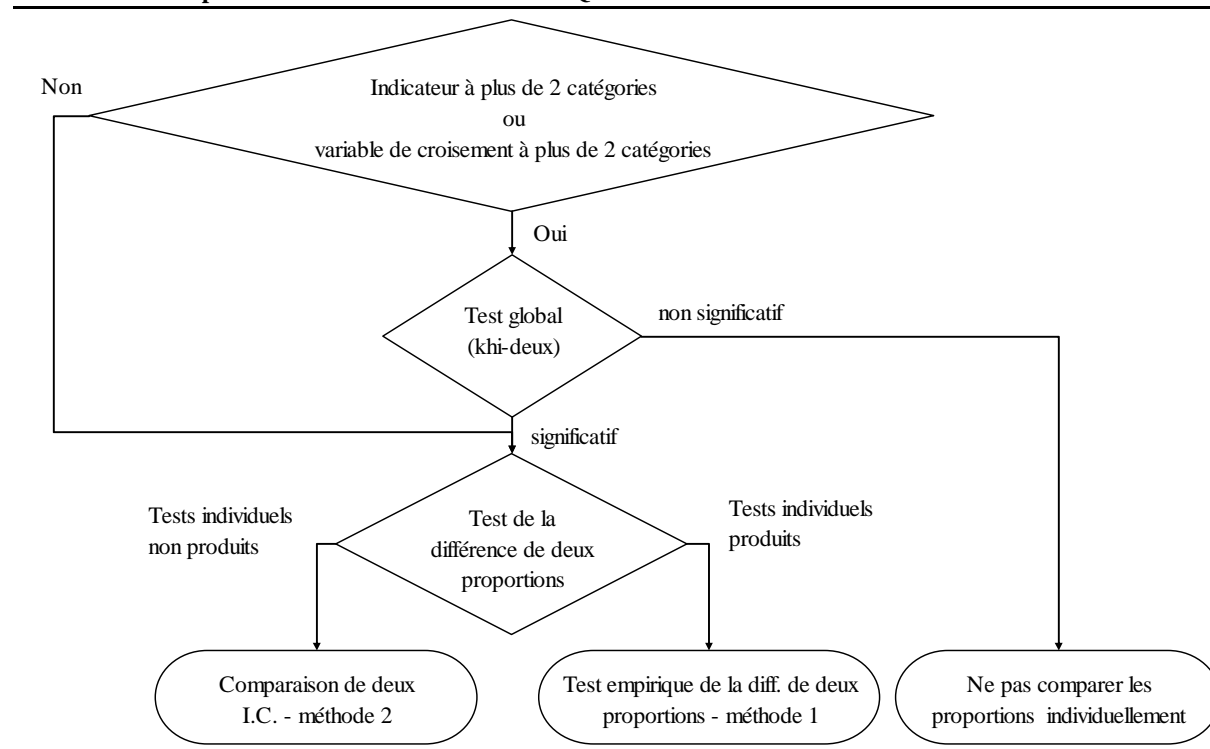
⁵⁸ Il est possible que malgré un test global significatif, les comparaisons 2 à 2 ne soient pas effectuées. Pour plus de détails, consulter la section 6.3.

⁵⁹ Il y aura autant de tests globaux du khi-deux, qu'il y a de sous-populations formées par la requête.

⁶⁰ Pour l'EQSP, une seule valeur pour le seuil de signification est permise. Pour plus de détails, consultez la section 3.3.4.

même comparer les intervalles de confiance fournis dans le tableau afin de déterminer si l'écart observé est significatif.

Diagramme 6-1 Tests statistiques recommandés et produits à l'Infocentre pour la comparaison des prévalences d'un indicateur de l'EQSP selon une variable de croisement



Paramètre « Calcul de la proportion ajustée, standardisation directe »

Ce paramètre permet d'obtenir les proportions ajustées lorsqu'un test statistique est demandé.

Paramètres pour les indicateurs portant sur les enfants de 3 à 14 ans

Pour les indicateurs portant sur les enfants de 3 à 14 ans de l'EQSP, certaines restrictions supplémentaires s'appliquent pour les requêtes paramétrables effectuées sur le portail de l'Infocentre. Elles sont résumées dans le prochain paragraphe. Seuls les paramètres affectés par ces contraintes supplémentaires sont abordés.

Lorsqu'une requête est effectuée pour le niveau géographique « Ensemble du Québec », une seule variable parmi l'âge et le sexe est permise pour définir la sous-population de cette requête. De plus, une seule variable de croisement est permise (contrairement à deux pour les autres indicateurs de ce niveau géographique) parmi, d'une part, les variables non sélectionnées comme sous-population et, d'autre part, les différentes variables de croisement indiquées dans la fiche de l'indicateur analysé. Il est à noter que si la sous-population choisie est le groupe d'âge, alors le choix de la région comme variable de croisement n'est disponible que pour le groupe des 6 à 14 ans. Lorsque la requête est effectuée pour le niveau géographique « Régional », une sous-population peut être définie (6-14 ans) et une seule variable de croisement est permise soit le sexe. Enfin, aucun tableau ne peut être généré au niveau géographique « Local » pour les indicateurs portant sur les enfants de 3 à 14 ans.

La prochaine section présente deux algorithmes de requêtes paramétrables faites sur le portail de l'Infocentre. Le premier concerne la production d'estimations chez les personnes de 15 ans et plus, alors que le second concerne la production d'estimations chez les enfants de 3 à 14 ans. Cette section sert à

connaître en détail les analyses permises lors d'une telle requête. Le lecteur est invité à consulter la section 6.2.3 pour des exemples précis de requêtes et de sorties de tableaux ou l'annexe 15 pour des tableaux récapitulatifs de ces algorithmes.

6.2.2. Algorithmes d'une requête paramétrable sur le portail pour les données de l'EQSP

Estimations portant sur les individus de 15 ans et plus :

1. Si paramètre « Niveau géographique » = « Ensemble du Québec » :
 - i. Les choix du paramètre « Sous-population » sont : « Sexe », « Âge », « Sexe et âge », « Totale » :
 - Si « Sexe » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe
 - Si « Âge » est choisi, un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
 - Si « Sexe et âge » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe et un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
 - ii. Paramètre « Première variable de croisement » permet de choisir :
 - « Aucune »
 - Il n'y a pas d'autres paramètres à choisir
 - Il n'y a aucun test statistique d'offert
 - La variable région
 - Si la variable région est choisie le paramètre « Test statistique » s'affiche avec comme choix « Aucun », « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population ». Si le paramètre sous-population = « Totale » alors le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement ». Ce dernier choix permet d'obtenir les résultats de la comparaison de chacune des RSS avec le reste du Québec et ainsi de comparer sous un angle d'analyse national l'indicateur selon les RSS
 - Il n'y a pas d'autres variables de croisement à choisir
 - La variable de sous-population non sélectionnée entre le sexe et l'âge (les deux si le paramètre sous-population = « Totale » et aucune si la valeur du paramètre est « Sexe et âge ») et les autres variables de croisement mentionnées dans la fiche de l'indicateur
 - Le paramètre « Deuxième variable de croisement » s'affiche (sauf si la sous-population est « Sexe et âge ») et permet de choisir :
 - « Aucune »
 - Afficher un paramètre « Test statistique » avec comme choix « Aucun » et « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population ». Si le paramètre sous-population = « Totale » alors le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement »
 - La variable région
 - Aucun test statistique
 - Les variables de croisement mentionnées dans la fiche de l'indicateur, à l'exception du sexe et de l'âge
 - Aucun test statistique
 - iii. Si « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou « Test de comparaison de l'indicateur selon la variable de croisement » est sélectionné :

- Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont 2 modalités
 - Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
 - Si l'indicateur a 2 modalités : l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour identifier la source de la différence, jusqu'à un maximum de 30 (32 si la variable de croisement est la région) comparaisons 2 à 2, sinon une note est affichée
 - Si l'indicateur a plus de deux modalités, donc s'il est une répartition : on ne fait pas de test pour identifier la source de la différence. Il est alors suggéré de comparer les intervalles de confiance
 - Si deux variables de croisement : aucun test statistique
- iv. Le choix d'obtenir une proportion ajustée n'est offert que si un test statistique est demandé, si l'âge n'est pas sélectionné comme sous-population ni comme variable de croisement, pour les indicateurs qui concernent les 15 ans et plus (à l'exception de cas particulier comme les indicateurs en santé au travail)
2. Si le paramètre « Niveau géographique » = « Régional » :
- i. Choisir la région à l'aide du paramètre « Région sociosanitaire »
 - ii. Les choix du paramètre « Sous-population » sont : « Sexe », « Âge », « Sexe et âge », « Totale »
 - Si « Sexe » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe
 - Si « Âge » est choisi, un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
 - Si « Sexe et âge » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe et un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
 - iii. Paramètre « Comparaison régionale » avec comme choix de réponse « Aucune », « Région comparée avec le reste du Québec », « Région comparée avec une autre région »
 - Si « Aucune » et que la sous-population est « Sexe et âge », on ne peut choisir de variables de croisement
 - Si « Aucune » et que la sous-population est « Totale » ou « Sexe » ou « Âge » :
 - Le paramètre « Variable de croisement » s'affiche (une seule variable de croisement est permise) avec comme choix :
 - « Aucune »
 - Aucun test statistique
 - La variable de sous-population non sélectionnée entre le sexe et l'âge (les deux si le paramètre sous-population = « Totale ») et les autres variables de croisement mentionnées dans la fiche de l'indicateur
 - Afficher un paramètre « Test statistique » avec comme choix « Aucun » et « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population ». Si le paramètre sous-population = « Totale » alors le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement »
 - Si « Région comparée avec le reste du Québec » :
 - On ne peut choisir d'autres variables de croisement, la région en 2 catégories (région sélectionnée et reste du Québec) étant considérée comme une variable de croisement
 - Le test statistique qui compare l'indicateur entre la région et le reste du Québec, stratifié selon la sous-population, est effectué
 - Si « Région comparée avec une autre région » :

- On ne peut choisir d'autres variables de croisement, la région en 2 catégories (région sélectionnée et la région de comparaison) étant considérée comme une variable de croisement
 - Le paramètre « Région de comparaison » permet de sélectionner la région de comparaison
 - Le test statistique qui compare l'indicateur entre les 2 régions, stratifié selon la sous-population, est effectué
- iv. Si « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou « Région comparée avec le reste du Québec » ou « Région comparée avec une autre région » est sélectionné :
- Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont 2 modalités
 - Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
 - Si l'indicateur a 2 modalités : l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour identifier la source de la différence, jusqu'à un maximum de 30 comparaisons 2 à 2, sinon une note est affichée
 - Si l'indicateur a plus de deux modalités, donc s'il est une répartition : on ne fait pas de test pour identifier la source de la différence. Il est alors suggéré de comparer les intervalles de confiance
- v. Le choix d'obtenir une proportion ajustée n'est offert que si un test statistique est demandé, si l'âge n'est pas sélectionné comme sous-population ni comme variable de croisement, pour les indicateurs qui concernent les 15 ans et plus (à l'exception de cas particulier comme les indicateurs en santé au travail)
3. Paramètre « Niveau géographique » = « Local » :
- i. Choisir la région pour laquelle on veut les RLS à l'aide du paramètre « Région sociosanitaire avec suréchantillon »
- ii. Les choix du paramètre « Sous-population » sont : « Sexe », « Âge », « Sexe et âge », « Totale »
- Si « Sexe » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe
 - Si « Âge » est choisi, un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
 - Si « Sexe et âge » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe et un paramètre « Groupe d'âge » permet de choisir une des catégorisations d'âge mentionnées dans la fiche de l'indicateur
- iii. Paramètre « Comparaison locale » avec comme choix de réponse « Aucune », « RLS comparés avec le reste de la région »
- Si « Aucune » et que la sous-population est « Sexe et âge », on ne peut choisir de variables de croisement
 - Si « Aucune » et que la sous-population est « Totale », « Sexe » ou « Âge » :
 - Le paramètre « Variable de croisement » s'affiche (une seule variable de croisement est permise) avec comme choix :
 - « Aucune »
 - Aucun test statistique
 - La variable de sous-population non sélectionnée entre le sexe et l'âge (les deux si le paramètre sous-population = « Totale ») et les autres variables de croisement mentionnées dans la fiche de l'indicateur
 - Afficher un paramètre « Test statistique » avec comme choix « Aucun » et « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population ». Si le paramètre sous-population = « Totale » alors le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement »
 - Si « RLS comparés avec le reste de la région »

- On ne peut choisir d'autres variables de croisement, le RLS étant considéré comme une variable de croisement
- Le test statistique correspondant est effectué
- iv. Si « RLS comparés avec le reste de la région » ou « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou « Test de comparaison de l'indicateur selon la variable de croisement » est sélectionné :
 - Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont 2 modalités
 - Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
 - Si l'indicateur a 2 modalités : l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour identifier la source de la différence, jusqu'à un maximum de 30 comparaisons 2 à 2, sinon une note est affichée
 - Si l'indicateur a plus de deux modalités, donc s'il est une répartition : on ne fait pas de test pour identifier la source de la différence. Il sera suggéré dans le guide de comparer les intervalles de confiance
 - Il n'est pas possible de comparer deux RLS entre eux
- v. Le choix d'obtenir une proportion ajustée n'est offert que si un test statistique est demandé, si l'âge n'est pas sélectionné comme sous-population ni comme variable de croisement, pour les indicateurs qui concernent les 15 ans et plus (à l'exception de cas particulier comme les indicateurs en santé au travail)

Estimations portant sur les enfants de 3 à 14 ans :

1. Si paramètre « Niveau géographique » = « Ensemble du Québec » :
 - i. Les choix du paramètre « Sous-population » sont : « Sexe de l'enfant », « Âge de l'enfant », « Totale » :
 - Si « Sexe de l'enfant » est choisi, un paramètre « Sexe » permet de choisir : masculin, féminin, par sexe
 - Si « Âge de l'enfant » est choisi, un paramètre « Groupe d'âge » permet de choisir une catégorisation d'âge de l'enfant :
 - 3-5 ans, 6-14 ans
 - 3-5 ans, 6-8 ans, 9-11 ans, 12-14 ans
 - 6-14 ans
 - ii. Paramètre « Variable de croisement » permet de choisir :
 - « Aucune » :
 - Il n'y a pas d'autres paramètres à choisir
 - Il n'y a aucun test statistique d'offert
 - Si le paramètre « Sous-population » = « Totale » ou « Sexe » ou si le paramètre « Sous-population » = « Âge de l'enfant » et que le paramètre « Groupe d'âge » = « 6-14 ans » alors la variable « région » est disponible comme variable de croisement :
 - Le paramètre « Test statistique » s'affiche avec comme choix « Aucun », « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou « Test de comparaison de l'indicateur selon la variable de croisement » tout dépendamment si le paramètre de sous-population est différent de la modalité « Totale ». Ce dernier choix, pour le paramètre « Test statistique », permet d'obtenir les résultats de la comparaison de chacune des RSS vs le reste du Québec et ainsi de comparer sous un angle d'analyse national l'indicateur selon les RSS
 - La variable de sous-population non sélectionnée entre le sexe de l'enfant et l'âge de l'enfant (les deux si sous-population = « Totale »), ainsi que les autres variables de croisement mentionnées dans la fiche de l'indicateur :
 - Le paramètre « Test statistique » affiche comme choix « Aucun » et « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-

- population ». Si le paramètre sous-population = « Totale » alors le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement »
- Si « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou « Test de comparaison de l'indicateur selon la variable de croisement » est sélectionné pour le paramètre « Test statistique » :
 - Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont 2 modalités
 - Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
 - Si l'indicateur a 2 modalités : l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour identifier la source de la différence, jusqu'à un maximum de 30 (32 si la variable de croisement est la région) comparaisons 2 à 2, sinon une note est affichée
 - Si l'indicateur a plus de deux modalités, donc s'il est une répartition : on ne fait pas de test pour identifier la source de la différence. Il est alors suggéré de comparer les intervalles de confiance
2. Si le paramètre « Niveau géographique » = « Régional » :
- i. Choisir la région à l'aide du paramètre « Région sociosanitaire »
 - ii. Les choix du paramètre « Sous-population » sont : « Âge de l'enfant », « Totale » :
 - Si « Âge de l'enfant » est choisi, le seul groupe d'âge offert est : 6-14 ans
 - iii. Paramètre « Comparaison régionale » avec comme choix de réponse « Aucune », « Région comparée avec le reste du Québec », « Région comparée avec une autre région » :
 - Si « Aucune », le paramètre « Variable de croisement » s'affiche avec comme choix :
 - « Aucune »
 - Aucun test statistique
 - « Sexe de l'enfant »
 - Afficher un paramètre « Test statistique » avec comme choix « Aucun » et « Test de comparaison de l'indicateur selon la variable de croisement »
 - Si « Région comparée avec le reste du Québec » :
 - On ne peut choisir d'autres variables de croisement, la région en 2 catégories (région sélectionnée et reste du Québec) étant considérée comme une variable de croisement
 - Le test statistique qui compare l'indicateur entre la région et le reste du Québec est effectué
 - Si « Région comparée avec une autre région » :
 - On ne peut choisir d'autres variables de croisement, la région en 2 catégories (région sélectionnée et la région de comparaison) étant considérée comme une variable de croisement
 - Le paramètre « Région de comparaison » permet de sélectionner la région de comparaison
 - Le test statistique qui compare l'indicateur entre les 2 régions est effectué
 - iv. Si « Test de comparaison de l'indicateur selon la variable de croisement » ou « Région comparée avec le reste du Québec » ou « Région comparée avec une autre région » est sélectionné :
 - Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont 2 modalités
 - Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
 - Si l'indicateur a 2 modalités : l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour identifier la source de la différence, jusqu'à un maximum de 30 comparaisons 2 à 2, sinon une note est affichée
 - Si l'indicateur a plus de deux modalités, donc s'il est une répartition : on ne fait pas de test pour identifier la source de la différence. Il est alors suggéré de comparer les intervalles de confiance

6.2.3. Exemples de résultats de requêtes paramétrables et recommandations

Pour les indicateurs du PCS dont la source est l'EQSP, l'Infocentre offre trois niveaux géographiques d'analyse : une production pour l'ensemble du Québec, une production à l'échelle régionale et une production à l'échelle locale⁶¹. C'est le paramètre « niveau géographique » de la fenêtre de saisie qui permet de faire ce choix.

Pour chacun de ces trois niveaux géographiques, la façon d'analyser un indicateur est quasiment identique. Il faut choisir la « sous-population » souhaitée, la « première variable de croisement », la « deuxième variable de croisement » (dans certaines situations seulement), et la sortie ou non de « tests statistiques ». Comme nous l'avons vu à la section précédente, qui présente la structure complète des analyses permises et des tests effectués, pour chacun des niveaux géographiques, seuls le nombre permis de variables de croisement et les tests permis peuvent différer selon le niveau géographique.

Les prochains paragraphes abordent, à l'aide d'exemples, chacun de ces éléments. Le premier groupe d'exemples permet de produire des prévalences pour l'ensemble des Québécois visés par l'enquête. Le second groupe permet de cibler les analyses à une région sociosanitaire précise, permettant de se comparer soit au reste du Québec, soit à une autre région. Le troisième groupe vise à permettre des analyses encore plus fines, au niveau des RLS, permettant de comparer les résultats obtenus à celui du reste de la région.

Premier groupe d'exemples : Ensemble du Québec

Comme premier exemple, analysons la perception du poids de l'ensemble des Québécois selon leur âge et leur sexe. Deux options s'offrent à nous : la première est de produire les prévalences selon le sexe comme variable de sous-population, et l'âge comme variable de croisement (tableau 6-1). La seconde est de choisir la combinaison sexe-âge comme variable de sous-population (tableau 6-2). La différence entre les deux options réside dans les analyses et les tests qu'on peut en tirer. Il incombe donc à l'utilisateur de produire le tableau qui correspond à ses besoins.

Pour des tableaux portant sur l'ensemble du Québec, il est aussi possible de choisir la valeur « région » comme première variable de croisement. Ce choix permet de faire une analyse des régions sous un angle national. Sous cette forme, et lorsque le test global est significatif, les prévalences régionales de l'indicateur sont comparées à celles du reste du Québec, et ce, pour chacune des RSS. Le tableau 6-3 offre le visuel de la sortie de l'Infocentre pour une telle requête sans les tests.

Pour ce niveau géographique, et seulement pour ce niveau⁶², il est possible, pour certaines situations (consulter l'algorithme à la section 6.2.2), de choisir une deuxième variable de croisement. Le cas échéant, les estimations sont produites pour le croisement des deux variables choisies, séparément pour chacune des sous-populations définies, mais aucun test statistique n'est effectué. Voir le tableau 6-4 pour un exemple de sortie de l'Infocentre.

⁶¹ Cette production ne sera disponible que pour les utilisateurs détenant un profil d'accès 20 ou 30 pour une période de six mois à partir de la date du déploiement initial des données de l'EQSP 2008 à l'Infocentre.

⁶² En général, pour les autres niveaux géographiques, la taille d'échantillon n'est pas suffisante pour permettre l'utilisation d'une deuxième variable de croisement.

Tableau 6-1 Répartition de la population selon la perception du poids pour l'ensemble du Québec, EQSP 2008 (extrait du tableau complet)

Sous-population		Indicateur		Statistiques				
Sexe	Variable de croisement	Perception du poids	Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	15-24 ans	Excès de poids	73 400	14,7	9,1	1,34	(12,0 - 17,5)	2,1
		Trop maigre	37 300	* 7,5	17,1	1,28	(5,0 - 9,9)	2,1
		À peu près normal(e)	388 200	77,8	2,3	1,76	(74,5 - 81,3)	2,1
	25-44 ans	Excès de poids	379 900	35,4	3,4	1,20	(33,2 - 37,6)	0,7
		Trop maigre	39 900	3,7	14,7	0,54	(2,7 - 4,9)	0,7
		À peu près normal(e)	654 700	60,9	2,0	1,23	(58,6 - 63,3)	0,7
	45-64 ans	Excès de poids	519 700	47,4	2,1	0,97	(45,4 - 49,4)	1,3
		Trop maigre	23 300	2,1	12,1	0,26	(1,6 - 2,7)	1,3
		À peu près normal(e)	554 000	50,5	1,9	0,97	(48,4 - 52,5)	1,3
	65 ans et +	Excès de poids	169 500	37,2	4,0	1,47	(34,2 - 40,1)	* 10,1
		Trop maigre	10 100	2,2	14,8	0,33	(1,6 - 2,9)	* 10,1
		À peu près normal(e)	276 300	60,6	2,4	1,46	(57,7 - 63,6)	* 10,1
Total	Excès de poids	1 143 000	36,6	1,7	0,63	(35,3 - 37,7)	2,5	
	Trop maigre	111 000	3,6	8,4	0,30	(3,0 - 4,2)	2,5	
	À peu près normal(e)	1 872 200	59,9	1,1	0,66	(58,7 - 61,2)	2,5	
Féminin	15-24 ans	Excès de poids	86 300	18,5	8,2	1,52	(15,7 - 21,7)	0,7
		Trop maigre	10 400	* 2,2	23,8	0,53	(1,2 - 3,4)	0,7
		À peu près normal(e)	369 400	79,3	2,0	1,55	(75,9 - 82,2)	0,7

Source : Infocentre de santé publique, octobre 2010.

