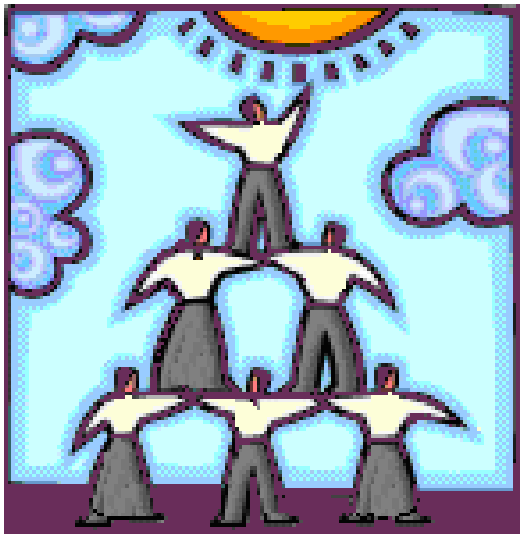


Plan commun de surveillance de l'état de santé de la population et de ses déterminants

Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires

– Enquête québécoise sur la santé des jeunes du secondaire 2010-2011 –

Version finale

Institut de la statistique du Québec

en collaboration avec
l'Institut national de santé publique du Québec

Mai 2013

Pour tout renseignement concernant l'ISQ
et les données statistiques dont il dispose,
s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec) G1R 5T4

Téléphone : 418 691-2401 ou 1 800 463-4090
(sans frais d'appel au Canada et aux États-Unis)

Site Web : www.stat.gouv.qc.ca

Dépôt légal
Bibliothèque et Archives Canada
Bibliothèque et Archives nationales du Québec
2e trimestre 2013
ISBN 978-2-550-67917-2 (PDF)

© Gouvernement du Québec, Institut de la statistique du Québec, 2013

Toute reproduction autre qu'à des fins de consultation personnelle
est interdite sans l'autorisation du gouvernement du Québec.
www.stat.gouv.qc.ca/droits_auteur.htm

Mai 2013

Le présent 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 générale adjointe de santé publique du ministère de la Santé et des Services sociaux.

L'Institut de la statistique du Québec a assuré la coordination des travaux du guide méthodologique et en a assumé la rédaction, en collaboration avec l'Institut national de santé publique du Québec. Le ministère de la Santé et des Services sociaux a, pour sa part, collaboré aux orientations et au suivi des travaux de ce guide.

Coordination et rédaction

Nathalie Plante, Direction de la méthodologie et de la qualité, ISQ

Collaboration à la rédaction

Mikaël Berthelot, Direction des statistiques de santé, ISQ
Véronique Boiteau, Infocentre, INSPQ
Robert Courtemanche, Direction de la méthodologie et de la qualité, ISQ
Annick Descormiers, Infocentre, INSPQ
Suzanne Gingras, Infocentre, INSPQ
Denis Hamel, Unité Connaissance-surveillance, INSPQ
Issouf Traoré, Direction des statistiques de santé, ISQ

Secrétariat

France Vaillancourt, Direction des statistiques de santé, ISQ
France Lozeau, Direction des statistiques du travail et de la rémunération, ISQ

Relecture

Sylvie Bernier, Direction de santé publique de l'ASSS de la Mauricie et du Centre-du-Québec
Monique Godin, Direction générale adjointe de santé publique, DGSP–MSSS
Jérôme Martinez, Unité Études et analyses de l'état de santé de la population, INSPQ
Julie Soucy, Direction générale adjointe de santé publique, DGSP–MSSS
Maria Constanza Street, Direction de santé publique de l'ASSS de Montréal

Enquête financée par

Ministère de la Santé et des Services sociaux et Institut de la statistique du Québec

Citation suggérée

INSTITUT DE LA STATISTIQUE DU QUÉBEC, en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC (2013). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*, version finale, Québec, Gouvernement du Québec, 124 p.

Table des matières

Introduction **11**

I. Description générale de l'*Enquête québécoise sur la santé des jeunes du secondaire 2010-2011* (EQSJS) **13**

1. Contexte général 13
2. Objectifs et thèmes de l'EQSJS 13
3. Accès aux données de l'EQSJS 14

II. Aspects méthodologiques et statistiques **17**

1. Plan d'échantillonnage et stratégie de collecte **19**

1.1 Population visée et population enquêtée 19
1.2 Bases de sondage 20
1.3 Méthode de sélection des élèves 22
1.3.1 Stratégie d'échantillonnage *22*
1.3.2 Classification des classes selon le niveau scolaire *23*
1.4 Taille et répartition de l'échantillon 23
1.5 Méthode de collecte 25
1.5.1 Instruments de collecte *25*
1.5.2 Stratégie de collecte *26*
1.5.3 Validation et saisie *28*
1.6 Résultats de collecte 29
1.6.1 Nombre de répondants *29*
1.6.2 Taux de réponse *30*
1.6.3 Période de collecte *31*

2. Estimation des proportions et des effectifs **33**

2.1 Définition 33
2.2 Pondération 33
2.3 Non-réponse partielle 35
2.3.1 Définition *35*
2.3.2 Taux de non-réponse partielle *35*
2.3.3 Analyse de la non-réponse partielle et imputation *37*
2.3.4 Traitement pour l'estimation des proportions *38*
2.3.5 Traitement pour l'estimation des effectifs *40*
2.3.6 Traitement de la non-réponse en cascade *41*

3. Mesures de précision pour les proportions **45**

3.1 Estimation de la variance 45
3.1.1 Définition et méthodes d'estimation *45*
3.1.2 Impact de l'imputation sur la variance *46*
3.2 Coefficient de variation 46
3.2.1 Définition et méthode de calcul *46*
3.2.2 Recommandations concernant les seuils de diffusion *47*

3.3 Intervalle de confiance 47
3.3.1 Définition *47*
3.3.2 Méthode de calcul pour les intervalles de confiance standard *48*
3.3.3 Méthode de calcul pour les intervalles de confiance « bootstrap » *49*
3.3.4 Recommandations relatives au choix du niveau de confiance ou au seuil de signification *50*