Tableau 6-2 Répartition de la population selon la perception du poids pour l'ensemble du Québec, EQSP 2008 (extrait du tableau complet)

Sous-population		Indicateur		Statistiques				
Sexe	Variable de croisement	Perception du poids	Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	15-24 ans	Excès de poids	73 400	14,7	9,1	1,34	(12,0 - 17,5)	2,1
		Trop maigre	37 300	* 7,5	17,1	1,28	(5,0 - 9,9)	2,1
		À peu près normal(e)	388 200	77,8	2,3	1,76	(74,5 - 81,3)	2,1
	25-44 ans	Excès de poids	379 900	35,4	3,4	1,20	(33,2 - 37,6)	0,7
		Trop maigre	39 900	3,7	14,7	0,54	(2,7 - 4,9)	0,7
		À peu près normal(e)	654 700	60,9	2,0	1,23	(58,6 - 63,3)	0,7
	45-64 ans	Excès de poids	519 700	47,4	2,1	0,97	(45,4 - 49,4)	1,3
		Trop maigre	23 300	2,1	12,1	0,26	(1,6 - 2,7)	1,3
		À peu près normal(e)	554 000	50,5	1,9	0,97	(48,4 - 52,5)	1,3
	65 ans et +	Excès de poids	169 500	37,2	4,0	1,47	(34,2 - 40,1)	* 10,1
		Trop maigre	10 100	2,2	14,8	0,33	(1,6 - 2,9)	* 10,1
		À peu près normal(e)	276 300	60,6	2,4	1,46	(57,7 - 63,6)	* 10,1
Total	Excès de poids	1 143 000	36,6	1,7	0,63	(35,3 - 37,7)	2,5	
	Trop maigre	111 000	3,6	8,4	0,30	(3,0 - 4,2)	2,5	
	À peu près normal(e)	1 872 200	59,9	1,1	0,66	(58,7 - 61,2)	2,5	
Féminin	15-24 ans	Excès de poids	86 300	18,5	8,2	1,52	(15,7 - 21,7)	0,7
		Trop maigre	10 400	* 2,2	23,8	0,53	(1,2 - 3,4)	0,7
		À peu près normal(e)	369 400	79,3	2,0	1,55	(75,9 - 82,2)	0,7

Source : Infocentre de santé publique, octobre 2010.

Tableau 6-3 Répartition de la population selon la perception du poids pour l'ensemble du Québec, EQSP 2008 (extrait du tableau complet)

Variable de croisement	Indicateur	Statistiques					
		Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
01 Bas-Saint-Laurent	Excès de poids	58 900	36,1	3,3	1,19	(33,8 - 38,4)	3,0
	Trop maigre	3 900	* 2,4	15,3	0,36	(1,7 - 3,1)	3,0
	À peu près normal(e)	100 300	61,5	2,0	1,22	(59,2 - 63,9)	3,0
02 Saguenay - Lac-Saint-Jean	Excès de poids	81 700	36,5	2,7	0,97	(34,8 - 38,4)	2,0
	Trop maigre	4 600	2,0	12,4	0,25	(1,6 - 2,6)	2,0
	À peu près normal(e)	137 800	61,5	1,6	0,99	(59,6 - 63,2)	2,0
03 Capitale-Nationale	Excès de poids	214 100	38,1	3,0	1,15	(35,8 - 40,5)	1,8
	Trop maigre	12 600	* 2,2	16,2	0,36	(1,5 - 3,0)	1,8
	À peu près normal(e)	335 000	59,6	2,0	1,17	(57,3 - 61,9)	1,8
04 Mauricie et Centre-du-Québec	Excès de poids	154 600	39,0	2,7	1,04	(37,0 - 40,9)	1,5
	Trop maigre	9 400	2,4	14,5	0,34	(1,7 - 3,1)	1,5
	À peu près normal(e)	232 800	58,7	1,8	1,04	(56,6 - 60,6)	1,5
05 Estrie	Excès de poids	98 100	39,9	2,9	1,16	(37,5 - 42,3)	2,4
	Trop maigre	7 400	3,0	14,8	0,44	(2,1 - 3,9)	2,4
	À peu près normal(e)	140 300	57,1	2,1	1,21	(54,7 - 59,5)	2,4
06 Montréal	Excès de poids	582 500	37,8	3,3	1,26	(35,3 - 40,1)	2,3
	Trop maigre	63 200	4,1	13,8	0,57	(3,0 - 5,2)	2,3
	À peu près normal(e)	896 100	58,1	2,2	1,30	(55,6 - 60,6)	2,3
07 Outaouais	Excès de poids	112 700	39,8	2,9	1,15	(37,6 - 42,2)	2,3
	Trop maigre	7 400	2,6	14,5	0,38	(1,9 - 3,5)	2,3
	À peu près normal(e)	163 300	57,6	2,0	1,18	(55,1 - 59,8)	2,3
08 Abitibi-Témiscamingue	Excès de poids	47 100	40,7	2,9	1,18	(38,5 - 43,0)	2,5
	Trop maigre	3 100	2,7	15,0	0,40	(1,9 - 3,5)	2,5
	À peu près normal(e)	65 500	56,6	2,1	1,20	(54,3 - 58,9)	2,5

Source : Infocentre de santé publique, octobre 2010.

Tableau 6-4 Répartition de la population selon la perception du poids pour l'ensemble du Québec, EQSP 2008 (extrait du tableau complet)

Variables de croisement		Indicateur	Statistiques					
Sexe	Statut pondéral	Perception du poids	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	Poids insuffisant	Excès de poids	*** 100	** 0,2	103,0	0,17	(0,0 - 1,0)	0,2
		Trop maigre	18 300	60,5	13,5	8,18	(42,6 - 74,1)	0,2
		À peu près normal(e)	11 900	* 39,3	20,7	8,16	(25,6 - 57,1)	0,2
	Poids normal	Excès de poids	55 600	4,7	9,4	0,44	(3,9 - 5,6)	0,3
		Trop maigre	79 800	6,8	9,1	0,61	(5,6 - 8,0)	0,3
		À peu près normal(e)	1 045 300	88,5	0,8	0,75	(87,0 - 89,9)	0,3
	Embonpoint	Excès de poids	617 700	46,8	2,1	0,99	(44,8 - 48,8)	0,2
		Trop maigre	6 000	** 0,5	36,1	0,16	(0,2 - 0,8)	0,2
		À peu près normal(e)	696 400	52,8	1,9	0,99	(50,7 - 54,8)	0,2
	Obésité	Excès de poids	436 800	87,5	1,3	1,12	(85,2 - 89,6)	0,1
		Trop maigre	*** 1 700	** 0,4	98,3	0,34	(0,0 - 1,2)	0,1
		À peu près normal(e)	60 400	12,1	9,0	1,09	(10,1 - 14,3)	0,1
Féminin	Poids insuffisant	Excès de poids	300	** 0,3	78,7	0,20	(0,0 - 0,8)	0,0
		Trop maigre	30 200	26,2	11,1	2,91	(20,5 - 32,4)	0,0
		À peu près normal(e)	84 900	73,5	4,0	2,92	(67,3 - 79,1)	0,0
	Poids normal	Excès de poids	246 800	14,8	4,1	0,60	(13,7 - 16,1)	0,1
		Trop maigre	34 900	2,1	12,7	0,27	(1,6 - 2,7)	0,1
		À peu près normal(e)	1 380 800	83,1	0,8	0,63	(81,8 - 84,3)	0,1
	Embonpoint	Excès de poids	546 600	69,6	1,5	1,05	(67,5 - 71,6)	0,2
		Trop maigre	*** 200	** 0,0	57,7	0,02	(0,0 - 0,1)	0,2
		À peu près normal(e)	238 600	30,4	3,5	1,05	(28,3 - 32,5)	0,2
	Obésité	Excès de poids	417 900	94,2	0,7	0,63	(92,9 - 95,5)	0,2
		Trop maigre	*** 0	** 0,0	102,9	0,01	(0,0 - 0,0)	0,2
		À peu près normal(e)	25 500	5,8	11,0	0,63	(4,5 - 7,1)	0,2

Source : Infocentre de santé publique, octobre 2010.

Deuxième groupe d'exemples : Régional (RSS)

Pour les analyses à l'échelle régionale, le paramètre « Comparaison régionale » permet de définir le type d'analyse souhaitée. Les choix sont : « Aucune », « Région comparée au reste du Québec » et « Région comparée à une autre région ». Selon la situation, la sélection d'une seule variable de croisement est permise. Consulter l'algorithme à la section 6.2.2 pour identifier ces situations. Le tableau 6-5 rapporte la présence d'idées suicidaires au cours de la vie selon l'âge pour la région de l'Abitibi-Témiscamingue. Dans un deuxième temps, l'utilisateur peut souhaiter vérifier si cette proportion de personnes rapportant des idées suicidaires est différente du reste du Québec (tableau 6-6) ou d'une autre région (tableau 6-7).

Tableau 6-5 Proportion de la population qui a songé sérieusement au suicide au cours de sa vie, EQSP 2008

Sous-population	Territoire		Statistiques					
	Groupe d'âge	Région	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
15-24 ans	08 Abitibi-Témiscamingue		1 800	* 10,0	21,5	2,16	(5,8 - 14,7)	1,6
	Ensemble du Québec		79 900	8,5	9,3	0,80	(7,0 - 10,1)	1,2
25-44 ans	08 Abitibi-Témiscamingue		5 400	15,8	10,2	1,60	(12,5 - 19,1)	1,5
	Ensemble du Québec		219 100	11,0	4,8	0,53	(10,0 - 12,1)	1,1
45-64 ans	08 Abitibi-Témiscamingue		5 500	13,5	10,2	1,38	(10,9 - 16,2)	2,1
	Ensemble du Québec		261 100	12,3	3,7	0,45	(11,4 - 13,3)	1,5
65 ans et +	08 Abitibi-Témiscamingue		800	* 4,9	21,6	1,05	(2,8 - 7,2)	9,9
	Ensemble du Québec		51 300	5,1	7,6	0,39	(4,4 - 5,9)	* 8,7
Total	08 Abitibi-Témiscamingue		13 700	12,4	6,7	0,82	(10,8 - 14,0)	3,1
	Ensemble du Québec		615 200	10,2	2,6	0,27	(9,7 - 10,7)	2,5

Source : Infocentre de santé publique, octobre 2010.

Tableau 6-6 Proportion de la population qui a songé sérieusement au suicide au cours de sa vie, EQSP 2008

Territoire		Statistiques					
Région	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	13 700	12,4	6,7	0,82	(10,8 - 14,0)	3,1	sig
Reste du Québec	601 400	10,1	2,7	0,27	(9,6 - 10,7)	2,5	

Source : Infocentre de santé publique, octobre 2010.

Tableau 6-7 Proportion de la population qui a songé sérieusement au suicide au cours de sa vie, EQSP 2008

Territoire		Statistiques					
Région	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	13 700	12,4	6,7	0,82	(10,8 - 14,0)	3,1	sig
03 Capitale-Nationale	51 300	9,4	7,1	0,67	(8,1 - 10,8)	2,0	

Source : Infocentre de santé publique, octobre 2010.

Troisième groupe d'exemples : Local (RLS)

De façon similaire à l'échelle régionale, un paramètre « Comparaison locale » doit être défini pour les analyses à l'échelle locale. Les choix sont cependant plus limités. En effet, seuls les choix « Aucune » et « RLS comparé au reste de la région » sont possibles. Encore une fois, selon la situation, la sélection d'au plus une variable de croisement peut être permise. Se référer à la section 6.2.2 pour identifier ces situations à l'aide des algorithmes. Le tableau 6-8 présente la sortie d'une requête à l'échelle locale, pour laquelle la comparaison de chacun des RLS avec le reste de la région est demandée. L'utilisateur de l'Infocentre pourra remarquer que malgré un résultat significatif du test global, les comparaisons deux à deux ne sont pas affichées, car il s'agit d'un indicateur en répartition (voir la section 6.3.1 et la note du tableau 6-8).

Tableau 6-8 Répartition de la population ayant au moins une dent naturelle au maxillaire supérieur ou inférieur selon la fréquence du brossage des dents, région 02 Saguenay - Lac-Saint-Jean, EQSP 2008 (extrait du tableau complet)

Territoire RLS (CSSS)	Indicateur Fréquence du brossage des dents	Statistiques					
		Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
0201 - RLS de La Baie (CSSS Cléophas-Claveau)	Au moins 2 fois/jour	13 100	81,8	2,2	1,84	(78,2 - 85,5)	2,1
	1 fois/jour	2 800	16,3	11,1	1,81	(12,7 - 19,9)	2,1
	Moins d'une fois/jour	300	** 1,9	31,7	0,59	(0,8 - 3,2)	2,1
0202 - RLS de Chicoutimi (CSSS de Chicoutimi)	Au moins 2 fois/jour	48 800	88,8	1,8	1,45	(85,7 - 91,7)	2,4
	1 fois/jour	5 700	10,4	13,7	1,42	(7,6 - 13,4)	2,4
	Moins d'une fois/jour	400	** 0,8	46,9	0,37	(0,1 - 1,6)	2,4
0203 - RLS de Jonquière (CSSS de Jonquière)	Au moins 2 fois/jour	38 700	85,0	2,0	1,67	(81,4 - 88,2)	1,1
	1 fois/jour	6 000	13,2	11,8	1,56	(10,0 - 16,5)	1,1
	Moins d'une fois/jour	900	** 1,9	40,8	0,77	(0,4 - 3,7)	1,1
0204 - RLS du Domaine-du-Roy (CSSS du Domaine-du-Roy)	Au moins 2 fois/jour	17 200	83,4	2,2	1,84	(79,5 - 87,1)	1,3
	1 fois/jour	3 200	15,7	11,5	1,81	(12,2 - 19,6)	1,3
	Moins d'une fois/jour	200	** 0,9	45,7	0,39	(0,1 - 1,9)	1,3
0205 - RLS de Maria-Chapelaine incluant Saint-Ludger-de-Milot (CSSS de Maria-Chapelaine)	Au moins 2 fois/jour	14 500	82,0	2,4	1,93	(78,0 - 85,6)	1,9
	1 fois/jour	2 800	14,9	11,7	1,74	(11,7 - 18,5)	1,9
	Moins d'une fois/jour	500	** 3,1	27,8	0,86	(1,4 - 5,1)	1,9
0206 - RLS de Lac-Saint-Jean-Est excluant Saint-Ludger-de-Milot (CSSS de Lac-Saint-Jean-Est)	Au moins 2 fois/jour	29 700	87,4	1,8	1,57	(83,9 - 90,1)	3,5
	1 fois/jour	3 800	11,0	14,0	1,55	(8,2 - 14,4)	3,5
	Moins d'une fois/jour	500	** 1,6	29,8	0,48	(0,6 - 2,7)	3,5
Total de la région 02 Saguenay - Lac-Saint-Jean	Au moins 2 fois/jour	161 900	85,8	0,8	0,72	(84,3 - 87,2)	2,1
	1 fois/jour	24 000	12,7	5,4	0,68	(11,3 - 14,1)	2,1
	Moins d'une fois/jour	2 900	* 1,5	16,1	0,24	(1,0 - 2,1)	2,1

Test global d'association de la répartition de la fréquence du brossage des dents de la population ayant au moins une dent naturelle au maxillaire supérieur ou inférieur selon les RLS Région 02 Saguenay - Lac-Saint-Jean

Statistique
Valeur p ^a
0,0409

Notes

Dans le cas d'un indicateur en répartition, seuls les tests globaux du khi-deux sont effectués, soit un pour chacune des sous-populations demandées. Pour les résultats globaux significatifs au seuil de 5 %, il est suggéré d'effectuer un test t de la différence ou de comparer les intervalles de confiance deux à deux afin d'identifier la source de la différence.

Source : Infocentre de santé publique, octobre 2010.

6.3. Particularités des analyses produites à l'Infocentre

Les prochaines sections décrivent certaines particularités des indicateurs qui ont un impact significatif sur l'analyse des données sur le portail de l'Infocentre. Chaque sous-section présente un cas précis d'indicateur où des décisions qui ont dû être prises par le Comité tripartite, ainsi que l'impact ou les mises en garde pour les analyses.

Ces sections seront particulièrement utiles pour l'analyse des données de l'EQSP sur le portail de l'Infocentre.

6.3.1. Indicateur en répartition

Lorsque l'indicateur analysé est une répartition, c'est-à-dire qu'il n'est pas dichotomique, certains résultats de tests ne sont pas produits lors de l'exécution de la requête. En effet, malgré qu'un test global soit significatif, aucun test de comparaison des proportions (comparaisons 2 à 2) n'est effectué. Le nombre trop élevé de tests à effectuer et le temps de compilation important qu'une telle requête représente justifient cette décision.

Dans une telle situation, une note au bas du tableau sera affichée mentionnant que les comparaisons 2 à 2 ne sont pas effectuées. Un exemple est présenté au tableau 6-9. L'utilisateur est alors invité à réduire l'ampleur de sa demande ou, pour un test du khi-deux significatif, à utiliser les intervalles de confiance de chacune des proportions, qui sont présentées dans les tableaux, afin d'identifier la source de la différence observée (méthode 2). Pour plus de détails sur la méthode 2, consulter la section 5.1.2.

Soulignons qu'il n'est pas permis de faire une requête comparant l'ensemble des régions lorsque l'indicateur analysé est une répartition. Par contre, cette requête est permise dans le cas où l'indicateur est une proportion (indicateur dichotomique).

Tableau 6-9 Répartition de la population active sexuellement au cours des 12 derniers mois selon le nombre de partenaires pour l'ensemble du Québec, EQSP 2008

Variable de croisement	Indicateur	Statistiques					
		Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	1	2 246 400	86,6	0,6	0,50	(85,5 - 87,5)	3,7
	2 à 4	273 400	10,5	4,4	0,47	(9,7 - 11,5)	3,7
	5 et plus	74 800	2,9	9,3	0,27	(2,4 - 3,5)	3,7
Féminin	1	2 154 000	91,8	0,5	0,42	(90,9 - 92,5)	3,3
	2 à 4	171 000	7,3	5,4	0,40	(6,5 - 8,1)	3,3
	5 et plus	22 200	* 0,9	15,7	0,15	(0,7 - 1,3)	3,3
Total	1	4 400 100	89,0	0,4	0,33	(88,4 - 89,7)	3,5
	2 à 4	444 200	9,0	3,5	0,31	(8,4 - 9,6)	3,5
	5 et plus	96 900	2,0	8,2	0,16	(1,7 - 2,3)	3,5

Test global d'association de la répartition de la population active sexuellement au cours des 12 derniers mois, selon le nombre de partenaires, selon le sexe

Statistique
Valeur p ¹
<0,0001

Notes

Dans le cas d'un indicateur en répartition, seuls les tests globaux du khi-deux sont effectués, soit un pour chacune des sous-populations demandées. Pour les résultats globaux significatifs au seuil de 5 %, il est suggéré d'effectuer un test t de la différence ou de comparer les intervalles de confiance deux à deux afin d'identifier la source de la différence.

Source : Infocentre de santé publique, octobre 2010.

6.3.2. Nombre trop important de tests demandés par une requête

Lorsque le nombre de comparaisons 2 à 2 est supérieur à 30⁶³, celles-ci ne seront pas produites lors de l'exécution de la requête et seuls les tests globaux du khi-deux le seront, un pour chacune des sous-populations définies par le paramètre « Sous-population », et ce, malgré un test du khi-deux significatif.

Lorsqu'une telle situation se présentera, une note au bas du tableau sera affichée mentionnant que le nombre de comparaisons 2 à 2 est trop élevé et que celles-ci ne seront pas effectuées. Un exemple de cette note est présenté au tableau 6-10. L'utilisateur est alors invité à réduire l'ampleur de sa demande ou, pour les tests du khi-deux significatifs, à utiliser les intervalles de confiance de chacune des proportions, qui sont présentés dans les tableaux, afin d'identifier la source de la différence observée (méthode 2). Pour plus de détails sur la méthode 2, consulter la section 5.1.2.

⁶³ Pour les comparaisons selon la région sociosanitaire, le nombre maximal de comparaisons permises est plutôt de 32. Cette valeur permet d'offrir les comparaisons d'un tel croisement en utilisant le sexe comme variable de sous-population.

Tableau 6-10 Proportion de la population se brossant les dents au moins deux fois par jour, parmi ceux ayant au moins une dent naturelle au maxillaire supérieur ou inférieur pour l'ensemble du Québec, EQSP 2008 (extrait du tableau complet)

Sous-population		Variable de croisement	Statistiques					
Sexe	Groupe d'âge	Niveau de revenu	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	15 à 17 ans	Quintile 1 (faible)	19 300	74,3	9,2	6,88	(59,3 - 87,1)	3,2
		Quintile 2	33 500	82,3	5,5	4,54	(72,1 - 90,6)	3,6
		Quintile 3	21 100	66,2	12,1	8,04	(50,1 - 82,5)	5,4
		Quintile 4	35 100	86,7	4,9	4,26	(77,4 - 94,1)	0,3
		Quintile 5 (élevé)	29 700	79,4	7,4	5,89	(66,9 - 90,3)	0,0
	18 à 24 ans	Quintile 1 (faible)	31 000	68,6	8,7	5,99	(56,4 - 80,3)	0,9
		Quintile 2	40 700	71,1	8,5	6,05	(59,2 - 82,9)	3,0
		Quintile 3	47 000	76,8	5,1	3,88	(67,6 - 84,5)	1,1
		Quintile 4	56 900	75,4	7,0	5,27	(65,1 - 85,0)	0,6
		Quintile 5 (élevé)	38 900	73,8	6,3	4,61	(64,1 - 82,2)	0,9
	25 à 44 ans	Quintile 1 (faible)	114 600	69,8	4,9	3,42	(62,6 - 76,6)	0,5
		Quintile 2	123 000	67,6	4,4	3,01	(61,2 - 73,5)	1,9
		Quintile 3	161 700	73,0	3,1	2,29	(68,4 - 77,5)	0,2
		Quintile 4	162 700	79,5	2,4	1,87	(75,5 - 83,1)	0,5
		Quintile 5 (élevé)	206 200	82,7	2,1	1,70	(79,3 - 86,1)	0,1
	45 à 64 ans	Quintile 1 (faible)	59 100	57,1	5,5	3,17	(50,5 - 63,2)	4,1
		Quintile 2	84 300	57,2	5,3	3,01	(51,4 - 63,2)	2,1

Test global d'association de la proportion de la population se brossant les dents au moins deux fois par jour, parmi ceux ayant au moins une dent naturelle au maxillaire supérieur ou inférieur selon le niveau de revenu, stratifié selon la sous-population

Sous-population		Statistique
Sexe	Groupe d'âge	Valeur p ^a
Masculin	15 à 17 ans	0,1780
	18 à 24 ans	0,8111
	25 à 44 ans	<0,0001
	45 à 64 ans	<0,0001
	65 ans et plus	0,0002
Féminin	15 à 17 ans	0,3445
	18 à 24 ans	0,3076
	25 à 44 ans	0,0007
	45 à 64 ans	0,0094
	65 ans et plus	0,3197

Notes

Étant donné que le nombre de comparaisons deux à deux découlant de cette requête est supérieur à 30, elles ne sont pas présentées. Pour les résultats globaux significatifs au seuil de 5 %, il est suggéré d'effectuer un test *t* de la différence ou de comparer les intervalles de confiance deux à deux afin d'identifier la source de la différence.

Message
Étant donné le nombre élevé de comparaisons 2 à 2, ces dernières ne sont pas présentées.

Source : Infocentre de santé publique, octobre 2010.

6.3.3. Non-réponse partielle élevée et sous-estimation des effectifs

Proportion de non-réponse partielle supérieure au seuil critique

Comme déjà été mentionné à la section 2.3, lorsqu'on analyse des données d'enquête, il est en général raisonnable de faire l'hypothèse que, pour une non-réponse partielle inférieure à 5 % pour l'ensemble du Québec et à 10 % à l'échelle régionale ou locale, les risques de biais occasionnés par un déséquilibre entre les caractéristiques des non-répondants et celles des répondants, sont plutôt faibles. En effet, le biais des estimations est en général peu important comparativement à l'erreur d'échantillonnage (c'est-à-dire la précision de l'estimation) et peut être considéré comme négligeable.

Les questions pour lesquelles le taux de non-réponse partielle dépasse le seuil critique doivent en général être analysées plus à fond au regard des biais. Les non-répondants peuvent présenter des caractéristiques différentes des répondants, ce qui à son tour peut entraîner un impact non négligeable sur les estimations présentées (sous ou surestimation)⁶⁴.

Chacun des tableaux produits sur le portail de l'Infocentre affiche les pourcentages de non-réponse partielle de l'indicateur analysé. Afin d'aider les utilisateurs, une note est présentée au bas du tableau lorsque cette non-réponse dépasse le seuil critique correspondant au niveau géographique du tableau. Les tableaux 6-11 et 6-12 présentent des exemples de tableaux sans croisement affichant chacune de ces notes.

Tableau 6-11 Proportion de la population victime de blessure attribuable à des mouvements répétitifs au cours des 12 derniers mois dont la blessure la plus grave est survenue durant le travail rémunéré pour l'ensemble du Québec, EQSP 2008

Sous-population	Statistiques						
	Sexe	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin		179 100	47,6	4,1	1,97	(43,7 - 51,4)	3,8
Féminin		159 100	42,8	4,5	1,91	(39,2 - 46,4)	* 5,1
Total		338 400	45,2	3,0	1,34	(42,5 - 47,9)	4,4

Notes

^a : Pourcentage de non-réponse partielle supérieur à 5 % au niveau de l'ensemble du Québec.

Pour cette raison, il est recommandé d'effectuer un examen plus approfondi de la non-réponse pour cet indicateur.

Source : Infocentre de santé publique, octobre 2010.

⁶⁴ Pour savoir comment procéder au diagnostic et à l'analyse de la non-réponse partielle, consultez l'annexe 5.

Tableau 6-12 Proportion de la population victime de blessure attribuable à des mouvements répétitifs au cours des 12 derniers mois dont la blessure la plus grave est survenue durant le travail rémunéré, EQSP 2008

Sous-population	Territoire	Statistiques					
		Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
15-19 ans	13 Laval	*** 100	** 3,7	148,9	5,52	(0,0 - 23,3)	0,7
	Ensemble du Québec	11 800	** 21,5	25,8	5,55	(11,4 - 33,3)	* 6,7
20-24 ans	13 Laval	*** 1 000	** 52,5	50,3	26,42	(0,0 - 100,0)	** 24,7
	Ensemble du Québec	22 500	51,5	11,5	5,93	(39,0 - 63,2)	1,4
25-44 ans	13 Laval	6 200	51,0	13,4	6,84	(37,2 - 64,4)	5,2
	Ensemble du Québec	150 400	52,1	4,4	2,30	(47,3 - 56,6)	2,9
45-54 ans	13 Laval	5 200	51,3	14,7	7,51	(35,0 - 67,1)	6,6
	Ensemble du Québec	104 200	54,8	4,6	2,55	(49,9 - 59,9)	* 5,2
55-64 ans	13 Laval	2 200	* 43,8	18,5	8,10	(26,9 - 61,7)	9,9
	Ensemble du Québec	45 400	37,7	8,0	3,00	(32,1 - 43,7)	5,0
65 ans et +	13 Laval	*** 100	** 2,5	108,0	2,75	(0,0 - 11,1)	** 15,9
	Ensemble du Québec	2 400	** 4,8	25,3	1,21	(2,5 - 7,7)	* 9,3
Total	13 Laval	14 900	42,6	8,9	3,80	(34,8 - 50,2)	8,2
	Ensemble du Québec	338 400	45,2	3,0	1,34	(42,5 - 47,9)	4,4

Notes

^a : Pourcentage de non-réponse partielle supérieur à 5 % au niveau de l'ensemble du Québec. Pour cette raison, il est recommandé d'effectuer un examen plus approfondi de la non-réponse pour cet indicateur.

^{aa} : Pourcentage de non-réponse partielle supérieur à 10 % au niveau régional ou local. Pour cette raison, il est donc recommandé d'effectuer un examen plus approfondi de la non-réponse pour cet indicateur.

^{aaa} : Le nombre de répondants dans cette cellule ou dans la cellule complémentaire est inférieur à 5. Dans le cas de diffusion de ces données, il est nécessaire de procéder au masquage de données, selon les normes établies dans le guide méthodologique de l'EQSP.

Source : Infocentre de santé publique, octobre 2010.

Bien que la non-réponse partielle d'un indicateur soit faible, il est possible que celle de la variable utilisée comme croisement soit non négligeable. La proportion de non-réponse partielle affichée dans un tableau avec croisement correspond à la non-réponse de l'indicateur pour chacune des catégories de la variable de croisement. La taille des domaines d'étude formés par ces catégories est beaucoup plus faible que celle de l'indicateur sans croisement. Par conséquent, la précision des estimations est également plus faible. Pour cette raison, le seuil critique de 5 % pour l'ensemble du Québec est augmenté à 10 % pour les tableaux avec croisement et les notes des tableaux sont légèrement modifiées (voir le tableau 6-13).

Par ailleurs, pour les tableaux avec croisement, une autre note au bas du tableau affiche le pourcentage de non-réponse partielle de la variable de croisement (tableau 6-13).

Tableau 6-13 Proportion de la population ayant eu des symptômes de rhinite allergique au cours des 12 derniers mois qui ont eu des symptômes en été pour l'ensemble du Québec, EQSP 2008

Sous-population	Variable de croisement	Statistiques					
		Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	Oui	7 400	* 60,1	18,6	11,20	(36,1 - 80,4)	4,7
	Non	210 700	50,2	3,8	1,90	(46,3 - 53,9)	6,5
Féminin	Oui	8 900	66,8	13,1	8,77	(46,4 - 83,9)	* 11,8
	Non	288 800	49,9	3,2	1,61	(46,8 - 53,0)	6,1
Total	Oui	16 300	63,4	11,8	7,49	(47,5 - 77,1)	8,4
	Non	499 600	50,0	2,4	1,21	(47,6 - 52,3)	6,3

Notes

^a : Pourcentage de non-réponse partielle supérieur à 10 % au niveau de l'ensemble du Québec avec au moins une variable de croisement. Pour cette raison, il est donc recommandé d'effectuer un examen plus approfondi de la non-réponse pour cet indicateur.

- Pour ce tableau, la non-réponse partielle pour la variable « Prise de médicaments non prescrits » est de 4,0 %.
- En présence de non-réponse partielle à la variable de croisement, les effectifs de population présentés dans les tableaux croisés sont sous-estimés. En effet, la non-réponse partielle observée pour la variable de croisement n'est pas prise en compte dans le calcul des effectifs de l'indicateur. Pour obtenir des estimations non-biaisées des effectifs de l'indicateur analysé, référer au tableau sans croisement.

Source : Infocentre de santé publique, octobre 2010.

Sous-estimation des effectifs de population

Un autre élément impliquant la non-réponse partielle de la variable de croisement est que, à l'Infocentre, celle-ci n'est pas répartie comme il est suggéré de le faire à la section 2.3. Seule la non-réponse de l'indicateur l'est. Ceci a pour seule conséquence que les effectifs qui sont présentés dans les tableaux avec croisement sous-estiment les effectifs réels de la population⁶⁵. Afin de faire état de cette limite, une note est affichée, le cas échéant, au bas du tableau. Le tableau 6-14 présente un exemple de tableau affichant cette note.

⁶⁵ L'utilisateur qui souhaite obtenir des estimations non-biaisées des effectifs d'un indicateur selon une variable de croisement peut procéder de la façon suivante : produire les tableaux de l'indicateur avec et sans le croisement. Multiplier la proportion de la catégorie du croisement souhaité par l'effectif du tableau sans croisement pour cette même catégorie. Par exemple, si l'on souhaite obtenir le nombre de personnes qui ont déjà consommé des médicaments non prescrits parmi ceux qui ont déjà songé sérieusement au suicide au cours de leur vie, il faut prendre la proportion de personnes qui ont déjà consommé des médicaments non prescrits parmi ceux qui ont déjà songé sérieusement au suicide dans le tableau croisé et multiplier cette proportion par l'effectif des personnes qui ont consommé des médicaments non prescrits du tableau sans croisement.

Tableau 6-14 Proportion de la population ayant consommé des médicaments qui ne lui étaient pas prescrits, pour avoir un effet, au cours de la vie pour l'ensemble du Québec, EQSP 2008

Variable de croisement	Statistiques					
	Avoir songé sérieusement au suicide ou tenté de se suicider au cours de sa vie	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %
Oui	58 500	6,7	8,6	0,57	(5,5 - 7,8)	1,8
Non	50 700	1,0	9,4	0,09	(0,8 - 1,1)	2,5

Notes

- Pour ce tableau, la non-réponse partielle pour la variable « » est de 2,3 %.
 - En présence de non-réponse partielle à la variable de croisement, les effectifs de population présentés dans les tableaux croisés sont sous-estimés. En effet, la non-réponse partielle observée pour la variable de croisement n'est pas prise en compte dans le calcul des effectifs de l'indicateur. Pour obtenir des estimations non-biaisées des effectifs de l'indicateur analysé, référer au tableau sans croisement.
- Source : Infocentre de santé publique, octobre 2010.

6.3.4. Indice du niveau de revenu

Sur le portail de l'Infocentre, l'indice du niveau de revenu ajusté pour le seuil de faible revenu selon la taille du ménage et de la communauté (ou niveau de revenu) est disponible sous deux formes : une version détaillée présentée en quintiles (1-faible à 5-élevé) et une version dichotomique distinguant le quintile inférieur (1-faible) des autres quintiles (2 et plus). Un utilisateur peut donc produire des tableaux à partir de l'un ou l'autre de ces indices. Puisque l'indice repose sur les quintiles de la distribution des ratios de revenu du ménage au seuil de faible revenu, il est donc normal d'obtenir des proportions globales avoisinant les 20 % pour chacun des niveaux.

Comme discuté aux sections 2.4.3 et 3.1.2, une proportion trop importante de répondants a refusé ou n'était pas en mesure de fournir le revenu total du ménage. Cette non-réponse a donc dû être traitée par imputation. Pour produire des estimations de variances adéquates, cette imputation devrait être considérée dans ces calculs. Or, ce n'est pas le cas pour les données de l'EQSP disponibles à l'Infocentre. Ce qui fait que la variance des estimations impliquant le niveau de revenu peut-être sous-estimée de manière importante. Il a également été mentionné à la section 2.4.3 que les principales sous-populations affectées par cette imputation, donc celles pour lesquelles la sous-estimation est la plus importante, sont les jeunes de 15 à 24 ans et les femmes de 65 ans et plus. Or, l'analyse effectuée sur les non-répondants partiels pour la variable du revenu a également montré que la sous-estimation de la variance touchait principalement les quintiles 2 à 5 de l'indicateur et que, par le fait même, l'impact sur le quintile inférieur était beaucoup moins important.

Considérant cela, le Comité tripartite a convenu que :

- Pour les tableaux produits sur le portail à partir de l'indicateur de niveau de revenu dichotomique, aucune mention particulière n'est nécessaire lors d'une telle production.
- Lorsque l'indicateur de niveau de revenu est celui présenté en quintiles, une note sera produite au bas du tableau produit afin de mettre en garde l'utilisateur contre un résultat qui pourrait s'avérer erroné. Cette note suggèrera à l'utilisateur d'interpréter avec prudence les différences significatives dont le seuil observé (valeur-p) est près du seuil de signification, ou que les deux intervalles de confiance sont près du chevauchement, en particulier pour les 15-24 ans et les femmes de 65 ans et plus. Le tableau 6-15 présente un exemple de tableau pour lequel cette note apparaît.

Tableau 6-15 Répartition de la population se percevant pauvre ou très pauvre selon la durée dans cette situation pour l'ensemble du Québec, EQSP 2008

Variable de croisement	Indicateur	Statistiques					
		Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Quintile 1 (faible)	Moins d'un an	55 800	14,9	11,2	1,66	(11,7 - 18,1)	5,9
	Entre 1 et 4 ans	122 300	32,6	5,7	1,85	(29,0 - 36,4)	5,9
	Entre 5 et 9 ans	76 200	20,3	7,5	1,51	(17,3 - 23,4)	5,9
	10 ans et plus	121 200	32,3	5,2	1,68	(29,0 - 35,5)	5,9
Quintile 2	Moins d'un an	27 800	15,6	13,7	2,14	(11,5 - 20,2)	6,5
	Entre 1 et 4 ans	64 000	36,0	7,9	2,85	(30,4 - 41,6)	6,5
	Entre 5 et 9 ans	30 500	17,2	11,5	1,97	(13,2 - 21,3)	6,5
	10 ans et plus	55 400	31,2	8,3	2,58	(25,9 - 36,5)	6,5
Quintile 3	Moins d'un an	14 800	* 22,0	16,7	3,68	(14,9 - 29,7)	4,1
	Entre 1 et 4 ans	24 100	36,0	11,3	4,05	(28,0 - 44,4)	4,1
	Entre 5 et 9 ans	9 500	* 14,1	20,5	2,89	(8,6 - 20,2)	4,1
	10 ans et plus	18 700	27,9	12,9	3,58	(20,7 - 35,4)	4,1
Quintile 4	Moins d'un an	7 200	* 25,2	20,8	5,22	(15,0 - 35,7)	1,4
	Entre 1 et 4 ans	7 900	* 27,6	18,5	5,11	(18,0 - 38,5)	1,4
	Entre 5 et 9 ans	5 600	* 19,5	22,0	4,29	(11,8 - 29,0)	1,4
	10 ans et plus	7 900	* 27,7	15,5	4,30	(19,5 - 37,9)	1,4
Quintile 5 (élevé)	Moins d'un an	3 200	** 20,4	32,6	6,67	(8,4 - 35,2)	0,5
	Entre 1 et 4 ans	5 300	** 33,6	25,3	8,51	(16,9 - 51,3)	0,5
	Entre 5 et 9 ans	1 900	** 11,9	36,8	4,39	(3,4 - 22,4)	0,5
	10 ans et plus	5 400	* 34,0	22,0	7,50	(20,1 - 50,4)	0,5

Notes

· La variance des estimations est sous-estimée, car elle ne tient pas compte de l'imputation effectuée sur l'indicateur du revenu. L'impact de l'imputation sur la précision des proportions est variable d'une sous-population à une autre et peut être important. Il est suggéré d'interpréter avec prudence les différences significatives dont le seuil observé (valeur-p) est près du seuil de signification, ou dont les 2 intervalles de confiance sont près du chevauchement, en particulier pour les 15-24 ans et les femmes de 65 ans et plus.

Source : Infocentre de santé publique, octobre 2010.

6.3.5. Diffusion par Réseaux Locaux de Services (RLS)

Disponibilité pour six régions sociosanitaires seulement

Comme brièvement mentionné à la section 2.2, le niveau géographique utilisé pour la pondération correspond à celui des objectifs de l'enquête, soit l'échelle locale pour les régions qui ont fait l'achat d'un échantillon supplémentaire (Saguenay–Lac-Saint-Jean, Mauricie–Centre-du-Québec, Abitibi-Témiscaminque, Gaspésie–Îles-de-la-Madeleine, Chaudière-Appalaches et Lanaudière) et l'échelle régionale pour les autres RSS. Il a donc été décidé de se limiter aux objectifs de l'enquête, en raison principalement du fait qu'un ajustement plus fin pour les régions sans suréchantillon aurait eu pour impact de diminuer la qualité de l'ajustement de la pondération au niveau de l'ensemble du Québec. En fait, l'ensemble des opérations méthodologiques développées pour l'EQSP n'avait pas pour objectif de fournir des données de qualité à l'échelle locale pour les régions qui n'ont pas fait d'achat d'unités supplémentaires

Par conséquent, le Comité tripartite a décidé de ne pas diffuser d'estimations à l'échelle locale pour ces régions sur le portail de l'Infocentre.

Découpage non officiel des RLS

La Direction de santé publique de l'ASSS de Chaudière-Appalaches a demandé de scinder en deux parties le RLS du Grand Littoral (1202), soit l'actuelle ville de Lévis (Grand Littoral Urbain – 1202A) et le reste de ce territoire (Grand Littoral Rural – 1202B) tout en conservant les mêmes objectifs de précision que les RLS. Cette demande a été reçue favorablement et deux découpages infrarégionaux sont donc offerts pour cette région sur le portail de l'Infocentre, soit le niveau local comme défini par le MSSS, qui regroupe les parties urbaines et rurales du RLS du Grand Littoral, et le niveau local non officiel, qui distingue les parties urbaines et rurales du RLS du Grand Littoral. Ainsi, lors d'une requête paramétrable à l'échelle locale, lorsque la région 12 est sélectionnée, un paramètre supplémentaire apparaît dans la fenêtre de la requête permettant de sélectionner l'un des deux découpages suivants :

- 12 Chaudière-Appalaches
- 12 Chaudière-Appalaches (découpage RLS non-officiel)

Pour ce dernier, une note apparaît au bas du tableau produit afin d'indiquer que le découpage ne correspond pas au découpage officiel du MSSS. Le tableau 6-16 présente un exemple de cette situation.

Tableau 6-16 Répartition de la population selon la perception du poids pour les RLS de la région 12 Chaudière-Appalaches, EQSP 2008

Territoire RLS (CSSS)	Indicateur Perception du poids	Statistiques					
		Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
1201 - RLS de Lac-Échemin (CSSS des Échemins)	Excès de poids	5 300	37,5	6,3	2,35	(32,7 - 42,1)	2,5
	Trop maigre	400	** 2,7	29,5	0,60	(1,2 - 4,5)	2,5
	À peu près normal(e)	8 400	59,8	4,0	2,39	(55,1 - 64,7)	2,5
1202A - Grand Littoral Urbain	Excès de poids	39 800	36,8	6,2	2,29	(32,2 - 41,2)	0,9
	Trop maigre	1 200	** 1,1	35,7	0,40	(0,3 - 2,1)	0,9
	À peu près normal(e)	67 200	62,1	3,7	2,29	(57,8 - 66,6)	0,9
1202B - Grand Littoral Rural	Excès de poids	29 400	39,2	5,9	2,30	(34,3 - 43,9)	1,9
	Trop maigre	900	** 1,2	44,9	0,52	(0,2 - 2,4)	1,9
	À peu près normal(e)	44 800	59,7	3,8	2,29	(55,1 - 64,5)	1,9
1203 - RLS de Beauce (CSSS de Beauce)	Excès de poids	19 000	34,4	6,9	2,36	(29,9 - 39,3)	3,2
	Trop maigre	600	** 1,1	47,1	0,51	(0,1 - 2,3)	3,2
	À peu près normal(e)	35 600	64,5	3,7	2,40	(59,5 - 69,1)	3,2
1204 - RLS de l'Amiante (CSSS de la région de Thetford)	Excès de poids	13 500	37,8	6,4	2,41	(33,1 - 42,6)	4,5
	Trop maigre	700	** 2,1	29,4	0,61	(0,9 - 3,5)	4,5
	À peu près normal(e)	21 400	60,1	4,0	2,42	(55,3 - 64,9)	4,5
1205 - RLS de Montmagny-L'Islet (CSSS de Montmagny-L'Islet)	Excès de poids	12 000	35,2	6,8	2,39	(30,5 - 39,8)	5,0
	Trop maigre	700	** 2,0	38,2	0,78	(0,6 - 3,7)	5,0
	À peu près normal(e)	21 400	62,8	3,9	2,44	(58,1 - 67,5)	5,0
Total de la région 12 Chaudière-Appalaches	Excès de poids	119 000	36,9	3,0	1,10	(34,8 - 39,1)	2,4
	Trop maigre	4 500	* 1,4	16,8	0,23	(0,9 - 1,9)	2,4
	À peu près normal(e)	198 900	61,7	1,8	1,09	(59,5 - 63,9)	2,4

Notes

· En raison de la stratégie de suréchantillonnage retenue par la région 12 visant à prendre en considération les caractéristiques spécifiques des milieux de vie, le territoire de RLS 1202 a été divisé en deux parties, soit « 1202A Grand Littoral Urbain » et « 1202B Grand Littoral Rural ».