4. Ajustement des proportions selon l'âge 51
4.1 Définition 51
4.2 Méthode de calcul 51
4.3 Variance et intervalle de confiance pour les proportions ajustées 52

5. Comparaison de proportions 53
5.1 Tests statistiques 53
5.1.1 Test global d'indépendance *53*
5.1.2 Test d'égalité de deux proportions *53*

6. Exploitation des données de l'EQSJS 2010-2011 55
6.1 Comparabilité des résultats de l'EQSJS 2010-2011 avec d'autres sources de données 55
6.2 Production de tableaux et comparaison des prévalences sur le portail de l'Infocentre de santé publique 56
6.2.1 Paramètres d'une requête sur le portail de l'Infocentre de santé publique *56*
6.2.2 Particularités des paramètres pour certains indicateurs *60*
6.2.3 Algorithme d'une requête paramétrable sur le portail pour les données de l'EQSJS 2010-2011 ..... *61*
6.2.4 Exemples de résultats de requêtes sur le portail de l'Infocentre de santé publique et recommandations *66*
6.3 Particularités des analyses produites à l'Infocentre de santé publique 72
6.3.1 Indicateur en répartition *73*
6.3.2 Nombre trop important de tests demandés pour une requête *74*
6.3.3 Non-réponse partielle élevée et sous-estimation des effectifs *75*
6.3.4 Diffusion par réseau local de services (RLS) ou par commission scolaire *78*
6.3.5 Analyse des données par niveau scolaire *79*
6.3.6 Région sociosanitaire de l'école *80*
6.3.7 Écarts d'âge régionaux *80*
6.3.8 Indicateurs d'âge d'initiation *80*
6.3.9 Emploi durant l'année scolaire *81*
6.3.10 Point de césure des indicateurs selon les centiles *81*
6.3.11 Biais de désirabilité sociale *82*

7. Protection des renseignements personnels 83
7.1 Confidentialité des fichiers de microdonnées 83
7.2 Confidentialité des tableaux destinés à la diffusion 83
7.3 Règles de masquage par suppression en présence de données à risque d'identification 85

III. Recommandations pour l'interprétation des résultats 87

1. Lien de causalité 87

2. Inférence à la population visée 87

3. Référence à un nombre **87**
4. Estimations et non valeurs exactes **88**
5. Différence non statistiquement significative **88**
6. Présentation de résultats sous forme de ratio **89**
7. Présentation de résultats selon l'ordre des estimations **89**
8. Interprétation sous forme de tendance **90**
9. Interprétation des résultats pour les indicateurs basés sur les centiles **90**
Références bibliographiques **91**

Liste des annexes

Annexe 1 Comment utiliser la pondération lors de la production d'estimations à partir des données de l'EQSJS 2010-2011

Annexe 2 Diagnostic et analyse de la non-réponse partielle pour les données de l'EQSJS 2010-2011

Annexe 3 Exemples de calcul pour l'estimation d'un effectif

Annexe 4 Information supplémentaire concernant la méthode du « bootstrap »

Annexe 5 Intervalle de confiance binomial

Annexe 6 Niveau de confiance

Annexe 7 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 8 Tests statistiques – un bref survol

Annexe 9 Validité du test du khi-deux

Annexe 10 Tests d'égalité de deux proportions – deux méthodes

Annexe 11 Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre de santé publique

Introduction

L'*Enquête québécoise sur la santé des jeunes du secondaire 2010-2011* (EQSJS) s'inscrit parmi une nouvelle génération d'enquêtes planifiées à des fins de surveillance et faisant partie du *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'EQSJS. 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'EQSJS. Cette enquête est la troisième[1] à être traitée à l'Infocentre de santé publique dans le cadre de la mise en œuvre du *Plan commun de surveillance de l'état de santé de la population et de ses déterminants 2004-2007* (PCS).

Le *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé des jeunes du secondaire 2010-2011* 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'établir des façons de faire et des balises communes relativement à ces aspects aux fins des activités de surveillance.

Ce guide vise à soutenir les utilisateurs dans l'analyse et l'interprétation des résultats des indicateurs d'enquêtes sociosanitaires du PCS, accessibles sur le portail de l'Infocentre de santé publique, 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 document a été rédigé en s'inspirant du *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance – Enquête québécoise sur la santé de la population 2008* (ISQ, 2010). La première partie donne une description générale de l'EQSJS. 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 pour l'enquête; 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 comparaison des proportions fait l'objet de la section 5. Les sections 6 et 7 portent respectivement sur l'exploitation des données de l'EQSJS à l'Infocentre de santé publique et sur la protection de la confidentialité des données des répondants ainsi que des écoles ayant participé à l'enquête. Finalement, des recommandations sur l'interprétation des résultats sont proposées dans la troisième et dernière partie du document.

Le guide se veut un outil dynamique et vise à identifier les balises méthodologiques communes afin d'optimiser l'utilisation de l'information qui provient de l'EQSJS, et ce, afin de mieux soutenir la prise de décision et d'informer la population sur son état de santé et ses déterminants. Il est ainsi appelé à s'enrichir et à évoluer en fonction, notamment, des commentaires des utilisateurs, des prochaines éditions de l'EQSJS et de la mise au point de nouvelles méthodologies. Il est à noter que la présente version constitue une mise à jour du guide paru en octobre 2012 lors de la diffusion des données sur la santé physique et les habitudes de vie.

1. Les deux premières étant l'Enquête sur la santé dans les collectivités canadiennes (ESCC), de Statistique Canada, et l'Enquête québécoise sur la santé de la population 2008 (EQSP), réalisée par l'ISQ.

I. Description générale de l'*Enquête québécoise sur la santé des jeunes du secondaire 2010-2011* (EQSJS)

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 réalisée, suivie des deux enquêtes sociales et de santé de 1992-1993 et de 1998, puis de l'Enquête sociale et de santé auprès des enfants et des adolescents de 1999. 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.