Source : Infocentre de santé publique, octobre 2010.

La Direction de santé publique de l'ASSS du Saguenay-Lac-Saint-Jean a demandé à retirer la municipalité de Saint-Ludger-de-Milot de son RLS officiel (Lac-Saint-Jean-Est - RLS206) pour l'inclure dans le RLS Maria-Chapdelaine (RLS205), car cette municipalité est, dans les faits, desservie par les

établissements de ce RLS. Cette demande a été accueillie favorablement⁶⁶, bien que ce reclassement a pour effet de rendre le découpage par RLS dans cette région non-conforme au découpage officiel du MSSS. Par conséquent, le découpage officiel ne peut être recréé pour la région sociosanitaire du Saguenay–Lac-Saint-Jean et un seul découpage infrarégional est offert sur le portail. Ce découpage contient les deux RLS concernés sous les noms de :

- RLS de Maria-Chapdelaine incluant Saint-Ludger-de-Milot (CSSS de Maria-Chapdelaine)
- RLS de Lac-Saint-Jean-Est excluant Saint-Ludger-de-Milot (CSSS de Lac-Saint-Jean-Est).

Encore une fois, une note au bas des tableaux produits sur le portail de l’Infocentre fera état de cette situation particulière (voir le tableau 6-17).

Tableau 6-17 Répartition de la population selon la perception du poids pour les RLS de la région 02 Saguenay - Lac-Saint-Jean, EQSP 2008

Territoire RLS (CSSS)	Indicateur Perception du poids	Statistiques					
		Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
0201 - RLS de La Baie (CSSS Cléophas-Claveau)	Excès de poids	7 500	39,1	5,1	1,99	(35,3 - 43,0)	2,3
	Trop maigre	500	** 2,5	25,6	0,63	(1,3 - 3,9)	2,3
	À peu près normal(e)	11 100	58,4	3,5	2,04	(54,4 - 62,2)	2,3
0202 - RLS de Chicoutimi (CSSS de Chicoutimi)	Excès de poids	22 800	36,3	6,0	2,19	(32,0 - 40,8)	2,1
	Trop maigre	800	** 1,3	34,3	0,46	(0,4 - 2,4)	2,1
	À peu près normal(e)	39 200	62,3	3,5	2,18	(57,9 - 66,6)	2,1
0203 - RLS de Jonquière (CSSS de Jonquière)	Excès de poids	19 400	35,9	5,3	1,90	(32,0 - 40,0)	1,4
	Trop maigre	1 200	** 2,2	26,7	0,60	(1,1 - 3,7)	1,4
	À peu près normal(e)	33 400	61,9	3,2	1,97	(57,8 - 65,9)	1,4
0204 - RLS du Domaine-du-Roy (CSSS du Domaine-du-Roy)	Excès de poids	8 900	34,6	5,5	1,90	(30,8 - 38,5)	1,7
	Trop maigre	300	** 1,2	35,3	0,42	(0,4 - 2,2)	1,7
	À peu près normal(e)	16 600	64,2	3,0	1,95	(60,3 - 68,0)	1,7
0205 - RLS de Maria-Chapdelaine incluant Saint-Ludger-de-Milot (Excès de poids	8 700	40,6	5,1	2,08	(36,5 - 44,8)	1,9
	Trop maigre	600	* 2,9	24,3	0,69	(1,5 - 4,4)	1,9
	À peu près normal(e)	12 100	58,6	3,7	2,10	(52,2 - 60,7)	1,9
0206 - RLS de Lac-Saint-Jean-Est excluant Saint-Ludger-de-Milot	Excès de poids	14 400	35,2	5,7	2,02	(31,4 - 39,5)	2,8
	Trop maigre	1 100	* 2,7	23,6	0,64	(1,5 - 4,2)	2,8
	À peu près normal(e)	25 500	62,1	3,3	2,07	(57,8 - 66,1)	2,8
Total de la région 02 Saguenay - Lac-Saint-Jean	Excès de poids	81 700	36,5	2,7	0,97	(34,8 - 38,4)	2,0
	Trop maigre	4 600	2,0	12,4	0,25	(1,6 - 2,6)	2,0
	À peu près normal(e)	137 800	61,5	1,6	0,99	(59,6 - 63,2)	2,0

Notes

La municipalité de Saint-Ludger-de-Milot, bien que se trouvant officiellement dans le RLS 0206 (Lac-Saint-Jean-Est), est desservie en pratique par les établissements de santé du territoire Maria-Chapdelaine (RLS 0205). Les territoires de RLS 0205 et 0206 sont non-officiels en raison de la stratégie de suréchantillonnage retenue par la région 02.

Source : Infocentre de santé publique, octobre 2010.

6.3.6. Point de césure de l’indice de détresse psychologique

Les écrits scientifiques montrent généralement que le seuil choisi pour cibler la détresse psychologique ne peut être repris tout simplement d’une enquête à l’autre et d’une population à l’autre. Des tests de sensibilité et spécificité seraient nécessaires pour déterminer le meilleur seuil. Toutefois, seules les enquêtes portant spécifiquement sur la santé mentale peuvent se permettre de tels tests.

⁶⁶ Ce reclassement a été fait pour l’ensemble des opérations méthodologiques.

Le point de césure sur l'échelle variant d'une enquête à l'autre, le quintile semble un compromis acceptable. Pour l'EQSP, il a été convenu de retenir le quintile supérieur de l'échelle de détresse psychologique afin de décrire les sous-groupes les plus vulnérables. Ce choix a toutefois comme limite de ne pouvoir interpréter l'indicateur à titre de prévalence de la détresse psychologique, car le seuil utilisé n'est pas un seuil clinique et ne réfère pas à une pathologie spécifique.

Puisque l'indicateur repose sur les quintiles de la distribution de l'échelle de détresse psychologique, il est donc normal d'obtenir une proportion de détresse psychologique élevée avoisinant les 20 %.

Dans ces circonstances, il est essentiel de remettre en question la comparabilité de cet indicateur avec d'autres enquêtes, telles l'*Enquête québécoise sur des conditions de travail, d'emploi et de santé et sécurité du travail* (EQCOTESST) de l'ISQ et les cycles généraux et 1.2 de l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC) de Statistique Canada.

6.3.7. Indicateurs portant sur les enfants de 3 à 14 ans

Choix du déclarant pour la section enfant

Le choix de la personne à laquelle on confie le mandat de répondre à propos d'un tiers peut avoir un impact sur les estimations et sur la comparabilité avec d'autres enquêtes. Pour l'EQSP, on a choisi de mesurer la perception de l'un des parents plutôt que celle des mères ou des pères seulement, ou celle de la personne « la mieux placée » dans le ménage pour répondre à des questions sur l'enfant. Cela occasionne principalement deux limites : d'abord, il n'est pas approprié de croiser les questions de la section sur les enfants selon des caractéristiques individuelles du déclarant. Ensuite, ce choix du déclarant peut restreindre sensiblement la comparabilité des données avec d'autres enquêtes. En effet, si une enquête utilise un autre type de déclarant pour recueillir l'information au sujet d'un tiers que celui privilégié pour l'EQSP, les estimations ne seront pas comparables. Il faut donc être prudent lors de la comparaison des résultats de l'EQSP avec ceux d'autres enquêtes.

Par ailleurs, les indicateurs des diagnostics médicaux ne s'appuient pas sur la perception, mais sur la connaissance objective du parent du fait qu'un professionnel de la santé a posé un diagnostic sur la santé de l'enfant. Cette déclaration d'un diagnostic ne varie donc pas en fonction du répondant, que celui-ci soit le père ou la mère. Par conséquent, une seule des deux limites énoncées au paragraphe précédent s'applique aux indicateurs de diagnostics médicaux, soit celle stipulant qu'on ne peut croiser ces indicateurs avec les caractéristiques individuelles du répondant.

Le SDQ a une sensibilité et une spécificité supérieures lorsqu'il est possible de se fier à plus d'un répondant, par exemple en combinant les réponses du parent et de l'enseignant, ou mieux encore du parent, de l'enseignant et l'enfant (Goodman et autres, 2000). Dans le cas de l'EQSP, seul le questionnaire au parent a été utilisé, ce qui constitue une limite qu'il faut considérer dans l'interprétation de cet indicateur.

Regroupement permis pour la variable de l'âge pour chaque niveau géographique

Comme présenté au tableau 1-9 de la section 1.5.1, le nombre de répondants à la section enfants s'élève à 5 248 pour l'ensemble du Québec. Ce nombre varie entre 217 et 524 répondants selon la région sociosanitaire, soit une moyenne d'environ 330 répondants par région.

Lors de la planification de l'enquête, les objectifs, en matière de précision des estimations, se limitaient à des objectifs régionaux, tous groupes d'âge confondus. Pour cette raison, et parce que le nombre de répondants par région à la section enfant est plutôt limité, les groupes d'âge offerts sur le portail de l'Infocentre pour les indicateurs 18.1 à 18.9 sont les suivants :

- Pour l'ensemble du Québec : 3 à 5 ans, 6 à 8 ans, 9 à 11 ans et 12 à 14 ans
- Pour le niveau régional : 6 à 14 ans

Aucun tableau n'est disponible à l'échelle locale des RLS.

6.3.8. Indicateurs portant sur la période « au cours de la vie »

Dans le cas des indicateurs construits à partir de questions portant sur la période « au cours de la vie », il apparaît évident que certains renseignements impliquant des caractéristiques contemporaines des répondants (âge, revenu, scolarité, etc.) ne se prêtent pas bien aux analyses. Par conséquent, certaines variables de croisement n'ont pas été retenues pour ces indicateurs. Il s'agit des variables :

- Niveau de scolarité
- Niveau de revenu
- Indices de défavorisation
- Occupation principale
- Statut d'emploi
- Composition du ménage
- Indice de détresse psychologique

Cependant, le paramètre « Âge » a été conservé pour permettre l'étude du phénomène dans des sous-populations.

6.3.9. Indicateur construit à partir de la médiane des scores

Pour les indicateurs construits à partir de la médiane des scores de l'échantillon, par exemple les indicateurs 9.2 (proportion des travailleurs ayant une latitude décisionnelle faible au travail) et 9.4 (proportion des travailleurs ayant des exigences psychologiques élevées au travail), la comparaison avec d'autres sources de données est impossible. Une note au bas du tableau est automatiquement produite pour ces indicateurs (voir le tableau 6-18). Plus de détails sont fournis dans la fiche des indicateurs impliqués.

Tableau 6-18 Proportion des travailleurs ayant une latitude décisionnelle faible au travail, EQSP 2008

Territoire	Statistiques							
	Région de résidence	Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
05 Estrie		68 800	44,3	3,7	1,63	(41,1 - 47,5)	1,9	ns
04 Mauricie et Centre-du-Québec		112 200	47,2	3,2	1,49	(44,3 - 50,2)	2,6	

Notes

- Les travailleurs ont une latitude décisionnelle faible au travail par rapport à la médiane des scores des répondants. Par conséquent, il n'est pas possible de comparer l'indicateur avec d'autres enquêtes.

Source : Infocentre de santé publique, octobre 2010.

7. Protection des renseignements personnels⁶⁷

La protection des renseignements personnels (PRP) doit se faire à deux niveaux : celui des données individuelles des répondants (masquage du fichier de microdonnées - section 7.1) et celui des données agrégées (masquage des tableaux – section 7.2).

7.1. Confidentialité des fichiers de microdonnées

Le traitement de confidentialité (masquage) appliqué aux fichiers de microdonnées est de trois types : la suppression de valeur, le regroupement de catégories et l'élimination de variables. L'importance du masquage peut varier selon le moyen utilisé pour diffuser les fichiers de microdonnées (au CADRISQ, dans les locaux d'un organisme public, en accès à distance). En effet, étant donné que les mesures de contrôle informatique et physique sont différentes d'un mode de diffusion à un autre, les mesures de contrôle statistique (masquage) sont également différentes, et varient de façon à obtenir un équilibre entre ces deux mesures, permettant d'en venir à un contrôle de la divulgation qui est acceptable et équivalent pour chacun des modes de diffusion.

Le masquage appliqué au fichier disponible en accès à distance (FMII) est moins sévère que le masquage appliqué au fichier disponible dans les locaux d'un organisme public (FARE), ce qui permet un potentiel analytique supérieur pour le FMII (voir le tableau 1-1 à la section 3 de la première partie).

Il peut être intéressant pour l'utilisateur d'un fichier de microdonnées masqué de connaître le potentiel analytique du fichier par rapport au fichier sans masquage (fichier maître). Cette information pourrait, par exemple, lui indiquer la perte en matière d'analyse des données qu'il accepte en utilisant un fichier masqué. Il existe plusieurs mesures du potentiel analytique d'un fichier. L'ISQ a produit un nombre important de ces mesures pour les deux fichiers masqués de l'EQSP 2008 par rapport au fichier maître de l'enquête. Nous avons choisi de rapporter au tableau 1-1, la plus intelligible d'entre elles, soit la « mesure d'association – p-value ».

Cette mesure du potentiel analytique des fichiers masqués renseigne sur le pourcentage de conclusions des tests ayant changées au seuil de 5 %, pour un sous-ensemble de variables, par rapport aux conclusions obtenues à partir du fichier maître pour ces mêmes variables. La valeur de la mesure d'association – p-value, présentée dans le tableau 1-1, porte spécifiquement sur le sous-ensemble des variables affectées par le masquage des données⁶⁸. Or, la valeur de cette même mesure, évaluée sur l'ensemble des variables du fichier, serait évidemment inférieure à celle présentée dans ce tableau. En fait, celle-ci serait tout près de zéro étant donné le nombre très important de variables du fichier. Sur la base de cette information, nous pouvons conclure que le potentiel analytique des deux fichiers masqués est excellent, vu la proportion plutôt faible de tests qui changent de conclusion au seuil de 5 %.

Nous limiterons à ces quelques renseignements les éléments relatifs aux fichiers masqués. Les utilisateurs qui ont accès à l'un de ces fichiers peuvent consulter la documentation jointe avec les fichiers pour en connaître davantage.

⁶⁷ Certains éléments de cette section ont été extraits intégralement de documents de l'ISQ sur la confidentialité des microdonnées et des tableaux.

⁶⁸ Il s'agit des variables masquées dans au moins un des fichiers FARE et FMII, à l'exception des variables continues et des variables géographiques des RSS et des RLS.

7.2. Confidentialité des tableaux pour diffusion

Bien que le fichier de microdonnées de l'EQSP disponible à l'Infocentre soit un fichier non masqué, en raison de l'accord du répondant à partager ses renseignements avec cet organisme, il existe tout de même une obligation contractuelle pour le diffuseur de tableaux de respecter le caractère confidentiel des données diffusées sous forme de tableaux. Il est important de savoir que la divulgation d'information ne survient pas seulement lorsque l'on donne le nom et l'adresse d'un répondant, elle peut également survenir de façon indirecte lorsque l'on croise différentes caractéristiques du répondant. Statistique Canada, dans son document « *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*⁶⁹ », donne une définition assez complète du risque de divulgation et présente de nombreux exemples de divulgation. L'encadré présenté à la fin de la section est un extrait tiré de ce document. Il fournit de l'information supplémentaire sur la divulgation d'information.

L'ISQ s'est doté, il y a quelques années, d'un ensemble de règles de confidentialité pour parer à la divulgation de renseignements confidentiels résultant de la diffusion de tableaux. L'application de ces règles permet d'identifier quels tableaux ou résultats comportent un risque de divulgation et quels traitements de confidentialité doivent être appliqués à ces tableaux. L'Infocentre est soumis aux mêmes obligations que l'ISQ pour les données de l'EQSP, c'est pourquoi les utilisateurs de données de cette enquête doivent appliquer des règles équivalentes à celles développées par l'ISQ lors de la diffusion de tableaux.

Ainsi, avant toute diffusion⁷⁰ d'un tableau ou d'un résultat produit à partir des données de l'EQSP sur le portail de l'Infocentre, l'utilisateur de niveau 20 ou 30⁷¹ doit s'assurer qu'aucune cellule du tableau n'est jugée confidentielle. Pour ce faire, celui-ci doit s'assurer qu'il y ait un minimum de 5 répondants dans chacune des cellules. Étant donné que le nombre de répondants n'est pas indiqué dans les tableaux produits sur le portail, une note s'affichera au bas de ceux-ci lorsque des cellules seront jugées confidentielles selon ce critère. Le tableau 7-1 présente la note affichée dans le cas d'un indicateur en répartition, alors que le tableau 7-2 présente celle affichée dans le cas d'une proportion.

Dans l'éventualité où l'un de ces messages apparaît au bas du tableau, donc qu'il y a des cellules confidentielles, l'utilisateur devra :

- Renoncer à la diffusion de ce résultat.
- Utiliser un indicateur pour lequel les catégories problématiques sont regroupées avec d'autres catégories.
- Dans le cas d'un tableau avec croisement, utiliser une variable de croisement pour laquelle les catégories problématiques sont regroupées avec d'autres catégories.
- Masquer la (les) cellule(s) problématique(s) du tableau.

Bien entendu, dans le cas de cette dernière option, étant donné qu'une cellule confidentielle, même masquée, peut être déduite par une simple soustraction des valeurs des autres résultats, il est obligatoire de masquer, pour chaque cellule confidentielle, une autre cellule du tableau (une pour chaque marginale présentée dans le tableau).

⁶⁹ Statistique Canada, Centres de données de recherche, *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*, octobre 2005, p. 23-24.

⁷⁰ Le terme diffusion comprend, entre autres, la publication, la sortie des locaux sécurisés de l'organisme, la transmission du résultat qu'elle soit sous forme papier, informatique ou verbale, etc.

⁷¹ Les utilisateurs des niveaux 40 et 50 n'ont pas à vérifier la confidentialité des tableaux qu'ils diffusent, car ceux-ci ont déjà été vérifiés et masqués avant de leur être remis.

Tableau 7-1 Note s'affichant au bas du tableau lorsqu'une cellule est jugée confidentielle pour un indicateur en répartition

Le nombre de répondants servant à estimer la population et la proportion dans cette cellule est inférieur à 5. Dans le cas de diffusion de ces données, il est nécessaire de procéder au masquage de données, selon les normes établies dans le guide méthodologique de l'EQSP.

Tableau 7-2 Note s'affichant au bas du tableau lorsqu'une cellule est jugée confidentielle pour un indicateur en proportion (indicateur dichotomique)

Le nombre de répondants servant à estimer la population et la proportion dans cette cellule, ou dans la cellule complémentaire, est inférieur à 5. Dans le cas de diffusion de ces données, il est nécessaire de procéder au masquage de données, selon les normes établies dans le guide méthodologique de l'EQSP.

L'encadré de la page suivante fournit de l'information supplémentaire sur la confidentialité et sur la divulgation d'information confidentielle lors de la diffusion de tableaux de résultats.

La divulgation signifie qu'un lien peut être établi entre des données diffusées et des répondants particuliers (personnes, ménages, organisations).

Types de divulgation

Il y a trois types de divulgation : la divulgation de l'identité, la divulgation d'attributs et la divulgation par recoupement.

La **divulgation de l'identité** se produit lorsqu'un individu peut être identifié à partir de données diffusées, ce qui entraîne la révélation d'information à son sujet.

La **divulgation d'attributs** survient quand de l'information confidentielle est révélée et peut être attribuée à un individu. Il n'est pas nécessaire qu'un individu précis soit identifié ou qu'une valeur précise soit révélée pour que cela se produise. Par exemple, le fait de diffuser une fourchette de salaires étroite concernant une profession particulière pour une région donnée peut constituer un cas de divulgation.

La **divulgation par recoupement** survient lorsque de l'information diffusée peut être combinée pour obtenir des données confidentielles. Il y a lieu de s'assurer d'examiner toutes les données qui doivent être diffusées. Alors qu'un tableau en soi peut ne pas révéler d'information confidentielle, la divulgation peut survenir par la combinaison de l'information provenant de plusieurs sources, dont des sources externes (les données supprimées dans un tableau peuvent être dérivées à partir d'autres tableaux, par exemple).

Quelques exemples de divulgation dans les données d'enquête

- Une personnalité (un athlète professionnel, par exemple) est sélectionnée dans une enquête et l'information diffusée à propos de sa localité, notamment le revenu le plus élevé dans cette localité, a presque assurément été déclarée par elle (divulgation de l'identité).
- Les résultats d'une enquête longitudinale mettent en évidence un ménage qui a un profil migratoire inusité, ce qui mène à son identification (divulgation de l'identité).
- Les parents d'un jeune de 16 ans sélectionné dans une enquête voient un tableau montrant que toutes les personnes de 16 ans de l'échantillon dans leur région ont consommé des drogues (divulgation d'attributs).
- Un article de journal fait état d'une plainte déposée par un veuf de 37 ans à propos d'une enquête à laquelle il participait, alors qu'un tableau croisé montre qu'il y a seulement deux veufs dans la trentaine qui font partie de l'échantillon (cela mène éventuellement à la divulgation de l'identité ou d'attributs).
- En combinant plusieurs résultats, il est possible d'obtenir une information volontairement exclue d'un fichier de microdonnées à grande diffusion parce que cela présentait un risque de divulgation trop élevé (le pays de naissance d'un immigrant récent, par exemple).

À noter que même l'apparence de divulgation peut ternir la réputation d'un bureau de la statistique relativement à la confidentialité. Un tort peut être causé même si l'on s'aperçoit que la mauvaise personne (ou ménage) a été identifiée dans les deux premiers exemples. En outre, réfuter une identification erronée peut accroître le risque d'exposer l'identité des répondants réels.

III Recommandations pour l'interprétation des résultats

Les recommandations présentées dans ce chapitre s'inspirent des travaux produits dans le cadre des enquêtes de santé de l'Institut de la statistique du Québec; elles ont été particulièrement adaptées à l'interprétation des résultats d'une enquête transversale par sondage, c'est-à-dire lorsque la collecte des données se fait auprès d'un échantillon de la population.

1. Lien de causalité

Les tests du khi-deux servent à établir s'il existe ou non un lien entre deux variables catégorielles. Dans le cas des enquêtes transversales, on ne peut pas parler de lien de causalité entre les variables. Les résultats doivent être présentés en considérant cette restriction. Ainsi, les verbes « influencer », « modifier », « avoir un impact », « avoir pour effet » ou « être affecté » ne peuvent être utilisés puisqu'ils sous-entendent une relation causale. Des expressions du genre « être associé », « être lié » sont plus pertinentes. On ne doit pas dire :

- « Le statut d'activité et la catégorie professionnelle influencent aussi la qualité de l'alimentation... »
- « Les données confirment le fort effet de sexe... »

Il faut plutôt écrire :

- « Le statut d'activité et la catégorie professionnelle sont aussi liés à la qualité de l'alimentation... » ou encore « La qualité de l'alimentation diffère selon le statut d'activité et la catégorie professionnelle. »
- « L'association avec le sexe est très élevée... »

2. Inférence à la population visée

Il ne faut pas décrire les résultats en parlant des répondants; comme les données sont pondérées, on doit plutôt discuter des résultats en les associant à la population visée. On ne doit pas écrire :

- « Les répondantes à l'enquête révèlent qu'environ 4 % des enfants ne vivaient pas avec leur père biologique... »

Il faut plutôt écrire :

- « Les mères québécoises révèlent qu'environ 4 % des enfants ne vivaient pas avec leur père biologique... »

Cependant, il est approprié de parler de « répondants » dans un cas comme suit :

- Le répondant devait indiquer le nombre de fois au cours duquel il s'est enivré au cours des 12 derniers mois.

3. Référence à un nombre

La présentation des estimations de résultats ne doit pas laisser supposer qu'on travaille avec des nombres, mais plutôt avec des proportions; il ne faut pas utiliser des formulations comme :

- « Un peu plus de Québécois se disent... » ou
- « Moins de femmes que d'hommes... »

Il faut plutôt utiliser des expressions comme :

- « Sont plus nombreux, en proportion, à... »
- « Il y a proportionnellement plus de femmes que d'hommes... »
- « Toute proportion gardée, il y a plus de... »
- « La proportion de personnes de tel groupe ayant telle caractéristique est plus grande que celle observée dans tel autre groupe. »

4. Estimations et non des valeurs exactes

Le texte doit refléter le fait que les statistiques présentées sont des estimations, pas des valeurs exactes. Il faut utiliser des expressions comme « environ », « près de », « on estime », « on évalue à », « au moins », etc. Par exemple, la formulation suivante n'est pas adéquate :

- « Dans la population, 31 % des gens souffrent de la maladie... »

Il faut plutôt dire :

- « Dans la population, environ 31 % des gens souffrent de la maladie... »
- « Selon cette étude, 31 % des gens souffrent de la maladie... »

5. Différence statistiquement non significative

Un test statistique effectué dans le cadre d'une enquête par sondage permet de déterminer si deux caractéristiques sont différentes ou s'il existe une association entre elles, mais il ne permet cependant pas de déclarer qu'il n'y a pas de différence ou d'association entre les caractéristiques. On peut toutefois mentionner que l'enquête ne permet pas de détecter ou de déceler de différence ou d'association. On ne doit pas écrire :

- « Cependant lorsque les données sont ventilées selon le sexe, il n'y a aucun écart selon l'âge. »

Il faut plutôt écrire :

- « Cependant lorsque les données sont ventilées selon le sexe, on ne détecte aucun écart selon l'âge. »
- « Cependant lorsque les données sont ventilées selon le sexe, il n'y a aucun écart significatif selon l'âge. »
- « Cependant lorsque les données sont ventilées selon le sexe, il n'y a aucun écart statistique apparent selon l'âge. »

Lorsqu'on compare des proportions et que les estimations ne sont pas statistiquement différentes, on ne peut pas dire que les deux proportions sont égales même si leurs estimations sont proches.

À l'inverse, si une différence est détectée, on mentionne qu'il y a une différence ou une association. Des verbes comme « sembler », « avoir tendance » sont alors à éviter, ces expressions faisant référence à des tests non significatifs. Par contre, on peut écrire :

- « Il y a une différence selon le sexe pour telle caractéristique. »
- « Il existe une association entre telle variable et telle autre variable. »

6. Présentation de résultats sous forme de ratio

Il est parfois hasardeux de présenter les résultats sous forme d'un ratio. Bien qu'une estimation de proportion puisse être, par exemple, six fois plus grande pour une sous-population que pour une autre, l'intervalle de confiance pour ce ratio peut être très grand; à titre d'exemple, il se pourrait que ce ratio ne soit pas significativement différent de trois. Par exemple, on ne devrait pas dire :

- « Les proportions sont deux fois supérieures chez les femmes que chez les hommes (38 % vs 20 %)... »

Il faudrait plutôt dire :

- « Le pourcentage est plus élevé chez les femmes que chez les hommes (38 % vs 20 %)... »

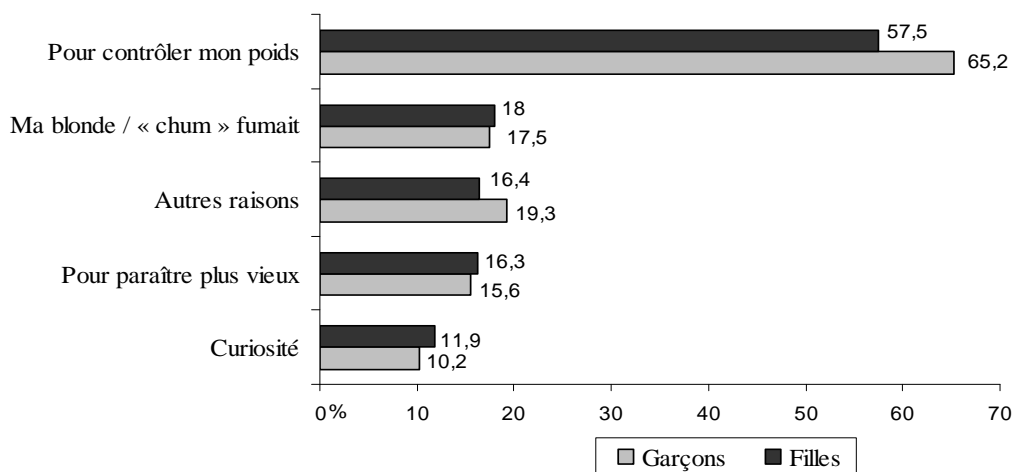
Si la borne inférieure vaut 1,5, on pourrait dire :

- « Le pourcentage est au moins 1,5 fois plus élevé chez les femmes que chez les hommes... »

7. Présentation de résultats selon l'ordre des estimations

Lorsque nous décrivons certaines caractéristiques, nous pouvons être tentés de les présenter selon l'ordre décroissant des estimations. Ordonner ainsi les résultats selon la valeur des proportions (estimations des proportions observées) peut laisser croire qu'il existe une différence significative entre chacun de ces résultats. Dès lors, si l'on souhaite présenter les résultats de cette façon, on devrait s'assurer que des tests de différences entre chacun des éléments ordonnés aient été effectués, qu'ils soient significatifs et que l'interprétation de ces tests soit adéquate. Au minimum, une identification des différences qui sont significatives ou un court texte devrait accompagner un tel tableau.

Figure 1 Raisons invoquées pour avoir commencé à fumer selon le sexe, Québec, 2008



Note : La seule différence significative entre les garçons et les filles s'observe au niveau de la raison « Pour contrôler mon poids ». Par ailleurs, cette dernière est la seule raison évoquée qui se distingue significativement des autres par sa fréquence plus élevée.

Source : Données fictives.

8. Tendance

Lorsque le seuil observé d'un test du khi-deux, servant à établir s'il existe ou non un lien entre deux variables catégoriques, est proche du seuil fixé et qu'une ou des enquêtes similaires ont déjà détecté un lien significatif entre ces mêmes variables, alors il est possible de parler de tendance. En effet, même si notre enquête n'a pu déceler un lien significatif (peut-être dû à la taille trop réduite de l'échantillon), une tendance dans le sens de la relation détectée par les autres études, peut être discutée.

De plus, dans le cadre d'un modèle statistique, on peut désirer présenter un résultat dont le seuil est légèrement plus élevé que le seuil théorique pourvu que ce résultat aille dans le sens attendu. La personne connaissant bien le sujet peut juger utile de présenter un tel résultat s'il est intéressant, en mentionnant toutefois que ce résultat n'est pas significatif.

Références bibliographiques

- BAULNE, J. et R. COURTEMANCHE (2008). *Y a-t-il vraiment des bénéfiques à l'envoi de lettres d'information dans les enquêtes à génération aléatoire de numéros de téléphone (GANT)?* Présentation au 24^e Symposium international de Statistique Canada.
- COCHRAN, W.G. (1977). *Sampling Techniques*. Third Edition. John Wiley & Sons, New York, 428 p.
- GILMORE, J. (2002). *Rapport sur l'usage de la cigarette au Canada, de 1985 à 2001* (Statistique Canada, n° 82F0077XIF au catalogue), Statistique Canada.
- GOODMAN, R., T. FORD, H. SIMMONS, R. GATWARD et coll. (2000). *Using the Strength and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample*, British Journal of Psychiatry, 177.
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2009) en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC et le MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX DU QUÉBEC, *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête sur la santé dans les collectivités canadiennes cycles 1.1, 2.1 et 3.1*, Québec, Gouvernement du Québec, 99 p.
- KORN, E.L. et B.I. GRAUBARD (1999), *Analysis of Health Surveys*, New York : John Wiley & Sons.
- LOHR, S. (1999). *Sampling: Design and Analysis*. Duxbury Press, USA.
- RUST, K.F. et J.N.K. RAO (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 5: 283-310.
- SCHENKER, N. et J. F. GENTLEMAN (2001). *On judging the significance of differences by examining the overlap between confidence intervals*, *The American statistician*, août, vol. 55, no. 3, p. 182-186.
- ST-PIERRE, M. et Y. BÉLAND (2004). *Mode effects in the Canadian Community Health Survey : a comparison of CAPI and CATI*, Proceedings of the American Statistical Association Meeting, Survey research Methods Section, Toronto: American Statistical Association, 8 p.
- TRAORÉ, I., B. BEAUVAIS, D. DU MAYS, M. DUMONT et G. MAROIS (2010). *L'Enquête québécoise sur la santé de la population, 2008. Cahier technique : Livre de codes et définition des indices (Fichier maître)*, Québec, Institut de la statistique du Québec, 704 p.
- VOSTANIS, P. (2006). Strengths and Difficulties Questionnaire: research and clinical applications. *Current opinion in psychiatry*. 19. 4.
- WOLTER, K.M. (1985). *Introduction To Variance Estimation*. Springer-Verlag, New York.

Annexe 1

Lettres d'information expédiées aux ménages correspondant aux numéros de téléphone échantillonnés pour lesquels une adresse a pu être obtenue



Février 2008

Objet : Invitation à collaborer à L'Enquête québécoise sur la santé de la population

Madame,
Monsieur,

Votre foyer a été choisi au hasard, comme 38 000 autres foyers, pour participer à L'**Enquête québécoise sur la santé de la population (EQSP)**. D'ici quelques jours, vous recevrez l'appel d'un intervieweur qualifié vous invitant à participer à cette étude d'envergure sur votre santé et votre bien-être.

L'Institut de la statistique du Québec mène cette étude à la demande du ministère de la Santé et des Services sociaux afin :

- **de mieux connaître l'état de santé de la population;**
- **d'améliorer les services de promotion de la santé et de prévention des maladies;**
- **de repérer les problèmes en émergence chez les jeunes et les moins jeunes;**
- **d'informer la population sur son état de santé.**

Des numéros de téléphone ont été choisis au hasard dans toutes les régions du Québec pour que chacune d'elles soit représentée. Les numéros sélectionnés sont les seuls qui pourront être utilisés c'est pourquoi **votre participation est très importante afin que le portrait de la santé de la population soit complet.**

L'Institut de la statistique du Québec est un organisme gouvernemental qui a pour mission de fournir des informations statistiques fiables sur la situation du Québec. Par ailleurs, conformément à l'article 25 de la *Loi sur l'Institut de la statistique du Québec*, l'Institut assure l'entière confidentialité des données dont il est dépositaire. Seuls des résultats globaux, portant sur l'ensemble des répondants, seront diffusés.

En tout temps, nous sommes prêts à répondre à vos questions au (418) 691-2404 (appels de la région de Québec) ou au 1-800-561-0213. Vous pouvez également visiter le www.stat.gouv.qc.ca/repondants pour obtenir plus d'information à propos de l'enquête.

Vous remerciant à l'avance de votre précieuse collaboration, je vous prie d'agréer, Madame, Monsieur, mes plus sincères salutations.

Le directeur général,

Stéphane Mercier

February 2008

Object: Invitation to participate in the Enquête québécoise sur la santé de la population (Québec population health survey)

Dear Madam or Sir:

(translation)

Your household was chosen at random, together with 38,000 other households, to participate in the **Enquête québécoise sur la santé de la population (EQSP)**. An interviewer will be calling you in a few days inviting you to participate in this province-wide survey on health and welfare.

The Institut de la statistique du Québec is conducting this survey at the request of the ministère de la Santé et des Services sociaux in order to:

- **gain an understanding of the health status of the population;**
- **improve health promotion and disease prevention services;**
- **identify problems in the young and the old;**
- **provide the public with information on its health status.**

Only a limited number of telephone numbers were chosen at random in all Québec regions so as to ensure the representation of each. **Therefore, we are counting on your participation for a complete portrait of the health status of the population.**

The Institut de la statistique du Québec is a government agency whose mission is to provide reliable statistics on the situation of Québec. Moreover, under section 25 of the *Act respecting the Institut de la statistique du Québec*, the Institut ensures the complete confidentiality of the data of which it is the depositary. Only the overall results on the respondents will be disseminated.

We would be pleased to answer any questions you may have at (418) 691-2404 (calls in the Québec City area) or at 1-800-561-0213. For further information on the survey, you may also visit the following website:
www.stat.gouv.qc.ca/repondants.

We thank you in advance for your invaluable cooperation.

Yours truly,

Stéphane Mercier
Director General

Octobre 2008

**Objet : Invitation à collaborer à l'Enquête québécoise sur la santé de la population
dans la région de Montréal**

Madame,
Monsieur,

Votre foyer a été choisi au hasard pour participer à l'**Enquête québécoise sur la santé de la population (EQSP)**. Dans la région de Montréal, la participation de 2000 personnes permettra à l'Agence de la santé et des services sociaux de mieux connaître l'état de santé de la population qu'elle dessert et ainsi d'adapter les services qui lui sont destinés.

D'ici quelques jours, vous recevrez l'appel d'un intervieweur qualifié vous invitant à participer à cette étude d'envergure sur votre santé et votre bien-être.

L'Institut de la statistique du Québec mène cette étude auprès de 38 000 personnes à l'échelle du Québec à la demande du ministère de la Santé et des Services sociaux afin :

- **de mieux connaître l'état de santé de la population;**
- **d'améliorer les services de promotion de la santé et de prévention des maladies;**
- **de repérer les problèmes en émergence chez les jeunes et les moins jeunes;**
- **d'informer la population sur son état de santé.**

Des numéros de téléphone ont été choisis au hasard dans toutes les régions du Québec pour que chacune d'elles soit représentée. Les numéros sélectionnés sont les seuls qui pourront être utilisés c'est pourquoi **votre participation est très importante afin que le portrait de la santé de la population soit complet.**

L'institut de la statistique du Québec est un organisme gouvernemental qui a pour mission de fournir des informations statistiques fiables sur la situation du Québec. Par ailleurs, conformément à l'article 25 de la *Loi sur l'Institut de la statistique du Québec*, l'Institut assure l'entière confidentialité des données dont il est dépositaire. Seuls des résultats globaux, portant sur l'ensemble des répondants, seront diffusés.

En tout temps, nous sommes prêts à répondre à vos questions au 418 691-2404 (appels de la région de Québec) ou au 1-800-561-0213. Vous pouvez également visiter le www.stat.gouv.qc.ca/repondants pour obtenir plus d'information à propos de l'enquête.

Vous remerciant à l'avance de votre précieuse collaboration, je vous prie d'agréer, Madame, Monsieur, mes plus sincères salutations.

Le directeur général,



Stéphane Mercier

October 2008

Subject: Invitation to participate in the Québec Population Health Survey in the Montreal region

Dear Sir or Madam:

Your household has been chosen at random to participate in the **Québec Population Health Survey (EQSP)**. In the Montreal region, the participation of 2000 people will enable the health and social services agency to better understand the health status of the population it serves and to adapt its services accordingly. A qualified interviewer will be calling you in a few days to invite you to participate in this broad survey on your health and welfare.

The Institut de la statistique du Québec (ISQ) is conducting this study of 38 000 people throughout the province at the request of the ministère de la Santé et des Services sociaux in order to:

- **gain a better understanding of the state of health of the population;**
- **improve health promotion and disease prevention services;**
- **identify emerging problems in the young and old;**
- **provide the public with information on its state of health.**

Only a limited number of telephone numbers have been chosen at random in all Québec regions so as to ensure the representation of each. Since only the telephone numbers selected can be used, **your participation is very important in ensuring that the portrait of the health of the population is complete.**

The ISQ is a government agency whose mission is to provide reliable statistics on the situation of Québec. Moreover, under section 25 of the *Act respecting the Institut de la statistique du Québec*, the ISQ ensures the complete confidentiality of the data it holds. Only general results for all respondents will be disseminated.

We are available at any time to answer your questions; please call 418-691-2404 (calls from the Québec City area) or 1-800-561-0213. For further information on the survey, you can also visit the following Website: http://www.stat.gouv.qc.ca/repondants/default_an.htm.

Thank you in advance for your invaluable cooperation.

Yours truly,

Stéphane Mercier
Director General

Annexe 2

Information supplémentaire sur les taux de réponse

Pour les enquêtes utilisant un plan d'échantillonnage complexe possédant plusieurs niveaux de sélection, par exemple le ménage et l'individu, il est important de considérer chacun de ces niveaux dans l'établissement du taux de réponse global pondéré de l'enquête.

En utilisant les codes de résultats obtenus lors de la collecte des données, il est possible de classer les unités de l'échantillon dans l'une des catégories suivantes : entrevue complétée⁷², ménage non-répondant (admissibilité non déterminée), personne sélectionnée non répondante (admissibilité confirmée)⁷³, numéro de téléphone non valide⁷⁴ et ménage inadmissible⁷⁵. À partir de ces catégories, il est ensuite possible de définir le taux de réponse ménage pondéré en utilisant les nombres suivants :

A : le nombre pondéré d'entrevues complétées

B : le nombre pondéré de cas « ménages non-répondant (admissibilité non déterminée) »

C : le nombre pondéré de cas « personne sélectionnée non-répondante (admissibilité confirmée) »

D : le nombre pondéré de numéros de téléphone non valides

E : le nombre pondéré de ménages inadmissibles

Afin de définir le taux de réponse, il est nécessaire de définir d'abord le taux d'admissibilité.

Le taux d'admissibilité pondéré des ménages renseigne sur la proportion de ménages admissibles parmi tous ceux pour lesquels l'admissibilité peut être établie. Dans le cas de certains numéros de téléphone générés, le résultat de la collecte ne permet pas d'établir précisément s'il permet de joindre un ménage privé admissible. Ceux-ci ne sont donc pas considérés dans le calcul du taux d'admissibilité qui est fait avec la formule suivante :

$$T_{admMÉNAGE} = \frac{A + C}{(A + C + E)}$$

⁷² Une entrevue complétée correspond à un ménage répondant et à un individu répondant.

⁷³ Une unité classée dans cette catégorie correspond à un ménage dont l'admissibilité est confirmée, mais dont la personne sélectionnée n'a pas répondu à l'enquête.

⁷⁴ Ce sont les numéros de téléphone qui ne sont pas en service ou qui correspondent à une ligne défectueuse pour toute la période de collecte, les numéros servant uniquement à un télécopieur et, enfin, les numéros de cellulaire.

⁷⁵ L'unité est inadmissible s'il n'y a aucune personne de 15 ans ou plus dans le ménage ou que celui-ci est localisé dans une réserve amérindienne ou bien qu'il correspond à un ménage collectif.

Enfin, le taux de réponse ménage pondéré⁷⁶, calculé ici en supposant que certains des cas dont l'admissibilité est inconnue sont en fait des cas admissibles, s'obtient par le rapport du nombre pondéré de répondants sur le nombre pondéré estimé de personnes admissibles :

$$T_{repMÉNAGE} = \frac{A + C}{[(A + C) + B * T_{admMÉNAGE}]}$$

La formule du taux de réponse individu pondéré est plus simple, soit le nombre pondéré d'individus ayant rempli le questionnaire divisé par le nombre total pondéré d'individus sélectionnés pour répondre à l'enquête.

$$T_{repINDIVIDU} = \frac{A}{(A + C)}$$

Le taux de réponse global pondéré de l'enquête s'obtient en prenant le produit des taux ménage et individu, soit

$$T_{repGLOBAL} = T_{repMÉNAGE} \times T_{repINDIVIDU}$$

⁷⁶ Dans ce calcul, on suppose que tous les numéros dont l'admissibilité est inconnue (B) sont en fait des numéros de téléphone valides. Cette définition du taux de réponse correspond à celle qui est en usage à l'ISQ. D'autres approches sont utilisées ailleurs pour définir le taux de réponse; les approches diffèrent essentiellement en ce qui concerne les hypothèses faites sur la validité et l'admissibilité des numéros de téléphone pour lesquels la collecte des données n'a pas permis de les fixer.

Annexe 3

Comment utiliser la pondération lors de la production d'estimations à partir des données de l'EQSP

Afin de sensibiliser l'utilisateur des données de l'EQSP à l'importance d'utiliser la pondération, voici un exemple faisant ressortir l'effet de la pondération sur les estimations.

La première partie présente deux tableaux produits à partir des données de l'EQSP. Le premier (tableau A3-1), affiche une estimation des fréquences pondérées de l'indicateur de rhinite allergique en été (RETE – a eu des symptômes de rhinite allergique en été) pour la région de la Côte-Nord, alors que le second (tableau A3-2) présente la version non pondérée du premier tableau. Les deuxième et troisième parties présentent respectivement la façon de pondérer les données à l'aide du logiciel SAS et la façon de pondérer les données à l'aide du logiciel SPSS.