Cette série d'enquêtes québécoises a ensuite été interrompue, Statistique Canada implantant au début des années 2000 l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC; Statistique Canada, 2011). Cette enquête transversale est constituée de cycles généraux (2000-2001, 2003, 2005, 2007-2008, 2009-2010, 2011-2012) représentatifs à l'échelon régional dans chacune des provinces et des territoires. À partir de ce moment, l'ESCC a comblé une partie des besoins d'information du MSSS sur la santé de la population du Québec.

Ces besoins 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 ainsi 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. Toutefois, d'autres objets et indicateurs inscrits au PCS ne sont pas couverts par le contenu commun de l'ESCC ou n'ont pas une précision satisfaisante à l'échelon régional. 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'EQSJS est une enquête transversale à portée régionale qui s'inscrit dans ce programme d'enquêtes. Complémentaire à l'ESCC pour ce qui est de la population des 12 à 17 ans, elle devrait être répétée aux cinq ans. Le MSSS a confié sa réalisation à l'Institut de la statistique du Québec (ISQ), comme le prévoit l'article 42 de la LSP.

2. Objectifs et thèmes de l'EQSJS

L'objectif général de l'EQSJS est d'alimenter en données les plans de surveillance afin d'améliorer l'état de santé des jeunes québécois du secondaire. Plus spécifiquement, l'enquête vise à :

- combler certains besoins d'information prioritaires sur l'état de santé et de bien-être des jeunes québécois du secondaire et sur ses déterminants;
- contribuer à une meilleure connaissance de leur état de santé physique, mentale et psychosociale et de leurs habitudes de vie;
- recueillir des données complémentaires sur les jeunes du secondaire qui facilitent l'étude des liens entre leur état de santé ou leurs habitudes de vie et leurs caractéristiques individuelles;

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

- produire des estimations à portée provinciale et régionale;
- guider les actions ou interventions visant à améliorer la santé et le bien-être des jeunes;
- suivre l'évolution dans le temps de l'état de santé des jeunes du secondaire et de ses déterminants (enquête répétée aux 5 ans).

Une consultation tenue au printemps 2007 par le MSSS au sein de la Direction générale de la santé publique et auprès des directions régionales de santé publique, des cinq tables de concertation nationales en santé publique et de l'Institut national de santé publique du Québec (INSPQ) a permis de faire un premier exercice de priorisation du contenu de l'EQSJS en fonction des besoins en information ainsi que de différents éléments et critères de faisabilité (prévalence, temps d'entrevue, périodicité, budget, etc.). Le choix final des indicateurs a ensuite été effectué par le Comité d'orientation de projet (COP)[3].

Les indicateurs sélectionnés sont rattachés aux thèmes suivants[4] :

Santé physique et habitudes de vie

- Perception de l'état de santé
- Poids et image corporelle
- Habitudes alimentaires
- Activité physique de transport et de loisir
- Santé respiratoire
- Santé buccodentaire
- Consommation de tabac, d'alcool et de drogue
- Comportements sexuels
- Expérience de travail

Santé mentale et psychosociale

- Santé mentale (ex. : détresse psychologique)
- Estime de soi, résilience et compétences sociales
- Problèmes d'adaptation sociale
- Environnement scolaire
- Environnement familial
- Caractéristiques du groupe de pairs

3. Accès aux données de l'EQSJS

L'accès aux résultats de l'EQSJS peut se faire par l'exploitation de deux ensembles de données, soit le fichier maître ou le fichier masqué contre l'identification involontaire (FMII).

Le **fichier maître** contient l'ensemble de l'information recueillie sur les répondants pour chacune des 16 régions sociosanitaires couvertes par l'enquête[5]. Lors de la collecte des données, le répondant était informé que les réponses qu'il fournirait seraient communiquées au MSSS et à 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 de santé publique à l'INSPQ l'exploite pour le déploiement des indicateurs de surveillance sur son portail informationnel et met des résultats statistiques à la disposition de ses utilisateurs. Le MSSS, quant à lui, utilise ce fichier 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

3. Le COP, sous la responsabilité partagée de l'ISQ et du MSSS, est composé de représentants des directions régionales de santé publique, des tables de concertation nationales en santé publique et de l'INSPQ.
4. Ces indicateurs ne sont pas couverts ou peu couverts au niveau régional par les sources de données existantes et ont été retenus parmi environ 80 indicateurs pour lesquels l'enquête constitue la meilleure source.
5. Les régions du Nunavik (17) et des Terres-Cries-de-la-Baie-James (18) ne sont pas couvertes par l'enquête.

propres besoins (rapports, publications diverses, etc.), tout en favorisant l'accès aux données à des chercheurs par l'intermédiaire 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. Aucun masquage de données n'est appliqué à ce fichier.

Le **FMII** est un fichier masqué qui peut être exploité au moyen d'un accès à distance accordé par l'ISQ. Il contient l'ensemble des répondants de l'EQSJS. Pour quelques répondants, toutefois, certaines données ont dû être masquées afin 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 pour des explications supplémentaires). Le tableau 1.1 présente les caractéristiques du FMII liées au masquage des données. Des détails sur les variables du fichier maître seront quant à eux disponibles dans le cahier technique de l'enquête (ISQ, 2013).

Tableau 1.1
Caractéristiques du FMII liées au masquage des données de l'EQSJS 2010-2011

Fichier	Nombre de répondants	Nombre de variables			
		Affectées par le masquage	Supprimées du fichier	Recatégorisées	Pour lesquelles des valeurs ont été supprimées
FMII (Fichier masqué contre l'identification involontaire)	63 196	1	11	4	17

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*, Fichier masqué contre l'identification involontaire.

II. Aspects méthodologiques et statistiques

Afin de faire une utilisation adéquate des données et des résultats issus de l'EQSJS 2010-2011, il importe de connaître la méthodologie utilisée. En effet, le plan d'échantillonnage de l'enquête, les procédures de collecte et de traitement des données ont tous une incidence sur les estimations produites. La connaissance de ces éléments permettra à l'utilisateur des données de l'enquête d'en faire une analyse judicieuse afin d'élaborer des politiques appropriées et adaptées aux populations visées. Cette partie du guide présente la méthodologie générale de l'enquête.

S'y trouve d'abord un aperçu des aspects méthodologiques importants de l'EQSJS 2010-2011 à prendre en considération, en particulier pour l'analyse des données de l'enquête. L'usager qui ne souhaite pas faire une lecture complète du guide peut s'y référer directement. Les numéros des sections du guide traitant de chacun des sujets y sont fournis. Précisons qu'en général, les tableaux apparaissant dans cette partie du guide sont issus du fichier maître et présentent les données pour l'ensemble du Québec ainsi que pour ses régions sociosanitaires.

Aperçu des principaux aspects méthodologiques

- L'EQSJS 2010-2011 porte sur une population de quelque 430 000 élèves du secondaire inscrits au secteur des jeunes dans les écoles québécoises à l'automne 2010, à l'exclusion principalement des centres de formation professionnelle, des écoles des Terres-Cries-de-la-Baie-James et du Nunavik ainsi que des écoles comptant 30 % ou plus d'élèves handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDAA) (section 1.1).
- L'enquête a une portée tant nationale que régionale. Elle vise la diffusion de statistiques de bonne précision pour toute proportion de 16 % ou plus parmi les élèves des combinaisons résultant du croisement du niveau scolaire, du sexe et de la région sociosanitaire associée à l'école (16 régions). Pour huit régions, des objectifs à l'échelle des réseaux locaux de services ou des commissions scolaires sont également visés (section 1.4).
- La sélection des élèves pour l'enquête a été faite séparément pour chaque combinaison de région et de niveau scolaire. En tout, 483 écoles admissibles à l'enquête ont d'abord été tirées aléatoirement avec une probabilité proportionnelle au nombre d'élèves inscrits, sur la base des données du ministère de l'Éducation, du Loisir et du Sport (MELS) de l'année scolaire 2009-2010. Une liste des classes des écoles sélectionnées a été dressée avec un responsable de l'école, par interview téléphonique. Une ou plusieurs classes ont ensuite été choisies aléatoirement au sein de chacune de ces écoles, et tous les élèves ont été invités à participer; 2 723 classes ont ainsi été sélectionnées. Un plus grand nombre de classes a été sélectionné au sein des régions ayant des objectifs de diffusion sous-régionaux (section 1.4).
- Les élèves participants ont rempli un questionnaire autoadministré informatisé d'une durée moyenne de 34 minutes sur miniportable durant une période de cours, après quoi leurs données ont été transmises à l'ISQ de façon électronique et sécurisée. Deux questionnaires comptant une majorité de questions communes ont été distribués, chacun à la moitié des élèves. L'enquête s'est déroulée dans les écoles du 2 novembre 2010 au 17 mai 2011 (section 1.5.2).
- Les élèves de 2 651 classes réparties dans 470 écoles ont participé à l'enquête, pour un total de 63 196 élèves répondants. Le taux de réponse global pondéré, mesure de la qualité des données, est de 88 % (section 1.6.2).
- La classification des élèves selon le niveau scolaire utilisée pour l'EQSJS 2010-2011 suit la nomenclature du MELS. Néanmoins, le niveau scolaire attribué à certaines classes ne correspond pas nécessairement à celui auquel on s'attendrait pour des élèves d'un groupe d'âge donné. C'est le cas, principalement, pour les classes du parcours de formation axée sur l'emploi offert à partir de la 3e secondaire (sections 1.3.2 et 6.3.5).

- Afin d'inférer les résultats à la population visée, toutes les estimations obtenues sur le portail de l'Infocentre de santé publique ont été pondérées pour tenir compte, d'une part, du fait que certains élèves avaient plus de chances d'être sélectionnés que d'autres et, d'autre part, de la non-réponse plus importante observée chez certains groupes d'élèves. En raison de l'existence de deux versions du questionnaire, trois pondérations sont disponibles pour les analyses (sections 2.1 et 2.2). Afin de tenir compte du plan de sondage, les poids « bootstrap » associés à chacune de ces trois pondérations sont utilisés aux fins des estimations de précision et des tests statistiques globaux. Pour la comparaison de proportions prises deux à deux, lorsque la requête de l'usager sur le portail ne permet pas le recours aux poids « bootstrap », une comparaison des intervalles de confiance est suggérée (section 3).
- Recommandations pour l'analyse des données de l'EQSJS 2010-2011 et l'interprétation des résultats :
  - Les risques de biais dus à la non-réponse partielle, relativement à la précision des estimations, sont faibles pour la presque totalité des indicateurs (proportion ou répartition). Quelques variables de croisement affichent toutefois une non-réponse partielle élevée, soit l'indicateur du plus haut niveau de scolarité des parents (9 %) et les indices de défavorisation matérielle et/ou sociale (15 % pour chacun des indices). En raison du taux de non-réponse partielle plus élevé observé pour ces derniers indices, des recommandations pour l'analyse et l'interprétation ont été énoncées dans le premier tome du rapport de l'*Enquête québécoise sur la santé des jeunes du secondaire 2010-2011* (Plante et autres, 2012). Par ailleurs, toutes les estimations d'effectifs doivent prendre en compte l'effet de la non-réponse partielle, s'il y a lieu, et ce, peu importe le taux observé (section 2.3).
  - L'enquête tend davantage à une analyse de données par niveau scolaire plutôt que par âge de l'élève. Bien que l'analyse par âge demeure pertinente, elle implique une perte de précision et appelle la prudence dans son interprétation, puisque l'échantillon de l'enquête n'est pas représentatif de l'ensemble des Québécois de chacun des groupes d'âge, surtout dans le cas des jeunes de 12 ans et moins ou de 17 ans et plus (section 6.3.5).
  - Le plan d'échantillonnage a été élaboré de manière à produire des estimations régionales de bonne précision, selon la région sociosanitaire de l'école fréquentée par l'élève et non selon la région de résidence de celui-ci (section 6.3.6).
  - La distribution estimée de l'âge des élèves varie quelque peu entre les régions sociosanitaires. La décision de standardiser ou non les statistiques régionales selon l'âge, aux fins de comparaison, relève par conséquent de l'angle d'analyse souhaité (section 6.3.7 et section 4).
  - Pour certains indicateurs sur les comportements à risque disponibles sur le portail de l'Infocentre de santé publique, l'initiation de ces comportements est mesurée uniquement sur la période antérieure à l'âge, au moment de l'enquête, de l'ensemble des élèves pris en compte pour l'estimation. Par exemple, l'indicateur de la proportion des élèves de 13 ans et plus ayant amorcé un comportement avant l'âge de 13 ans exclut de son dénominateur tous les élèves n'ayant pas atteint l'âge de 13 ans (section 6.3.8).
  - Pour mesurer le travail des élèves du secondaire durant l'année scolaire, la question utilisée fait référence à un emploi actuel. Ainsi, la proportion des élèves occupant un emploi durant l'année scolaire pourrait être sous-estimée (section 6.3.9).
  - Une proportion estimée pour un indicateur défini sur la base de centiles n'est pas une mesure de prévalence et ne doit pas être présentée comme un résultat de l'enquête. Par exemple, la proportion estimée pour un indicateur construit sur la base du quintile supérieur se situe par définition à 20 %. Cette proportion peut varier légèrement en raison de scores égaux obtenus par certains groupes d'élèves. Les proportions estimées pour ce type d'indicateur sont utiles pour comparer des groupes d'élèves entre eux (section 6.3.10).
  - Les mesures autorapportées peuvent être sujettes à un biais de désirabilité sociale. Certains thèmes étudiés dans l'EQSJS, tels que la consommation de drogues, les relations sexuelles et les comportements d'adaptation sociale, pourraient être plus sensibles à ce type de biais (section 6.3.11).
  - L'interprétation des résultats devrait suivre les recommandations présentées à la partie III.