Partie 1 – Exemples de tableaux produits à partir des données du fichier maître de l'EQSP 2008

Tableau A3-1 Proportions pondérées de l'indicateur sur les symptômes de rhinite allergique en été après répartition de la non-réponse partielle pour les proportions, région de la Côte-Nord

RETE (A eu des symptômes de rhinite allergique en été)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP
Oui	3 472	41,9	3 472	41,9
Non	4 811	58,1	8 283	100,0

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau A3-2 Proportions non-pondérées de l'indicateur sur les symptômes de rhinite allergique en été après répartition de la non-réponse partielle pour les proportions, région de la Côte-Nord

RETE (A eu des symptômes de rhinite allergique en été)	Nombre de répondants	Proportion	Nombre de répondants cumulatif	Proportion cumulative
NRP
Oui	84	37,8	84	37,8
Non	138	62,2	222	100,0

Source : Institut de la statistique du Québec, EQSP 2008.

Les tableaux A3-1 et A3-2 montrent bien l'importance d'utiliser la pondération lors de la production d'estimations à partir des données de l'EQSP, alors que la proportion des personnes ayant eu des symptômes en été passe de 37,8 % (non pondéré) à 41,9 % (pondéré); **une différence de 4,1 %**. Cette différence pourrait avoir un impact sur les conclusions tirées de ces résultats.

Partie 2 - Utilisation de la pondération à l'aide du logiciel SAS - fichier maître de l'EQSP 2008

Syntaxe en SAS

```
libname lib "...répertoire où se trouvent les données...";  
proc freq data= lib. Nom du fichier;  
tables RETE ;  
weight poids ;  
where RSS_estim=9;  
run.
```

Note : L'absence de spécification pour la pondération (option « weight ») aura pour conséquence de produire un résultat non pondéré.

Partie 3 - Utilisation de la pondération à l'aide du logiciel SPSS - fichier maître de l'EQSP 2008

Syntaxe en SPSS

Ouverture de la base de données :

Utiliser la syntaxe suivante : « get file= Emplacement\Nom du fichier.sav'. »

Pour pondérer les données :

Utiliser la syntaxe suivante : « weight by nom de la variable. »

- le nom de la variable de pondération pour les individus de 15 ans et plus est « POIDS ».

Les données sont maintenant pondérées pour effectuer les analyses.

Menus et Souris

Ouverture de la base de données :

Dans le menu, choisir « File »;

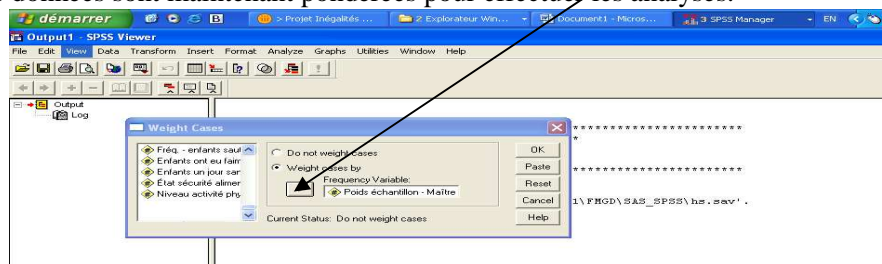
Ensuite, « Open » et « Data... ». Choisir la base de données.

Pour pondérer les données :

Dans le menu, choisir « Data » :

- ensuite, « Weight Cases... » ;
- choisir la variable de pondération dans la liste déroulante « Poids 15 ans et plus [POIDS] »;
- noircir le cercle « Weight cases by » et cliquer sur la petite flèche (la variable se déplace dans le rectangle);
- Cliquer sur « OK ».

Les données sont maintenant pondérées pour effectuer les analyses.



Annexe 4

Étape de calage aux marges dans la pondération de l'EQSP

La présente annexe fournit de l'information supplémentaire concernant l'étape d'ajustement au compte de la population, aussi appelée calage aux marges, abordée à la section 2.2.1.

La stratégie de pondération développée pour l'EQSP a pour but de permettre des inférences adéquates à la population visée tout en minimisant la variabilité des estimations. Or, il est possible d'améliorer le processus d'estimation en rajustant les poids d'échantillonnage à l'aide de données auxiliaires, lorsqu'elles sont disponibles. Le calage aux marges est une procédure qu'on peut appliquer pour incorporer des données auxiliaires.

Cette procédure rajuste les poids d'échantillonnage au moyen de multiplicateurs appelés les facteurs de calage. Ces facteurs font correspondre les estimations aux totaux connus de la population. Les poids obtenus sont appelés les poids de calage ou les poids d'estimation finaux. De façon générale, ces poids permettront de réduire la variance d'échantillonnage.

Pour la pondération du répondant de 15 ans et plus de l'EQSP, deux étapes de calage ont été effectuées; une pour l'échelle régionale et une pour l'échelle locale. La première a fait intervenir seulement une marge, alors que deux marges ont été utilisées pour la seconde.

À l'échelle régionale, les estimations de la population en 2008⁷⁷ définies pour chacun des quatorze groupes d'âge-sexe d'intérêt, c'est-à-dire les sept groupes d'âge 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65-74 ans et 75 ans et plus, pour chacun des deux sexes, ont été utilisées pour le calage aux marges. Une seule marge a donc été utilisée; celle-ci possédait 14 catégories.

En d'autres mots, le calage aux marges à l'échelle régionale a pour but d'ajuster les poids des répondants d'une même région, afin que la somme des poids, pour chacune des 14 catégories, corresponde au nombre d'individus de la population de cette catégorie dans cette région (marge 1).

Pour les six régions qui ont fait l'achat d'un échantillon supplémentaire, le calage aux marges a été appliqué à l'échelle locale (RLS) sous la contrainte de respecter les tailles de la population des régions pour chacun des groupes. Ainsi, deux marges ont été utilisées. La première correspond aux six groupes d'âge 15-24⁷⁸ ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65 ans et plus, pour chacun des deux sexes. Cette marge possède donc 12 catégories. La deuxième correspond aux tailles de la population des régions concernées pour chacune des douze catégories définies pour la première marge.

Autrement dit, le calage aux marges à l'échelle locale a pour but d'ajuster les poids des répondants d'un même RLS, afin que la somme des poids des répondants corresponde au nombre d'individus de la population dans ce RLS pour chacune des 12 catégories (marge 1). Mais également que, pour chaque catégorie, la somme des poids des répondants de tous les RLS d'une même région corresponde au nombre d'individus de la population de cette catégorie dans cette région (marge 2).

⁷⁷ Les estimations de la population en 2008 vivant en ménage privé sont basées sur les comptes du Recensement de 2006 et d'une estimation du sous-dénombrement net, de même que sur les comptes de naissance, décès, immigration et émigration depuis ce temps. Ces estimations ont été produites par la Direction des statistiques sociodémographiques de l'ISQ.

⁷⁸ À l'exception de la région 12 pour laquelle les groupes 15-24 ans et 25-34 ans chez les femmes ont été regroupés en raison d'un nombre insuffisant d'unités.

La procédure de calage aux marges utilisée pour la pondération des enfants de 3 à 14 ans est similaire à celle décrite pour les 15 ans et plus. Pour cette raison, elle ne sera pas élaborée dans cette annexe. Mentionnons uniquement que les groupes d'âge utilisés sont : 3-5 ans, 6-8 ans, 9-11 ans et 12-14 ans.

Annexe 5

Diagnostic et analyse de la non-réponse partielle pour les données de l'EQSP

1. Analyse quantitative de la non-réponse partielle

On recommande, lorsque la non-réponse partielle dépasse 5 %⁷⁹ pour l'ensemble du Québec et 10 % à l'échelle régionale ou locale, d'effectuer un examen plus approfondi de la variable étudiée.

Or, la non-réponse partielle est assez faible pour l'EQSP. Il n'est donc pas nécessaire, sauf exception, d'effectuer un examen de la non-réponse partielle avec les données de l'EQSP 2008. Toutefois, dans le cas où cet examen s'avère nécessaire, la présente annexe sert à illustrer comment procéder au diagnostic et à l'analyse de la non-réponse partielle pour des proportions lorsque celle-ci est non négligeable.

La suite de l'annexe 5, incluant les exemples, est tirée intégralement du document ISQ et collab. (2009).

Deux exemples tirés de l'ESCC 2000-2001 (cycle 1.1) seront présentés. Bien qu'au niveau régional, les risques de biais importants associés à la non-réponse partielle soient, sauf exception⁸⁰, négligeables relativement à la précision des estimations, ces deux exemples ne sont présentés que pour illustrer la démarche associée au diagnostic de la non-réponse partielle et à son impact sur les estimations de proportions.

2. Taux de non-réponse partielle simple et en fonction de la variable de croisement

Exemple 1 :

Calculons d'abord le taux de non-réponse partielle pour l'indice de l'activité physique pour la région sociosanitaire de Chaudière-Appalaches au cycle 1.1. Le tableau A5-1 illustre cet exemple.

Le taux de non-réponse partielle de l'indice de l'activité physique s'obtient directement du tableau de fréquences simples de la variable avant la répartition de la catégorie « NRP ». Le tableau A5-1 indique que le taux de non-réponse partielle de l'indice de l'activité physique pour la région Chaudière-Appalaches est de 16,66 % au cycle 1.1. Une analyse plus approfondie s'avère donc nécessaire.

⁷⁹ Pour des estimations de proportions, il peut arriver que le taux de non-réponse partielle soit inférieur à cette borne et qu'il y ait quand même un biais. Cependant, on pense qu'en général, ce biais serait alors négligeable. Par contre, pour l'étude de paramètres correspondant à des totaux, la situation est différente et des corrections devraient être apportées pour produire des estimations non biaisées quel que soit le taux de non-réponse partielle.

⁸⁰ Par exemple, au cycle 1.1, la variable « nombre d'heures passées devant l'ordinateur » (SACA_1R). Lorsque croisée avec l'âge, une relation très forte est détectée dans la région Chaudière-Appalaches, les jeunes étant proportionnellement moins nombreux que les plus âgés à ne pas utiliser l'ordinateur. Or, le taux de non-réponse partielle est plus élevé chez les jeunes que chez les plus âgés. On surestime donc la proportion de personnes de 12 ans ou plus non-utilisatrices de l'ordinateur. Le biais, causé par cette surestimation, n'est pas négligeable lorsque comparé à l'erreur échantillonnale. Dans ce cas-ci, des mises en garde sont nécessaires.

Tableau A5-1 Fréquences et proportions pondérées de l'indice de l'activité physique avant répartition de la non-réponse partielle pour la région de Chaudière-Appalaches, cycle 1.1 de l'ESCC, Québec

PACADPAI (Indice de l'activité physique)	Fréquence	Pourcentage	Fréquence cumulative	Pourcentage cumulatif
NRP	55 142,31	16,66	55 142,31	16,66
Actif	34 438,16	10,41	89 580,47	27,07
Modéré	60 499,80	18,28	150 080,27	45,34
Inactif	180 895,40	54,66	330 975,67	100,00

Source : Statistique Canada, Fichier de microdonnées à grande diffusion de l'ESCC 1.1 (2000-2001).

Exemple 2 :

En général, pour faire le diagnostic concernant la non-réponse partielle, on vérifie si le taux de non-réponse partielle est uniforme selon l'âge, le sexe, la scolarité ou toute autre variable de croisement jugée pertinente. Dans notre exemple, nous voulons savoir si le taux de non-réponse partielle diffère ou non de 16,66 % selon les catégories de la variable « sexe » puisque cette variable est jugée très pertinente.

Le tableau A5-2 présente la sortie SAS du croisement de la variable d'analyse PACADPAI (indice de l'activité physique) avec la variable DHHA_SEX (sexe), pour la région sociosanitaire de Chaudière-Appalaches au cycle 1.1. Celui-ci comporte des fréquences, des proportions et de l'information au sujet du test d'indépendance. Il semble en effet que la variable « sexe » soit très associée à l'indice de l'activité physique comme en témoigne le test du khi-deux significatif ($p=0,0038$). On remarque que le tableau croisé est accompagné d'un message d'avertissement (ATTENTION) concordant à la proportion de données manquantes. Il se peut que le nombre de « NRP » pour la variable à l'étude diffère de la valeur de « Fréquence manquante », ce qui n'est pas le cas pour notre exemple. Cela se produit lorsqu'il y a non-déclaration pour les deux variables croisées dans le tableau. C'est pourquoi on réfère plutôt au tableau de fréquences simples pour établir le taux de non-réponse partielle.

Tableau A5-2 Fréquences pondérées de l'indice de l'activité physique APRÈS répartition de la non-réponse partielle pour la région de Chaudière-Appalaches, cycle 1.1 de l'ESCC, Québec

PACADPAI (Indice de l'activité physique)	DHHA_SEX (sexe)		Fréquence cumulative
Fréquence Khi-2 cell. Pourcentage Pct. en ligne Pct. en col.	Homme	Femme	Total
NRP	98.632 . . .	59.865
Actif	58.224 2.5917 7.34 58.82 15.41	40.762 2.3591 5.14 41.18 9.82	98.987 12.49
Modéré	93.43 1.3476 11.78 53.73 24.73	80.466 1.2266 10.15 46.27 19.39	173.9 21.93
Inactif	226.14 1.8872 28.52 43.49 59.86	293.82 1.7178 37.06 56.51 70.79	519.95 65.58
Total	377.792 47.65	415.044 52.35	792.836 100.00
Fréquence manquante = 158.49718271			

Statistique	DF	Valeur	Proba.
Khi-2	2	11.1301	0.0038

Taille réelle de l'échantillon = 792.83615062

Fréquence manquante = 158.49718271

ATTENTION : 17 % des données sont manquantes.

Source : Statistique Canada, Fichier de microdonnées à grande diffusion de l'ESCC 1.1 (2000-2001).

Comme on peut le noter, la seule information fournie par le logiciel statistique SAS concernant les « NRP » correspond aux fréquences observées dans les différents niveaux de la variable de croisement. On doit donc calculer les taux de non-réponse partielle par catégorie de la variable de croisement en utilisant ce nombre de « NRP ». On pourra ensuite faire la comparaison de ces taux avec le taux de non-réponse partielle associée à la variable étudiée. Le détail des calculs est présenté à la fin de cette annexe. Le tableau A5-3 montre les résultats obtenus pour chacune des catégories de la variable sexe.

Tableau A5-3 Exemple de taux de non-réponse partielle par catégorie de la variable de croisement

Sexe	Non-répondants (NR)	Répondants potentiels (NTOT)	Taux de non-réponse partielle (%)
Homme	98,632	476,424	20,7
Femme	59,865	474,909	12,6

Source : Statistique Canada, Fichier de microdonnées à grande diffusion de l'ESCC 1.1 (2000-2001).

On constate dans ce tableau que le taux de non-réponse partielle associée à l'indice de l'activité physique est beaucoup plus important dans la sous-population des hommes (20,7 %) que dans la population générale de 12 ans ou plus (16,66 %).

3. Biais potentiels et mises en garde

En général, si les taux de non-réponse partielle selon l'âge, le sexe, la scolarité, le fait d'avoir ou non une incapacité à travailler lors de la dernière semaine ou tout autre variable pertinente diffèrent, on doit faire un examen des biais potentiels et faire les mises en garde qui s'imposent. On fera alors l'hypothèse que les non-répondants ayant des caractéristiques semblables aux répondants auraient fourni des réponses semblables.

Dans notre exemple, nous supposons ici que les non-répondants de sexe masculin auraient un niveau actif à l'indice de l'activité physique plus élevé que les non-répondantes féminines. C'est ce qu'ont répondu les hommes de 12 ans ou plus en comparaison aux femmes.

Dans un premier temps, on peut identifier les taux de non-réponse partielle par catégorie de la variable de croisement qui diffèrent le plus du taux associé à la variable étudiée. On peut ensuite se faire une idée des biais introduits par la non-réponse partielle sur les estimations en s'intéressant à l'effet qu'aurait eu la présence des non-répondants sur les proportions estimées.

Dans notre exemple, les deux variables étudiées sont assez fortement associées. De façon générale, on estime que l'indice de l'activité physique est associé au sexe, les hommes étant proportionnellement plus nombreux que les femmes à se situer au niveau actif de l'indice de l'activité physique. Par ailleurs, le taux de non-réponse partielle pour cet indice vaut 16,66 % et ce taux s'élève respectivement à 20,7 % et 12,6 %, selon qu'on est de sexe masculin ou féminin.

Par conséquent, il serait justifié d'inclure une mise en garde à nos analyses qui indique le sens du biais (sous ou surestimation). Ainsi, pour notre exemple, on indiquerait que la proportion d'actifs est légèrement sous-estimée.

Détail des calculs des résultats présentés au tableau A5-3 :

Pour les hommes :

$$\begin{aligned} \text{NR} &= 98,632 \\ \text{NTOT} &= 98,632 + 377,792 = 476,424 \\ \% &= 98,632 / 476,424 = \mathbf{20,7 \%} \end{aligned}$$

Pour les femmes :

$$\begin{aligned} \text{NR} &= 59,865 \\ \text{NTOT} &= 59,865 + 415,044 = 474,909 \\ \% &= 59,865 / 474,909 = \mathbf{12,6 \%} \end{aligned}$$

Annexe 6

Information supplémentaire – imputation des questions 54 et 120 de l'EQSP

Imputation de la question 54

En raison d'une erreur dans les conditions menant au passage à la question 54, certains répondants n'ont pas été interrogés sur la question 54. Cette question porte sur les troubles musculo-squelettiques reliés à l'emploi principal actuel. Elle visait à savoir si les douleurs au cou, au dos et aux membres inférieurs ou supérieurs ressenties par les répondants étaient liées à leur travail principal actuel.

Rappelons que dans le cas de la question 54, certains répondants ne se sont tout simplement pas fait poser la question, malgré le fait qu'ils auraient dû y répondre. Les individus qui ont subi ce saut de question injustifié ont le profil suivant : ils ont déclaré une douleur à au moins deux des quatre parties du corps couvertes par la question 54 et ont déclaré qu'au moins l'une d'entre elles n'était pas reliée au travail.

Autrement dit, pour les individus correspondants à ce profil, nous avons l'information selon laquelle leur douleur est reliée à leur travail actuel ou antérieur (Q53), mais nous ne savons pas s'il s'agit, en particulier, de leur emploi principal actuel (Q54). Les répondants de tous les autres profils ont été administrés adéquatement dans le questionnaire.

Afin de minimiser le risque de biais engendré par la non-réponse partielle générée par cette erreur de programmation, l'ISQ a décidé de procéder à l'imputation de valeur pour ces quatre questions.

Les paragraphes qui suivent fournissent, très succinctement, de l'information sur l'imputation de la question 54.

Les taux pondérés de non-réponse partielle aux questions Q54A à Q54D se situent entre 22,0 % et 25,6 %, dont entre 17,9 % et 22,0 % provient de l'erreur de programmation. Les variables Q54A à Q54D ont été imputées séparément à l'aide de l'imputation aléatoire par donneur (« Hot deck ») à l'intérieur de classes d'imputation. L'algorithme de classification CHAID a été employé pour construire ces classes d'imputation.

- Les classes d'imputation retenues ont été créées à partir des variables suivantes :

- Pour la Q54A : le nombre d'emplois (Q41), l'opinion du répondant sur le fait qu'il doit effectuer ou non des gestes répétitifs au travail (Q49A) et l'âge du répondant;
- Pour la Q54B : le nombre d'emplois (Q41), l'opinion du répondant sur le fait qu'il doit manier ou non des charges lourdes au travail (Q49C) et l'opinion du répondant sur le fait qu'il doit effectuer ou non une quantité excessive de travail (Q48D);
- Pour la Q54C : le nombre d'emplois (Q41), l'opinion du répondant sur le fait qu'il doit effectuer ou non des gestes répétitifs au travail (Q49A), l'opinion du répondant sur le fait qu'il doit manier ou non des charges lourdes au travail (Q49C) et l'opinion du répondant sur le fait qu'il doit effectuer ou non une quantité excessive de travail (Q48D);
- Pour la Q54D : le nombre d'emplois (Q41), l'opinion du répondant sur le fait qu'il doit manier ou non des charges lourdes au travail (Q49C), l'opinion du répondant sur le fait qu'il

doive travailler ou non en position debout (Q50A) et l'opinion du répondant sur le fait qu'il doive effectuer ou non une quantité excessive de travail (Q48D).

La variable Q41 portant sur le nombre d'emplois a été incluse par défaut afin de tenir compte du fait que certaines questions faisaient référence à l'ensemble des emplois du répondant, alors que d'autres portaient uniquement sur l'emploi principal. En principe, les questions portant sur l'emploi principal actuel devraient être davantage liées aux questions 54A à 54D pour les répondants n'ayant qu'un seul emploi.

Les questions 54A à 54D servent principalement à créer un indicateur de la présence de traumatisme lié au travail principal actuel du répondant. Le taux pondéré de non-réponse partielle pour cet indicateur est plus faible que les taux des questions 54A à 54D puisqu'il est calculé sur l'ensemble des travailleurs. Ce taux est de 15,7 %; l'erreur de programmation comptant pour 13,6 %.

Les estimations de proportion obtenues pour les questions 54A à 54D avant et après imputation diffèrent au maximum de 0,8 %. Toutefois, au niveau de l'indicateur, on note une différence importante entre la proportion des travailleurs ayant un traumatisme lié au travail estimée avant et après imputation. En effet, les non-répondants aux questions 54A à 54D sont des gens ayant un traumatisme lié au travail actuel ou antérieur, et ce traumatisme est fort probablement causé par le travail principal actuel pour la majorité d'entre eux. Ainsi, l'estimation pondérée de la proportion de travailleurs ayant un traumatisme lié au travail passe de 40,2 % à 47,3 % après imputation.

Imputation de la question 120

La question 120 du questionnaire de l'enquête demandait d'estimer, pour les douze derniers mois, le revenu total de tous les membres du ménage provenant de toutes les sources, avant impôt et autres retenues. Le répondant avait la possibilité de fournir cette estimation soit par une valeur exacte, bien qu'approximative, soit sous forme de catégorie en indiquant la tranche de revenu⁸¹, dans laquelle elle se trouve.

Rappelons que sur les 38 154 répondants de l'EQSP, 4 411 ont refusé ou n'étaient pas en mesure de fournir le revenu du ménage en 2008, soit une proportion pondérée de 14,4 %. Par ailleurs, 12 190 ont déclaré le revenu du ménage sous forme de catégorie, ce qui correspond à une proportion de 32,3 %. En présence d'une non-réponse partielle aussi élevée, il est souhaitable de procéder à l'imputation de données de manière à minimiser les risques de biais occasionnés par une non-réponse caractérisée. L'ISQ a donc décidé de procéder à l'imputation du revenu total du ménage pour les non-répondants partiels de l'EQSP.

Les paragraphes qui suivent fournissent, très succinctement, l'information sur l'imputation du revenu du ménage.

Tout d'abord, mentionnons que le revenu du ménage des répondants en tant que variable continue est nécessaire afin de créer un indice du niveau de revenu. Comme mentionné précédemment, 32,3 % des répondants n'ont pas fourni le revenu exact de leur ménage, mais ont tout de même déclaré le revenu en catégorie, alors que 14,4 % n'ont donné aucune information sur le revenu de leur ménage. Le taux pondéré de non-réponse pour le revenu exact du ménage s'élève donc à 46,7 %.

La non-réponse partielle à la question sur le revenu est très présente chez les jeunes de moins de 25 ans. Entre autres, seulement 12,2 % des jeunes de moins de 18 ans ont été en mesure de fournir le revenu

⁸¹ Les tranches de revenus proposées sont les suivantes : moins de 10 000 \$, 10 000 \$ à 19 999 \$, 20 000 \$ à 29 999 \$, 30 000 \$ à 39 999 \$, 40 000 \$ à 49 999 \$, 50 000 \$ à 59 999 \$, 60 000 \$ à 79 999 \$, 80 000 \$ à 99 999 \$, 100 000 \$ et plus.

exact du ménage. Il semble donc que le revenu du ménage est inconnu de la majorité des jeunes et que cette information est très difficile à fournir pour eux.

Pour les répondants ayant fourni le revenu uniquement en catégorie, le revenu exact a été imputé séparément par catégorie de revenu à l'aide de l'imputation aléatoire par donneur (« Hot deck ») à l'intérieur de classes d'imputation. Le fait que le donneur soit dans la même catégorie de revenu que le receveur permet de s'assurer que le revenu exact imputé se situe à l'intérieur de la catégorie de revenu déclarée par le répondant.

Pour chaque catégorie de revenu, des classes d'imputation ont été constituées en employant des caractéristiques des répondants. Pour cette étape, les individus non-répondants au revenu de moins de 20 ans ont été traités séparément de ceux de 20 ans et plus. En effet, pour les moins de 20 ans, seules les caractéristiques du ménage ont servi à créer les classes d'imputation⁸², alors que pour les 20 ans et plus, les caractéristiques individuelles du répondant ont également été utilisées. Chaque classe d'imputation devait contenir un minimum de 50 donneurs pour être retenue.

Pour une catégorie de revenu donnée, l'imputation du revenu chez les individus de moins de 20 ans, a été effectuée à partir des classes d'imputation créées pour ce groupe; tous les répondants, peu importe leur âge, étaient considérés comme des donneurs.

Concernant l'imputation chez les individus de 20 ans et plus : lorsque les classes d'imputation retenues ne faisaient pas intervenir de caractéristiques individuelles du répondant, alors les donneurs potentiels étaient tous les répondants, peu importe leur âge⁸³. Par contre, lorsque des caractéristiques individuelles étaient retenues dans le modèle pour déterminer les classes d'imputation, alors seuls les répondants de 20 ans et plus étaient considérés comme donneurs potentiels⁸⁴.

Les classes d'imputation ont été construites en employant les variables étant le plus fortement liées significativement au revenu du ménage. Les classes d'imputation retenues ont été créées à partir des variables suivantes :

- Pour la catégorie « Moins de 10 000 \$ » :
 - Pour les moins de 20 ans : le nombre de personnes de 15 ans et plus dans le ménage
 - Pour les 20 ans et plus : l'âge du répondant
- Pour la catégorie « De 10 000 \$ à moins de 20 000 \$ » :
 - Pour les moins de 20 ans : la perception du répondant de la situation financière de son ménage par rapport aux gens de son âge (Q116)
 - Pour les 20 ans et plus : l'âge du répondant et le fait que le répondant occupe un emploi rémunéré ou non (Q40)
- Pour la catégorie « De 20 000 \$ à moins de 30 000 \$ » :
 - Pour les moins de 20 ans : le nombre de personnes de 20 ans et plus dans le ménage
 - Pour les 20 ans et plus : la scolarité du répondant (Q114) et le fait que le répondant occupe un emploi rémunéré ou non (Q40)
- Pour la catégorie « De 30 000 \$ à moins de 40 000 \$ » :
 - Pour les moins de 20 ans : le nombre de personnes de 20 ans et plus dans le ménage

⁸² Pour l'imputation du revenu chez les individus de moins de 20 ans, tous les répondants, peu importe leur âge, étaient considérés comme des donneurs.

⁸³ Le modèle est, dans ce cas, identique à celui utilisé pour l'imputation chez les moins de 20 ans.

⁸⁴ Pour les jeunes de moins de 20 ans, il a été jugé que les caractéristiques individuelles des répondants n'avaient pas de lien direct avec le revenu du ménage.

- Pour les 20 ans et plus : la scolarité du répondant (Q114) et le fait que le répondant occupe un emploi rémunéré ou non (Q40)
- Pour la catégorie « De 40 000 \$ à moins de 50 000 \$ » :
 - Pour l'ensemble des répondants : aucune variable additionnelle
- Pour la catégorie « De 50 000 \$ à moins de 60 000 \$ » :
 - Pour les moins de 20 ans : aucune variable additionnelle
 - Pour les 20 ans et plus : la scolarité du répondant (Q114)
- Pour la catégorie « De 60 000 \$ à moins de 80 000 \$ » :
 - Pour l'ensemble des répondants : la perception du répondant de la situation financière de son ménage par rapport aux gens de son âge (Q116)
- Pour la catégorie « De 80 000 \$ à moins de 100 000 \$ » :
 - Pour l'ensemble des répondants : le nombre de personnes de 15 ans et plus dans le ménage
- Pour la catégorie « 100 000 \$ et plus » :
 - Pour les moins de 20 ans : la perception du répondant de la situation financière de son ménage par rapport aux gens de son âge (Q116) et la région sociosanitaire (RSS)
 - Pour les 20 ans et plus : la scolarité du répondant (Q114) et l'âge du répondant

Pour les répondants n'ayant fourni aucune information concernant le revenu du ménage, une valeur exacte a été estimée à l'aide d'une régression linéaire. Trois modèles ont été employés, soit un pour les répondants de 20 ans et plus occupant un emploi rémunéré, un pour les répondants 20 ans et plus n'occupant pas d'emploi rémunéré et finalement un pour les répondants de moins de 20 ans. La décision de séparer les répondants occupant un emploi rémunéré de ceux n'occupant pas d'emploi rémunéré avait pour but d'incorporer les caractéristiques liées à l'emploi du répondant pour le premier groupe. Encore une fois, les répondants de moins de 20 ans ont été imputés à partir d'un modèle basé sur des caractéristiques du ménage uniquement dont les paramètres sont estimés à partir de l'ensemble des répondants tous âges confondus. Les variables explicatives retenues dans les différents modèles sont les suivantes :

- Pour les répondants de 20 ans et plus occupant un emploi rémunéré :
 - la région sociosanitaire (RSS)
 - le sexe
 - l'âge
 - la classification nationale des professions (CNP) correspondant à l'emploi principal du répondant
 - le nombre de personnes de 20 ans et plus dans le ménage
 - l'occupation principale du répondant (Q115)
 - la perception du répondant sur la situation financière de son ménage par rapport aux gens de son âge (Q116)
- Pour les répondants de 20 ans et plus n'occupant pas un emploi rémunéré :
 - la région sociosanitaire (RSS)
 - le sexe
 - l'âge
 - le nombre de personnes de 20 ans et plus dans le ménage
 - la scolarité du répondant (Q114)
 - la perception du répondant sur la situation financière de son ménage par rapport aux gens de son âge (Q116)

- Pour les répondants de moins de 20 ans :
 - la région sociosanitaire (RSS)
 - le nombre de personnes de 20 ans et plus du ménage
 - la perception du répondant sur la situation financière de son ménage par rapport aux gens de son âge (Q116)

Le revenu moyen estimé après imputation diminue de 5 237 \$.

Annexe 7

Exemples de calcul pour l'estimation d'un effectif

Estimation d'un effectif en présence de non-réponse partielle

Tableau A7-1 Effectif et proportion pondérés de l'indicateur MARI sur la consommation de marijuana / cannabis / haschich, Québec

MARI (A déjà pris ou essayé de la marijuana, du cannabis ou du haschich)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP	152 088	2,4	152 088	2,4
Oui	2 711 013	42,9	2 863 101	45,3
Non	3 463 422	54,7	6 326 523	100,0

Source : Institut de la statistique du Québec, EQSP 2008.

Tableau A7-2 Effectif et proportion pondérés de l'indicateur MARI sur la consommation de marijuana / cannabis / haschich (non-réponse répartie pour la proportion seulement), Québec

MARI (A déjà pris ou essayé de la marijuana, du cannabis ou du haschich)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP
Oui	2 711 013	43,9	2 711 013	43,9
Non	3 463 422	56,1	6 174 435	100,0

* Les valeurs présentées dans cette colonne ne sont pas des estimations sans biais de l'effectif de population, car la non-réponse partielle n'a pas été intégrée adéquatement aux calculs. Ces valeurs ne sont présentées que pour illustrer le traitement recommandé pour l'estimation des proportions. Pour une estimation sans biais de l'effectif, consulter la section 2.6 (tableau 2-6).

Source : Institut de la statistique du Québec, EQSP 2008.

Détails des calculs pour la catégorie « A déjà pris ou essayé de la marijuana, du cannabis ou du haschich » de l'exemple présenté à la section 2.6 :

Effectif biaisé de la catégorie « A déjà pris ou essayé de la marijuana, du cannabis ou du haschich » de la question d'analyse Q69		2 711 013
Divisé par	/	
Taux de réponse de l'indicateur MARI pour l'ensemble du Québec (1 – taux de non-réponse partielle)		97,596 % (1 – 2,404 %)
=	=	
Estimation non biaisée du nombre de personnes au Québec qui ont déjà pris ou essayé de la marijuana, du cannabis ou du haschich		2 777 791

Conclusion :

Le nombre de personnes au Québec qui ont déjà pris ou essayé de la marijuana, du cannabis ou du haschich est estimé à **2 777 800** (voir tableau A7-3⁸⁵).

Tableau A7-3 Proportion de la population ayant consommé du cannabis (marijuana ou haschich) au cours de la vie pour l'ensemble du Québec, EQSP 2008

Statistiques					
Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
2 777 800	43,9	1,0	0,43	(43,1 - 44,8)	2,4

Source : Infocentre de santé publique, octobre 2010.

Estimation d'un effectif en présence de non-réponse partielle « en cascade »

Question filtre (Q63) :

Avez-vous déjà sérieusement songé à vous suicider ou à vous enlever la vie?

Tableau A7-4 Effectif et proportions pondérées de la question 63 portant sur les idées suicidaires et les tentatives de suicide à vie, Québec

Q63 (Avez-vous déjà sérieusement songé à vous suicider ou à vous enlever la vie?)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP	149 630	.	149 630	.
Oui	877 165	14,2	1 026 795	14,2
Non	5 299 728	85,8	6 326 523	100,0

* La fréquence de la catégorie « NRP » n'a pas été redistribuée afin de présenter les valeurs à redistribuer.

Source : Institut de la statistique du Québec, EQSP 2008.

Question d'analyse (Q64) :

Est-ce que cela s'est produit au cours des 12 derniers mois?

Tableau A7-5 Effectif et proportions pondérées de la question 64 portant sur les idées suicidaires et les tentatives de suicide au cours des 12 derniers mois, Québec

Q64 (Est-ce que cela s'est produit au cours des 12 derniers mois?)	Effectif *	Proportion	Effectif * cumulatif	Proportion cumulative
NRP	2 132	.	2 132	.
Oui	171 209	19,6	173 341	19,6
Non	703 824	80,4	877 165	100,0

* La fréquence de la catégorie « NRP » n'a pas été redistribuée afin de présenter les valeurs à redistribuer.

Source : Institut de la statistique du Québec, EQSP 2008.

⁸⁵ Notez que les effectifs présentés sur le portail de l'Infocentre sont arrondis à la centaine.

Tableau A7-6 Proportions de la population qui a déjà songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois pour l'ensemble du Québec, EQSP 2008

Statistiques					
Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
175 800	2,8	5,5	0,15	(2,5 - 3,1)	2,6

Source : Infocentre de santé publique, octobre 2010

Détails des calculs pour la catégorie « A songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois » de l'exemple présenté à la section 2.6 :

Effectif biaisé de la catégorie « A songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois » de la question d'analyse Q64	171 209
+	+
Effectif estimé des personnes n'ayant pas répondu à Q64 (« NRP » de la question d'analyse seulement) qui auraient songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois	19,6 % de 2 132
+	+
Effectif estimé des personnes n'ayant pas répondu à Q63, mais qui auraient songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois (« NRP » de la question filtre qui aurait dû répondre à la question 64)	19,6 % de (14,2 % de 149 630)
=	=
Estimation non biaisée du nombre de personnes au Québec qui ont songé sérieusement au suicide ou qui ont tenté de se suicider au cours des 12 derniers mois	175 791

Conclusion :

Le nombre de personnes au Québec qui ont songé sérieusement au suicide ou qui ont tenté de se suicider au cours des 12 derniers mois est estimé à **175 800** (voir tableau A7-6⁸⁶).

⁸⁶ Notez que les effectifs présentés sur le portail de l'Infocentre sont arrondis à la centaine.

Annexe 8

Information supplémentaire concernant la méthode bootstrap

Dans l'EQSP 2008, la méthode bootstrap est utilisée à deux fins. Premièrement, pour l'estimation de la variance des estimations, pour laquelle 500 ensembles de poids bootstrap sont utilisés afin de produire 500 estimations ponctuelles du paramètre d'intérêt. La variance bootstrap est ensuite calculée en mesurant la variabilité de ces estimations. Deuxièmement, pour produire des estimations dites « par intervalle » des estimations ponctuelles, c'est-à-dire pour calculer des intervalles de confiance.

Seule l'utilisation première est abordée dans cette annexe. Le lecteur est invité à consulter la section 3.3.3 pour plus d'information sur les intervalles de confiance bootstrap utilisés dans l'EQSP 2008.

Voici en bref, en quoi consiste la méthode par ré-échantillonnage du bootstrap⁸⁷ : la méthode du bootstrap consiste à tirer indépendamment plusieurs sous-échantillons à partir de l'échantillon initial incluant les non-répondants, les inadmissibles et les unités non valides. Le nombre de sous-échantillons varie d'une enquête à l'autre selon les besoins et objectifs de celles-ci. Chaque sous-échantillon est sélectionné indépendamment et est le résultat d'un échantillonnage aléatoire simple (ÉAS) avec remise de $n-1$ grappes, parmi les n grappes présentes à l'intérieur de chaque strate de l'échantillon principal. Pour chacun des sous-échantillons ainsi obtenus, de nouvelles probabilités de sélection sont calculées pour chaque unité du sous-échantillon afin de tenir compte de la possibilité de sélectionner plusieurs fois une même unité dans l'échantillon bootstrap (sélection avec remise). Par la suite, l'ensemble des ajustements décrit à la section 2 sur la pondération est appliqué à chaque sous-échantillon, dont l'ajustement aux comptes de la population (calage aux marges), afin d'en arriver à la création d'un poids bootstrap. Dans le cas de l'EQSP 2008, 500 ensembles de poids bootstrap, tous indépendants et tenant compte de la non-réponse du sous-échantillon, ont été créés pour l'estimation de la précision des données.

Pour obtenir l'estimation de la variance d'une estimation ponctuelle (statistique calculée à partir du poids de sondage) à l'aide de la méthode bootstrap, il suffit de recalculer cette même estimation ponctuelle 500 fois en utilisant les 500 ensembles de poids bootstrap. La variabilité observée entre les 500 résultats représente l'estimation de la variance.

Du point de vue de la pratique, voici les principales étapes de l'estimation de la précision d'une estimation donnée à l'aide de la méthode du bootstrap :

- a) L'estimation du paramètre d'intérêt (proportion, total, etc.) est d'abord calculée en utilisant le poids de sondage inclus dans le fichier de données.
- b) La même estimation ponctuelle est calculée en utilisant cette fois chacun des 500 ensembles de poids bootstrap contenus dans le fichier. Ainsi, cinq cents estimations bootstrap, par exemple de la proportion, sont ainsi obtenues.

⁸⁷ La littérature contient un bon nombre de documents décrivant ou évaluant la méthode du bootstrap dans un contexte d'utilisation avec des données provenant d'enquêtes à plans complexes. Parmi ceux-ci, mentionnons seulement les ouvrages de Lohr (1999), Rust et Rao (1996) ou Wolter (1985).

- c) Finalement, la variance (selon la formule utilisée pour un plan aléatoire simple) de ces 500 estimations bootstrap est calculée. Cette variance correspond à l'estimation de la variance de l'estimation du paramètre d'intérêt calculée en a).

Annexe 9

Intervalle de confiance binomial

Lorsque l'approximation par la loi Normale n'est pas une bonne approximation de la loi de l'estimateur, c'est-à-dire lorsque la proportion est trop petite, compte tenu de la taille d'échantillon efficace⁸⁸ associée, l'intervalle de confiance basé sur la loi Normale doit être remplacé par un intervalle de confiance asymétrique basé sur la loi Binomiale.

Les détails de la construction des bornes de l'intervalle Binomial ne sont pas présentés dans ce document⁸⁹. Cependant, les conditions d'application de la loi Normale pour une proportion et une taille d'échantillon efficace minimale sont présentées dans le tableau suivant :

Tableau des conditions d'application de la loi Normale⁹⁰

Proportion	Taille efficace minimale
$p < 0,05$ ou $p > 0,95$	10 000
$0,05 \leq p < 0,10$ ou $0,90 < p \leq 0,95$	1 400
$0,10 \leq p < 0,20$ ou $0,80 < p \leq 0,90$	600
$0,20 \leq p < 0,30$ ou $0,70 < p \leq 0,80$	200
$0,30 \leq p < 0,40$ ou $0,60 < p \leq 0,70$	80
$0,40 \leq p < 0,50$ ou $0,50 < p \leq 0,60$	50
$p = 0,5$	30

⁸⁸ La taille efficace d'un échantillon correspond à la taille de cet échantillon divisé par l'effet du plan de sondage. Par exemple, un effet de plan égal à 2 associé à un échantillon de 20 000 répondants sélectionnés à partir d'un plan de sondage complexe fournit une précision identique à celle qu'on obtiendrait en sélectionnant un échantillon de taille 10 000 à l'aide d'un plan de sondage simple (échantillon aléatoire simple). On dit alors que la taille d'échantillon efficace est de 10 000.

⁸⁹ Le lecteur qui y est intéressé peut consulter Korn et Graubard (1998).

⁹⁰ Adaptation du tableau 3.3 de la page 58 de Cochran (1977).

Annexe 10

Niveau de confiance

Les niveaux de confiance les plus couramment utilisés sont les suivants : 99,9 %, 99 % et 95 %. Puisque les intervalles de confiance sont fréquemment utilisés pour réaliser des tests de comparaison de proportions, le choix du niveau de confiance à utiliser est étroitement associé au seuil de signification de ces tests. En fait, le niveau de confiance des intervalles produits correspond à $1-\alpha$, où α est le seuil du test de comparaison. Fixer ce seuil équivaut à choisir le niveau de confiance.

Au moment de construire un test de comparaison, on se préoccupe de sa puissance, c'est-à-dire sa capacité à détecter une différence entre les proportions, de manière à identifier les caractéristiques présentant des disparités significatives. Incidemment, plus le nombre de répondants est élevé, plus le test de comparaison est puissant, c'est-à-dire qu'il permet de détecter de petites différences entre les proportions. En fait, avec un test très puissant, plusieurs différences sont déclarées significatives, ce qui facilite d'autant l'interprétation des données pour les utilisateurs et admet des conclusions fortes.

Le contrôle de l'erreur de type I du test, c'est-à-dire le rejet de l'hypothèse d'égalité entre deux proportions alors qu'elles sont en fait égales, importe aussi dans la construction du test. Le seuil fixé du test (α) correspond en réalité au risque d'erreur (de type I) maximal qu'on est prêt à courir. Par exemple, si l'on fixe le seuil d'un test de comparaison à 5 %, c'est qu'on est prêt à risquer de déclarer les proportions différentes par erreur dans au plus 5 % des cas.

En pratique, lorsqu'on construit un test, on choisit de contrôler l'erreur de type I en fixant un seuil acceptable, tout en prenant en considération la puissance du test statistique. Plus la puissance du test est élevée, moins on acceptera de prendre un risque de commettre une erreur de type I. En effet, lorsqu'on abaisse le seuil du test, on réduit le nombre de différences qui seront détectées par le test. Par contre, on réduit également les chances de faire une erreur en prenant une mauvaise décision (c'est-à-dire de conclure que les proportions sont différentes alors qu'elles sont égales). Or, en présence d'un test puissant, la perte attribuable à la diminution du seuil sera compensée par la puissance élevée du test.

Annexe 11

Exemple de calcul d'une proportion ajustée selon l'âge, et dans le cas où l'utilisation des poids bootstrap n'est pas possible, de sa variance

Calcul de la proportion ajustée selon la méthode de la standardisation directe

Exemple fictif pour le calcul d'une proportion de la population se percevant en bonne santé.

Groupe d'âge	Proportion brute (A)	Erreur-type (B)	Population de référence, ensemble du Québec, sexes réunis, 2001 ¹ (C)	Poids d'ajustement (D)	Proportion ajustée (E)	Variance de la proportion ajustée (F)
15-24 ans	0,942	0,0192695	983 107	0,16140237	0,152041032	9,6730E-06
25-44 ans	0,977	0,0102453	2 253 498	0,36996982	0,361460512	1,4368E-05
45-64 ans	0,858	0,0221330	1 890 083	0,31030587	0,266242439	4,7169E-05
65-74 ans	0,762	0,0523700	549 361	0,09019178	0,068726134	2,2310E-05
75 ans et plus	0,659	0,0654112	414 983	0,06813016	0,044897777	1,9860E-05
Total	0,877	0,0117626	6 091 032		0,89336789	0,00011338

¹: Source : Ministère de la Santé et des Services sociaux, Service du développement de l'information, janvier 2010

Étape 1 : Calcul des proportions brutes (A) et de leurs erreur-type (B) pour chacun des groupes d'âge à partir du fichier de données de l'EQSP.