1. Plan d'échantillonnage et stratégie de collecte

Cette section aborde en premier lieu les aspects liés au plan d'échantillonnage, soit la description des élèves qui composent la population visée et la population enquêtée, les bases de sondage à partir desquelles sont sélectionnés les élèves à solliciter pour participer à l'enquête, la méthode de sélection de ceux-ci ainsi que la taille et la répartition de l'échantillon. Elle traite par la suite de la stratégie de collecte, tant du point de vue de la planification de l'enquête que de son déroulement et des résultats obtenus, notamment les taux de réponse.

1.1 Population visée et population enquêtée

La population visée par l'EQSJS 2010-2011 est constituée de tous les élèves de la 1re à la 5e année du secondaire qui sont inscrits au secteur des jeunes, dans les écoles québécoises publiques et privées, francophones et anglophones, à l'automne 2010. Sont toutefois exclus les élèves qui fréquentent :

- les centres de formation professionnelle;
- les écoles situées dans les régions sociosanitaires des Terres-Cries-de-la-Baie-James et du Nunavik;
- les écoles de langue d'enseignement autochtone;
- les établissements hors réseau (relevant du gouvernement fédéral ou d'autres ministères provinciaux);
- les écoles composées d'au moins 30 % de jeunes handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDAA).

Par conséquent, les résultats issus de l'enquête portent spécifiquement sur la population visée ou un sous-groupe de cette population, à l'exclusion des élèves des catégories susmentionnées. En tenant compte de ces exclusions, on peut affirmer que la population visée par l'EQSJS couvre environ 98,4 % de l'ensemble des élèves québécois inscrits au secondaire au secteur des jeunes.

Par ailleurs, bien qu'ils fassent partie de la population visée, les élèves de certaines classes ont été soustraits de la population enquêtée pour des raisons logistiques, soit parce que les coûts de déplacement des intervieweurs lors des visites dans les écoles auraient été trop élevés, soit parce que le nombre de participants à l'enquête risquait d'être trop faible relativement aux efforts investis.

Ont d'abord été exclus les élèves de certaines écoles de régions éloignées (District n° 1 Parent ville de La Tuque, Natashquan, Côte-Nord-du-Golf-Saint-Laurent, Blanc-Sablon, Bonne-Espérance, Saint-Augustin, Gros-Mécatina, Baie-James et l'Île d'Anticosti) ou d'écoles de très petite taille (écoles comptant moins de 25 élèves pour un niveau scolaire donné pour les régions sociosanitaires Saguenay–Lac-Saint-Jean, Capitale-Nationale, Mauricie et Centre-du-Québec, Estrie, Montréal, Outaouais, Chaudière-Appalaches, Laval, Lanaudière, Laurentides et Montérégie et écoles comptant moins de 15 élèves pour les régions Bas-Saint-Laurent, Abitibi-Témiscamingue, Côte-Nord et Gaspésie–Îles-de-la-Madeleine). La sous-couverture liée à l'exclusion de ces écoles est estimée à environ 1 % de la population visée; elle concerne principalement les régions Côte-Nord et Gaspésie–Îles-de-la-Madeleine, avec des taux respectifs de 7 % et 3 %. La population enquêtée compte ainsi 592 écoles pour l'année scolaire 2010-2011.

Ont également été retranchés de la population enquêtée les élèves des classes de moins de cinq élèves ainsi que ceux qui font partie d'une classe où la majorité des élèves ne sont pas en mesure de lire un questionnaire en français ou en anglais ou de manipuler un ordinateur. Il s'agit pour la plupart de classes d'immersion composées d'élèves allophones ou encore de classes où les élèves présentent des troubles d'apprentissage sévères (au sein d'écoles composées de moins de 30 % d'élèves EHDAA). Cette évaluation était faite par une personne-ressource nommée par la direction de l'école. Bien qu'il ait été sensiblement plus important en 1re secondaire, ce type d'exclusion est difficile à quantifier avec précision à partir des données disponibles et n'expliquerait pas par ailleurs l'ensemble de la sous-couverture observée.

En effet, en plus des exclusions volontaires, les difficultés rencontrées lors de la collecte peuvent contribuer à la sous-couverture de la population visée. Par exemple, cette dernière pourrait être en partie attribuable à l'oubli de dénombrer certaines classes préalablement à la sélection de l'échantillon, ou encore à l'oubli de comptabiliser certains élèves. À l'opposé, il est possible qu'une part de la sous-couverture soit surestimée dans le cas des élèves ayant abandonné l'école au courant de l'année scolaire; ces élèves ne sont plus visés mais demeurent inclus dans les effectifs scolaires du MELS. Sur cette base, on estime que la population enquêtée représente environ 93,5 % de la population visée[6].

1.2 Bases de sondage

En l'absence d'une liste exhaustive des classes de toutes les écoles secondaires québécoises, le plan d'échantillonnage de l'EQSJS 2010-2011 requiert deux degrés. Au premier degré, des écoles sont sélectionnées à partir d'une première base de sondage, tandis qu'au second degré, une liste des classes des écoles échantillonnées est créée aux fins de l'enquête.

La première base de sondage ayant servi à sélectionner des écoles, pour chaque niveau scolaire, est le fichier de déclaration des clientèles scolaires (DCS) du MELS de l'année scolaire 2009-2010. Ce fichier comprend, entre autres renseignements, les coordonnées de l'établissement scolaire, le nombre d'élèves inscrits par niveau scolaire ainsi que le réseau et la langue d'enseignement. Pour des raisons de confidentialité, les renseignements sur les inscriptions à chacun des niveaux du secondaire y sont regroupés pour l'ensemble des écoles comptant 30 % ou plus d'EHDAA (écoles dont les élèves sont exclus de la population visée[7]).

Une seconde base de sondage ayant servi à sélectionner les classes est formée par la liste de l'ensemble des classes de chacun des niveaux scolaires pour lesquels l'école a été sélectionnée. Cette liste a été élaborée par la personne-ressource de la direction de l'école, avec l'aide d'un intervieweur. Un cours de sciences humaines a été suggéré comme point de repère pour la formation des classes (cours d'*Histoire et éducation à la citoyenneté* pour la 1re à la 4e secondaire et *Monde contemporain* pour la 5e secondaire), lorsque cela était pertinent. Deux raisons justifient le choix de cette discipline. Le cours doit être obligatoire pour donner à chaque élève une probabilité non nulle d'être sélectionné, de telle sorte que l'échantillon soit représentatif de l'ensemble de la population visée. Pour assurer une bonne représentativité, on doit également éviter les matières faisant partie de programmes de performance, telles que les mathématiques ou l'anglais langue seconde. Le bloc des sciences humaines répond à ces deux critères. Dans les écoles où une autre matière a dû être considérée pour constituer la base de sondage des classes, on s'est assuré que tous les élèves étaient inscrits à un cours de la matière retenue.

Le tableau 1.2 montre la composition de la population des écoles enquêtées, au regard du nombre de classes estimé[8], par région sociosanitaire et niveau scolaire. Ces estimations sont basées sur les données du MELS de l'année 2009-2010 et ont servi à construire le plan d'échantillonnage. La population des écoles enquêtées compte environ 15 000 classes.

6. Cette proportion est estimée par l'inverse du taux de glissement, soit le rapport de la somme des poids de sondage avant calage (voir section 2.2) sur le nombre total d'élèves de la population visée.
7. Inclure les élèves de ces écoles dans la population visée demanderait par conséquent des efforts très importants, d'autant plus que plusieurs des élèves EHDAA présentent des incapacités à remplir le questionnaire.
8. La population des écoles enquêtées est obtenue en retranchant de la population visée les écoles correspondant aux critères d'exclusion listés à la section 1.1. Le nombre de classes des écoles de la population enquêtée a été estimé par école et niveau scolaire en arrondissant à l'entier supérieur le résultat de la division entre le nombre d'élèves estimé d'après les données de la déclaration des clientèles scolaires du MELS de l'année scolaire 2009-2010 et le nombre moyen d'élèves par classe estimé d'après l'Enquête québécoise sur le tabac, l'alcool, la drogue et le jeu chez les élèves du secondaire, 2008 (ETADJES; Dubé et autres, 2009), soit 28,3 élèves.

Tableau 1.2

Estimation du nombre de classes de la population des écoles enquêtées, par région sociosanitaire et niveau scolaire, Québec, 2009-2010

	Région sociosanitaire	1re sec.	2e sec.	3e sec.	4e sec.	5e sec.	Total
01	Bas-Saint-Laurent	72	86	83	73	60	**374**
02	Saguenay–Lac-Saint-Jean	110	115	121	102	95	**543**
03	Capitale-Nationale	233	262	254	228	208	**1 185**
04	Mauricie et Centre-du-Québec	194	189	207	181	156	**927**
05	Estrie	128	126	134	126	101	**615**
06	Montréal	683	698	718	647	582	**3 328**
07	Outaouais	146	144	156	139	116	**701**
08	Abitibi-Témiscamingue	68	58	70	55	48	**299**
09	Côte-Nord	38	37	37	31	31	**174**
10	Nord-du-Québec	7	8	8	7	7	**37**
11	Gaspésie–Îles-de-la-Madeleine	35	37	41	33	30	**176**
12	Chaudière-Appalaches	147	160	167	133	123	**730**
13	Laval	159	142	147	140	121	**709**
14	Lanaudière	184	219	210	181	152	**946**
15	Laurentides	240	254	263	235	196	**1 188**
16	Montérégie	574	601	625	536	474	**2 810**
Total		**3 018**	**3 136**	**3 241**	**2 847**	**2 500**	**14 742**

Source : Ministère de l'Éducation, du Loisir et du Sport, Déclaration des clientèles scolaires (DCS), 2009-2010.

Compilation : Institut de la statistique du Québec, 2010.

Le tableau 1.3 présente quant à lui le nombre d'élèves de la population visée, par région sociosanitaire et niveau scolaire, mis à jour d'après les données du MELS de l'année scolaire 2010-2011. La population visée compte un total de 431 100 élèves.