Étape 2 : Calcul des poids d'ajustement (D) = population de référence par groupe d'âge (C) divisé par la population totale de référence (C : Total).

Exemple : 25-44 ans, (D) = 2 253 498 ÷ 6 091 032 = 0,36996982.

Étape 3 : Calcul de la portion de la proportion ajustée par groupe d'âge (E) = proportion brute du groupe d'âge (A) multipliée par le poids d'ajustement du groupe d'âge (D).

Exemple : 25-44 ans, (E) = 0,977 X 0,36996982 = 0,361460512.

Étape 4 : Calcul de la proportion ajustée (E : total) = total des portions de la proportion ajustée (E).

Étape 5 : Calcul de la portion de la variance de la proportion ajustée par groupe d'âge (F) = le carré de l'erreur-type du groupe d'âge (B) multipliée par le carré du poids d'ajustement du groupe d'âge (D)

Exemple : 25-44 ans, (F) = (0,0102453)² X (0,36996982)² = 1,4368X10⁻⁵.

Étape 6 : Calcul de la variance de la proportion ajustée (F : total) = total des portions de la variance de la proportion ajustée (F). L'erreur-type est la racine carrée de cette valeur, soit 0,010648.

Annexe 12

Tests statistiques – un bref survol

Afin de mieux comprendre le rôle et l'importance des tests statistiques dans le contexte d'une analyse faite à partir de données d'enquête par échantillonnage, telle l'EQSP, voici un bref survol de certains concepts de base en inférence statistique.

On appelle *inférence statistique*, l'ensemble des méthodes développées pour permettre d'obtenir de l'information concernant une population sans avoir à faire de recensement⁹¹. Nous pouvons résumer la procédure de l'inférence statistique en quatre étapes :

- a) choisir un échantillon représentatif de la population;
- b) pondérer l'échantillon;
- c) étudier les variables statistiques dans l'échantillon pondéré;
- d) tirer des conclusions sur le comportement de ces variables statistiques dans la population.

Les méthodes employées à chacune de ces étapes sont respectivement l'échantillonnage, la pondération, l'estimation et les tests statistiques. Les prochains paragraphes porteront sur cette dernière catégorie.

Qu'est-ce qu'un test statistique?

Une *hypothèse statistique* est un énoncé ou une affirmation concernant une ou plusieurs populations. On parle de *test statistique* (ou *test d'hypothèse*) lorsque, par une expérience quelconque, on arrive à confirmer ou infirmer une hypothèse. Par exemple, si en observant un jeune homme jouer aux dés on constate qu'il gagne trop souvent, alors on formulera automatiquement l'hypothèse que les dés sont truqués et l'on voudra tester cette hypothèse en prenant nous-mêmes les dés et en les lançant un certain nombre de fois dans le but de comparer les résultats obtenus aux résultats théoriques de dés non truqués (c'est-à-dire lorsque chaque face a une chance sur six d'être obtenue). Donc, un test statistique est une procédure appliquée pour déterminer si les données de l'échantillon soutiennent les énoncés formulés au sujet de la population. La confirmation ou l'infirmerie d'une hypothèse est toujours faite avec une certaine probabilité que l'on voudra cependant aussi forte que possible.

À quoi sert un test statistique?

L'étape finale d'un test d'hypothèse consiste à prendre la décision de rejeter ou d'accepter l'hypothèse. Cette décision sera basée sur une estimation du paramètre concerné et cette estimation est calculée à partir de données recueillies dans un échantillon.

On ne doit donc pas s'attendre à ce qu'une estimation ait exactement la valeur prévue par l'hypothèse (par exemple 1/6 pour la face d'un dé). On acceptera quand même l'hypothèse si l'estimation du paramètre est relativement près de la valeur du paramètre prévue par l'hypothèse. On dit alors que l'écart n'est pas significatif et qu'il est vraisemblablement dû au hasard de l'échantillonnage.

⁹¹ C'est-à-dire recueillir l'information auprès de tous les individus de la population.

Comment interpréter un test statistique?

En matière d'interprétation, lorsqu'un test entre deux proportions est significatif, alors on peut affirmer que les proportions sont statistiquement différentes. Par contre, lorsqu'un test n'est pas significatif, alors on ne peut pas pour autant conclure que les proportions sont égales. Dans ce cas, le test nous indique seulement que les données de l'enquête, c'est-à-dire la précision obtenue à l'aide de l'échantillon, ne nous permettent pas de conclure que les proportions sont différentes. En fait, il est possible que, si 1 000 répondants de plus avaient été enquêtés, alors la précision des estimations aurait été « meilleure » et le test aurait été cette fois significatif. C'est pourquoi, on utilise par exemple la formulation suivante pour faire état d'un résultat non significatif : « Pour ce qui est de la proportion de la population qui a des symptômes de rhinite allergique en Estrie et en Montérégie, l'enquête ne nous permet pas de conclure à une différence significative » ou bien « qu'aucune différence significative n'a été observée pour la proportion de la population qui a des symptômes de rhinite allergique de ces deux régions ».

Exemple

Appliquons maintenant la théorie qui vient d'être énoncée à un exemple lié aux données de l'EQSP. Supposons qu'à partir des données de l'EQSP 2008, on désire comparer la proportion de la population de l'Abitibi-Témiscamingue qui a songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois avec celle du reste du Québec.

Les données de l'EQSP lui permettent d'estimer que la proportion de personnes qui a songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois en Abitibi-Témiscamingue s'élève à 3,4 % alors que celle du reste du Québec est de 2,8 %. À première vue, les résultats ne semblent pas très différents. Si, lors de l'enquête (EQSP), toutes les personnes de l'Abitibi-Témiscamingue avaient été enquêtées et qu'en plus, toutes avaient répondu, alors la valeur 3,4 % représenterait non pas une estimation, mais la proportion exacte de personnes qui a songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois en Abitibi-Témiscamingue. Et s'il en était de même pour le reste du Québec, alors la proportion exacte de personnes qui a songé sérieusement au suicide ou tenté de se suicider pour la même période pour le reste du Québec serait de 2,8 %. Dès lors, on pourrait conclure que les proportions sont différentes, car les valeurs ne sont pas identiques.

Malheureusement, ce n'est pas le cas. L'EQSP est une enquête par échantillon et les résultats qui en sont tirés correspondent à des estimations des valeurs exactes. Ainsi, l'estimation de la proportion de personnes qui ont songé sérieusement au suicide ou tenté de se suicider en Abitibi-Témiscamingue est associée à une mesure de précision qui reflète l'incertitude liée à l'échantillonnage (voir le tableau A12-1). La précision est souvent représentée par l'intervalle de confiance de l'estimation. Dans le cas de la proportion pour l'Abitibi-Témiscamingue, l'intervalle de confiance est [2,6 % ; 4,3 %], ce qui veut dire qu'on estime que la valeur exacte de la proportion se situe entre ces deux valeurs. De la même manière, on estime que la valeur exacte de la proportion de personnes qui a songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois mesurée sur le reste du Québec se situe entre 2,5 % et 3,1 %.

Rappelons que lorsqu'on désire comparer deux proportions, on doit utiliser un test statistique pour déterminer si les données de l'échantillon soutiennent les énoncés formulés au sujet de la population. Dans cet exemple, un test de la différence des proportions (voir section 5.1.2) nous permettrait de conclure que les deux estimations ne sont pas statistiquement différentes. Donc, qu'on ne peut pas conclure que la proportion de la population qui a songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois en Abitibi-Témiscamingue est différente de celle du reste du Québec.

Tableau A12-1 Proportion de la population qui a déjà songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois, EQSP 2008

Territoire		Statistiques					
Région	Nombre ¹	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	4 000	3,4	13,8	0,47	(2,6 - 4,3)	3,1	ns
Reste du Québec	171 800	2,8	5,7	0,16	(2,5 - 3,1)	2,6	

Source : Infocentre de santé publique, octobre 2010.

Annexe 13

Validité du test du khi-deux

Le test du khi-deux n'est valide qu'à certaines conditions. La condition utilisée pour les données du portail est liée au nombre de cellules du tableau analysé auxquelles correspondent des fréquences espérées inférieures à 5. Lorsque plus de 20 % des cellules ont une fréquence espérée inférieure à 5, on indique que le test pourrait ne pas être valide et une note apparaît au bas du tableau produit, à l'endroit où est normalement indiqué le résultat du test global. La note qui s'affiche est la suivante « *** ATTENTION : 20 % ou plus des cellules du tableau ont des valeurs espérées inférieures à 5, la validité du test n'est donc pas assurée. Si un regroupement de catégories est offert, il est recommandé de l'effectuer avant de faire l'interprétation du test. Sinon, interpréter les résultats avec prudence. *** ».

En général, lorsque le test d'indépendance du khi-deux n'est pas valide, il importe, comme le mentionne la note, de procéder au regroupement de quelques rangées ou colonnes du tableau avant de faire l'interprétation du test. Les rangées ou colonnes posant problème correspondent généralement à celles présentant de faibles effectifs. Autant que possible, ces rangées ou colonnes devront être regroupées avec d'autres qui sont assez homogènes par rapport à la variable d'étude.

Annexe 14

Tests de la différence de proportions – deux méthodes

Étant donné la nature catégorique⁹² de plusieurs variables et indicateurs étudiés, le test de la différence de proportions est l'un des outils importants de l'analyse de telles données. Bien entendu, pour être adéquates, les conclusions tirées des données de l'EQSP doivent être appuyées par un test statistique approprié qui tient compte du plan de sondage de l'enquête. Entre autres, les analyses doivent intégrer la pondération.

Pour l'exploitation des données de l'EQSP 2008 sur le portail de l'Infocentre, deux méthodes ont été implantées pour effectuer des comparaisons de proportions. Les deux méthodes proposées se servent du concept des poids bootstrap. Plus précisément, ces méthodes font intervenir la construction d'intervalle de confiance bootstrap (voir la section 3.3.3). Cette annexe présente les deux méthodes implantées à l'Infocentre pour la comparaison de proportions, ainsi que les avantages et les inconvénients de chacune d'elles.

Méthode 1 : test empirique de la différence de deux proportions

Comme abordé brièvement à la section 5.1.2, le test de la méthode 1 est basé sur l'utilisation des ensembles de poids bootstrap de l'enquête et a l'avantage de s'affranchir de l'hypothèse selon laquelle les données suivent approximativement une loi connue (Normale ou Binomiale). Cet avantage a pour effet de permettre l'utilisation de ce test, peu importe les circonstances, contrairement à un test t de la différence de deux proportions, comme c'est le cas pour les données de l'ESCC à l'Infocentre⁹³.

Le test de la méthode 1 repose sur la construction d'un intervalle de confiance bootstrap pour la différence des proportions comparées. Pour construire un intervalle de confiance bootstrap (voir la section 3.3.3), il suffit de calculer la différence des deux proportions pour chaque ensemble de poids bootstrap et, à partir de la distribution ordonnée des 2 000⁹⁴ valeurs obtenues, construire l'intervalle de la différence des proportions en prenant le 2,5^e percentile et le 97,5^e percentile de cette distribution, pour un intervalle de niveau de confiance de 95 %.

Le résultat du test de la méthode 1 s'obtient très simplement : on peut conclure que deux proportions sont différentes au seuil de 5 % lorsque l'intervalle de confiance de la différence ne contient pas la valeur zéro. Si par contre, l'intervalle contient la valeur nulle, alors on ne peut pas conclure que les deux proportions sont significativement différentes. La puissance du test de la méthode 1 est du même ordre de grandeur que celle d'un test de Student (test t).

⁹² Une variable catégorique (aussi appelée variable qualitative) est une variable où chaque réponse peut être classée dans une catégorie particulière.

⁹³ Pour les données de cette enquête à l'Infocentre, le test utilisé dépend de la méthode d'estimation de chacune des deux proportions comparées. Si la méthode d'estimation des deux proportions est la méthode normale, alors le résultat affiché est celui d'un test t (méthode 1). Si par contre, la méthode d'au moins une des deux proportions comparées est la méthode binomiale, alors la méthode 1 ne peut être utilisée, et le résultat d'un test plus conservateur est affiché, soit celui de la comparaison de deux intervalles de confiance (méthode 2). Pour plus de détails, consulter ISQ (2009).

⁹⁴ Rappelons que le nombre d'ensembles de poids bootstrap a été fixé à 2 000 dans le cas de l'estimation des percentiles d'une distribution, contrairement à 500 seulement pour l'estimation de la variance des estimations. Pour plus de détails, consulter la section 3.3.3.

La méthode 1 présente l'avantage de tenir compte de la corrélation entre les estimations (ou de la covariance entre \hat{p}_1 et \hat{p}_2), dans le calcul de la variance de la différence. En effet, l'utilisation d'une méthode par ré-échantillonnage comme le bootstrap pour le calcul d'une approximation de cette variance permet de tenir compte adéquatement du plan de sondage.

Méthode 2 : construction de deux intervalles de confiance

Il existe une autre avenue pour la comparaison de proportions lorsque les comparaisons ne sont pas affichées par la requête effectuée sur le portail de l'Infocentre (voir la section 6.3 pour les différentes situations menant à l'absence des comparaisons deux à deux) ou que la méthode 1 ne peut être utilisée. Celle-ci consiste à baser le test sur l'examen de deux intervalles de confiance calculés séparément pour chaque proportion plutôt que sur un intervalle de confiance pour la différence des proportions. Les intervalles calculés sont encore une fois des intervalles de confiance bootstrap (section 3.3.3).

Un test basé sur cette méthode fait l'hypothèse que la corrélation entre les deux proportions estimées est égale à -1, ce qui correspond à la valeur maximale de la variance de la différence. Par ailleurs, lorsqu'on utilise deux intervalles pour tester la différence entre deux proportions, on ne contrôle plus le seuil du test à 5 %. En effet, celui-ci sera au maximum de 5 %, ce qui veut dire que si la corrélation est supérieure à -1, alors le seuil réel sera inférieur à 5 % et des écarts significatifs au seuil désiré ne seront pas détectés par le test. Ce test est plus conservateur qu'un test basé sur l'intervalle de confiance de la différence (méthode 1), ce qui veut dire qu'il rejette moins souvent l'hypothèse d'égalité des proportions qu'il devrait au seuil de 5 %.

La façon d'interpréter ce test est la suivante : lorsque les intervalles bootstrap des deux proportions comparées ne présentent aucun chevauchement, on peut alors affirmer que les deux proportions sont significativement différentes, l'une étant plus grande que l'autre. À l'opposé, lorsque les intervalles bootstrap se chevauchent, c'est-à-dire lorsque la borne supérieure de l'intervalle de la plus petite estimation est plus grande que la borne inférieure de l'intervalle de l'estimation la plus grande, on ne peut affirmer que les deux proportions comparées sont statistiquement différentes. Cependant, en présence d'un faible chevauchement, étant donné que le test de la méthode 2 est très conservateur, il est possible qu'un test plus puissant permette de détecter une différence significative entre les proportions. Or, le test de la méthode 1 est plus puissant que celui de la méthode 2 pour tester la comparaison de proportions. Ce qui veut dire qu'elle détecte plus souvent des différences significatives.

Le tableau A14-1 montre un exemple de tableau produit sur le portail de l'Infocentre pour la méthode 1 décrite précédemment.

Tableau A14-1 Proportion de la population qui a déjà songé sérieusement au suicide ou tenté de se suicider au cours des 12 derniers mois, EQSP 2008

Territoire		Statistiques					
Région	Nombre ^a	Proportion brute	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	4 000	3,4	13,8	0,47	(2,6 - 4,3)	3,1	ns
Reste du Québec	171 800	2,8	5,7	0,16	(2,5 - 3,1)	2,6	

Source : Infocentre de santé publique, octobre 2010.

Annexe 15

Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre

L'annexe 15 expose une façon différente de présenter l'algorithme d'une requête paramétrable effectuée sur le portail de l'Infocentre de celle présentée à la section 6.2.2. Comme c'est le cas dans cette section, on retrouve deux parties à cette annexe, soit une première pour l'algorithme d'une requête pour un indicateur portant sur les individus de 15 ans et plus, et une seconde pour l'algorithme d'une requête pour un indicateur portant sur les enfants de 3 à 14 ans.

Requête pour un indicateur portant sur les individus de 15 ans et plus de l'EQSP

Paramètres : « Indicateurs/Indicateurs associés » :

Si le nom d'un indicateur associé est trop long, le nom de l'indicateur comprend du texte entre [] et ce texte n'est pas reproduit dans les indicateurs associés, il est plutôt remplacé par [...].

Autres paramètres :

Niveau géographique	Sous-population	Sexe	Groupe d'âge	Première variable de croisement	Deuxième variable de croisement	Test statistique	Calcul de la proportion ajustée, standardisation directe	
<i>Ensemble du Québec</i>	<i>Totale</i>			Aucune				
				Région		- Aucun		
						- Indicateur / var. de croisement	Non Oui ¹	
				Sexe Groupe d'âge Variables fiche	Aucune	- Aucun		
				Région Variables fiche				
	<i>Sexe</i>	Par sexe Masculin Féminin			Aucune			
					Région		- Aucun	
							- Indicateur / var. de croisement	Non Oui ¹
					Groupe d'âge Variables fiche	Aucune	- Aucun	
				Région Variables fiche				
	<i>Âge</i>			Groupes d'âge	Aucune			
					Région		- Aucun	
							- Indicateur / var. de croisement	
					Sexe Variables fiche	Aucune	- Aucun	
				Région Variables fiche				
	<i>Sexe et âge</i>	Par sexe Masculin Féminin		Groupes d'âge	Aucune			
Région Variables fiche						- Aucun		
						- Indicateur / var. de croisement		

¹ Sauf si l'âge est en variable de croisement, et pour certains indicateurs pour qui la proportion ajustée n'était pas demandée dans la fiche-indicateur.

Niveau géographique	Région sociosanitaire	Sous-population	Sexe	Groupe d'âge	Comparaison régionale	Région de comparaison	Variable de croisement	Test statistique	Calcul de la proportion ajustée, standardisation directe			
Régional	Choix d'une région	Totale			- Aucune		- Aucune - Sexe - Groupe d'âge - Variables fiche	- Aucun - Indicateur / var. croisement	Non Oui ¹			
					- Région vs reste du Québec				Non Oui			
					- Région vs une autre région		Choix région de comparaison		Non Oui			
					- Aucune		Choix région de comparaison	- Aucune - Groupe d'âge - Variables fiche	- Aucun - Indicateur / var. de croisement	Non Oui ¹		
					- Région vs reste du Québec				Non Oui			
					- Région vs autre région				Non Oui			
		Age				Groupes d'âge	- Aucune		- Aucune - Sexe - Variables fiche	- Aucun - Indicateur / var. de croisement		
							- Région vs reste du Québec					
							- Région vs autre région		Choix région de comparaison			
							- Aucune		Choix région de comparaison	- Région vs reste du Québec		
							- Région vs une autre région					
Sexe et âge			Par sexe Masculin Féminin	Groupes d'âge	- Aucune							
					- Région vs reste du Québec							
					- Région vs une autre région		Choix région de comparaison					

¹ Sauf si l'âge est en variable de croisement et pour certains indicateurs pour qui la proportion ajustée n'était pas demandée dans la fiche-indicateur.

Niveau géographique	Région sociosanitaire avec suréchantillon	Sous-population	Sexe	Groupe d'âge	Comparaison locale	Variable de croisement	Test statistique	Calcul de la proportion ajustée, standardisation directe		
Local	Choix d'une région avec sur-échantillon	Totale			- Aucune	- Aucune	- Aucun	Non Oui ¹		
					- Sexe		- Indicateur / var. de croisement			
					- Groupe d'âge					
					- Variables fiche					
							- RLS vs reste de la région			Non Oui
		Sexe	Par sexe Masculin Féminin				- Aucune	- Groupe d'âge	- Aucun	Non Oui ¹
							- Variables fiche		- Indicateur / var. de croisement	
							- RLS vs reste de la région			
		Age				Groupes d'âge	- Aucune	- Sexe	- Aucun	
							- Variables fiche		- Indicateur / var. de croisement	
							- RLS vs reste de la région			
Sexe et âge	Par sexe Masculin Féminin			Groupes d'âge	- Aucune	- Variables fiche				
					- RLS vs reste de la région					

¹ Sauf si l'âge est en variable de croisement et pour certains indicateurs pour qui la proportion ajustée n'était pas demandée dans la fiche-indicateur.

Requête pour un indicateur portant sur les enfants de 3 à 14 ans de l'EQSP

Paramètre : « Indicateurs/Indicateurs associés » :

Si le nom d'un indicateur associé est trop long, le nom de l'indicateur comprend du texte entre [] et ce texte n'est pas reproduit dans les indicateurs associés, il est plutôt remplacé par [...].

Autres paramètres :

Niveau géographique	Sous-population	Sexe	Groupe d'âge	Variable de croisement	Test statistique		
Ensemble du Québec	Totale			- Aucune			
				- Région	- Aucun - Indicateur / var. de croisement		
				- Sexe de l'enfant - Groupe d'âge - Variables fiche	- Aucun - Indicateur / var. de croisement		
	Sexe de l'enfant	- Par sexe - Masculin - Féminin			- Aucune		
					- Région	- Aucun - Indicateur / var. de croisement	
					- Groupe d'âge - Variables fiche	- Aucun - Indicateur / var. de croisement	
	Âge de l'enfant			- 3-5, 6-14 ans - 3-5, 6-8, 9-11, 12-14 ans	- Aucune		
					- Sexe de l'enfant - Variables fiche	- Aucun - Indicateur / var. de croisement	
				- 6-14 ans		- Aucune	
						- Sexe de l'enfant - Région - Variables fiche	- Aucun - Indicateur / var. de croisement

Niveau géographique	Région sociosanitaire	Sous-population	Comparaison régionale	Région de comparaison	Variable de croisement	Test statistique
<i>Régional</i>	<i>Choix d'une région</i>	<i>Totale</i>	Aucune		Aucune	
			Région vs reste du Québec			
			Région vs autre région	Choix région de comparaison		
		<i>6-14 ans</i>	Aucune		Aucune	
			Région vs reste du Québec			
			Région vs autre région	Choix région de comparaison		