Tableau 1.3

Nombre d'élèves de la population visée, par région sociosanitaire et niveau scolaire, Québec, 2010-2011

	Région sociosanitaire	1re sec.	2e sec.	3e sec.	4e sec.	5e sec.	Total
01	Bas-Saint-Laurent	2 087	2 110	2 380	1 973	1 786	**10 336**
02	Saguenay–Lac-Saint-Jean	2 846	3 187	3 407	3 041	2 576	**15 057**
03	Capitale-Nationale	6 606	6 997	7 450	6 471	6 105	**33 629**
04	Mauricie et Centre-du-Québec	5 354	5 336	6 124	5 189	4 611	**26 614**
05	Estrie	3 362	3 479	3 881	3 450	3 254	**17 426**
06	Montréal	19 687	21 065	21 151	19 227	18 217	**99 347**
07	Outaouais	4 148	4 415	4 444	3 866	3 345	**20 218**
08	Abitibi-Témiscamingue	1 770	1 758	2 283	1 500	1 288	**8 599**
09	Côte-Nord	1 127	981	1 131	874	785	**4 898**
10	Nord-du-Québec	185	205	197	155	137	**879**
11	Gaspésie–Îles-de-la-Madeleine	905	926	1 162	829	741	**4 563**
12	Chaudière-Appalaches	4 229	4 358	4 964	4 083	3 667	**21 301**
13	Laval	4 902	4 464	4 686	4 286	4 078	**22 416**
14	Lanaudière	5 463	6 337	6 169	5 086	4 364	**27 419**
15	Laurentides	6 645	7 163	7 998	6 799	5 965	**34 570**
16	Montérégie	16 839	17 429	18 729	16 258	14 573	**83 828**
Total		**86 155**	**90 210**	**96 156**	**83 087**	**75 492**	**431 100**

Source : Ministère de l'Éducation, du Loisir et du Sport, Déclaration des clientèles scolaires (DCS), 2010-2011.

Compilation : Institut de la statistique du Québec, 2011.

1.3 Méthode de sélection des élèves

1.3.1 Stratégie d'échantillonnage

L'échantillon a été sélectionné selon un plan d'échantillonnage par grappes stratifié à deux degrés. Au premier degré, des écoles ont été sélectionnées avec probabilité proportionnelle à leur taille (PPT), indépendamment pour chaque région sociosanitaire et chaque niveau scolaire. La mesure de taille utilisée pour la sélection est le nombre d'élèves inscrits par niveau à l'année scolaire 2009-2010, selon le MELS[9]. Le croisement des 16 régions et des cinq niveaux scolaires forme le premier niveau de stratification. Certaines régions ayant des objectifs additionnels de diffusion de statistiques sous-régionales ont également été stratifiées selon le réseau local de services (RLS) ou la commission scolaire (CS).

Puisqu'une portion élevée des écoles est sélectionnée pour chaque niveau scolaire, certaines combinaisons d'école et de niveau ont une probabilité de sélection initiale calculée supérieure à un. Pour un niveau scolaire donné, une école est considérée comme autoreprésentative lorsqu'elle est sélectionnée avec certitude. Autrement dit, si l'on répétait l'échantillonnage plusieurs fois, les écoles autoreprésentatives feraient partie de chacun des échantillons. Cela est dû au plan d'échantillonnage proportionnel à la taille de l'école ainsi qu'à la proportion élevée d'écoles sélectionnées parmi l'ensemble de la population en cause. Ces écoles dites « autoreprésentatives » pour un niveau scolaire ont été considérées en elles-mêmes comme strates d'échantillonnage (Statistique Canada, 2003). Pour un niveau scolaire donné, la sélection des écoles non autoreprésentatives (probabilité de sélection inférieure à un) a par conséquent été faite en excluant au préalable les écoles autoreprésentatives des strates d'échantillonnage. Afin d'assurer une bonne répartition de l'échantillon des écoles non autoreprésentatives, des variables de stratification supplémentaires ont été utilisées dans certaines régions, lorsque les effectifs le permettaient : la langue ou le réseau d'enseignement et l'indice de milieu socio-économique de l'école[10].

Étant donné que la base de sondage exploitée pour la sélection des écoles datait de l'année scolaire précédant l'enquête, certaines écoles sélectionnées ont dû être jugées inadmissibles à l'enquête pour le niveau scolaire pour lequel elles avaient été sélectionnées, soit en raison d'une fermeture (1 école), soit parce qu'il n'y avait plus de classe de ce niveau dans l'école. Dans ces rares cas, une nouvelle sélection a été faite.

Au second degré d'échantillonnage, pour un niveau scolaire donné, une ou plusieurs classes ont été sélectionnées aléatoirement pour chaque école échantillonnée au premier degré. Dans les écoles autoreprésentatives, le nombre de ces classes a été défini en arrondissant la probabilité de sélection initiale à l'entier inférieur ou supérieur (selon un nombre généré aléatoirement et tenant compte de l'écart entre cette probabilité et l'entier inférieur). Dans les écoles non autoreprésentatives, une seule classe a été retenue. Les classes ainsi sélectionnées constituent des grappes d'élèves, tous inclus à leur tour dans l'échantillon.

Il importe de mentionner que le plan d'échantillonnage a été élaboré de manière à produire des estimations régionales de bonne précision, selon la région sociosanitaire de l'école fréquentée par l'élève et non selon la région de résidence de celui-ci. Pour la grande majorité des régions, on observe toutefois une très forte concordance entre ces deux variables, de sorte que l'analyse selon la région associée à l'école devrait fournir un portrait similaire à celui qui prévaut selon la région de résidence. La concordance est moins bonne pour les élèves fréquentant les écoles de la région de Laval et, dans une moindre mesure, pour ceux fréquentant les écoles des Laurentides, de Lanaudière et de Montréal. Pour les régions de Laval, des Laurentides et de Lanaudière, on note qu'une part assez importante des élèves du réseau d'enseignement privé sont inscrits à l'école dans une autre région que celle où ils habitent.

9. Au moment de l'élaboration du plan d'échantillonnage, les données de la déclaration des clientèles scolaires de l'année 2010-2011 n'étaient pas disponibles.
10. L'indice de milieu socio-économique attribué à chacune des écoles secondaires publiques québécoises a été mis au point par le MELS; lors de l'étape de planification de l'enquête, il était disponible pour l'année 2008-2009.

Pour la région de Montréal, on observe la tendance inverse, c'est-à-dire qu'une part non négligeable des élèves qui fréquentent les écoles de Montréal habitent à l'extérieur de la région. Cet aspect est évoqué de nouveau à la section 6.3.6 en ce qui concerne les analyses produites à l'Infocentre de santé publique.

1.3.2 Classification des classes selon le niveau scolaire

Au 2e cycle d'enseignement secondaire (3e, 4e et 5e), l'élève peut avoir accès à l'un ou l'autre de deux parcours de formation générale[11] ou à un parcours de formation axée sur l'emploi. Dans ce dernier cas, deux types de formation sont proposés : la formation menant à l'exercice d'un métier semi-spécialisé (FMSS), d'une durée d'un an, et la formation préparatoire au travail (FPT), d'une durée de trois ans. Pour la classification des élèves inscrits au parcours de formation axée sur l'emploi, le MELS utilise la nomenclature administrative suivante dans son fichier de déclaration des clientèles scolaires :

- tous les élèves des classes de formation menant à l'exercice d'un métier semi-spécialisé (FMSS) sont classifiés en 3e secondaire, peu importe leur âge;
- tous les élèves des classes de 1re année de formation préparatoire au travail (FPT) sont classifiés en 3e secondaire, peu importe leur âge;
- tous les élèves des classes de 2e année de FPT sont classifiés en 4e secondaire, peu importe leur âge;
- tous les élèves des classes de 3e année de FPT sont classifiés en 5e secondaire, peu importe leur âge.

Au 1er cycle d'enseignement (1re et 2e secondaire), une formation régulière est offerte à l'ensemble des élèves, qui sera dorénavant désignée par l'expression « parcours de formation générale ».

Lors de l'entrevue téléphonique avec le responsable de l'école, le niveau scolaire attribué aux classes pour leur sélection aléatoire a été défini selon cette même nomenclature dans le cas des formations axées sur l'emploi. Cela était nécessaire car le niveau déclaré par le responsable, dans bon nombre de cas, ne pouvait correspondre à un niveau de la nomenclature du MELS[12]. Il arrivait également que le responsable ne puisse attribuer ni un parcours de formation générale ni le parcours de formation axée sur l'emploi (FMSS ou FPT) à certaines classes du 2e cycle. De même, le responsable ne pouvait parfois pas associer certaines classes du 1er cycle à la formation régulière[13]. Selon la description fournie lors de l'entrevue téléphonique, il s'agissait principalement dans ces deux situations de programmes adaptés pour élèves en difficulté (adaptation scolaire, cheminement adapté, programme préparatoire à la formation professionnelle, programme d'aide à la réussite, etc.). Les classes qui ne pouvaient être associées à une formation en particulier ont été regroupées sous l'appellation « autre parcours »[14]. Aux fins de l'échantillonnage, ces classes ainsi que celles des parcours de formation générale ont toutes été réparties selon le niveau scolaire déclaré par le responsable, soit de la 1re à la 5e secondaire.

1.4 Taille et répartition de l'échantillon

Pour chaque région, la taille de base de l'échantillon a été fixée de manière à obtenir des statistiques régionales fiables par niveau scolaire et par sexe : 3 550 élèves répondants par région devaient ainsi permettre d'obtenir une bonne précision (coefficient de variation ≤ 15 %), pour une prévalence de 16 % ou plus dans ces domaines d'études. Selon l'hypothèse que 90 % des élèves des classes participantes répondraient au questionnaire et que le nombre moyen d'élèves par classe est de 28,3, le nombre de répondants requis pour l'atteinte des objectifs représente environ 28 classes échantillonnées par combinaison de région et de niveau. En prévision du refus éventuel de participer à l'enquête de la part de certaines classes ou écoles, cette taille d'échantillon a été haussée à 30 classes afin d'atteindre les objectifs de précision ciblés.

---

11. La désignation « parcours de formation générale » regroupe ici le parcours de formation générale et le parcours de formation générale appliquée.
12. Le niveau scolaire associé aux effectifs du fichier de déclaration des clientèles scolaires du MELS doit correspondre à celui utilisé dans l'enquête puisque ce fichier constitue la référence pour définir la composition de la population visée.
13. Il demeure donc une incohérence, voire un flou, pour certaines classes, entre la classification faite par les responsables des écoles et la classification administrative du MELS.
14. Cette désignation n'est pas utile au plan d'échantillonnage, mais elle est utilisée dans certaines analyses.

Par ailleurs, huit régions ont choisi d'accroître la taille de l'échantillon de base afin d'obtenir une bonne précision à l'échelle de leurs RLS ou de leurs CS (pour les RLS : régions 03, 04, 05, 08, 11, 14 et 16; pour les CS : régions 05 et 12), en raison des objectifs de diffusion de statistiques sous-régionales. Pour ces régions, la taille d'échantillon a été accrue en fonction des objectifs ciblés propres à chacune d'elles, mais sans prévoir une non-réponse éventuelle de la part des écoles ou des classes. Il n'a en effet pas été possible de gonfler davantage la taille d'échantillon à l'échelle de ces RLS ou CS de façon à tenir compte d'une non-réponse éventuelle, comme cela a été fait à l'échelle régionale, car cette augmentation aurait été trop coûteuse[15]. L'ajout de classes en cours de collecte était par conséquent une option à envisager, en dernier recours, dans l'éventualité d'une non-réponse trop importante de la part des écoles ou des classes. En tout, 20 classes ont dû être ajoutées, mais seulement une fois que toutes les démarches et tentatives raisonnables pour convaincre une école ou une classe de participer ont été épuisées[16].

Les objectifs de précision ciblés conjugués aux effectifs scolaires des régions Nord-du-Québec et Gaspésie–Îles-de-la-Madeleine ont fait en sorte que celles-ci ont dû être recensées, c'est-à-dire que les élèves de toutes les classes de la population enquêtée de ces régions ont été sélectionnés. Pour des raisons similaires, quelques écoles des autres régions ont également été recensées. En raison des écarts entre les données d'effectifs scolaires de l'année 2009-2010, sur lesquelles s'est appuyé le plan d'échantillonnage, et celles de l'année scolaire en cours lors de l'enquête (2010-2011), certains ajustements des tailles d'échantillon ont dû être effectués durant la collecte.

Avec un tel échantillon, on pouvait s'attendre à obtenir quelque 65 300 élèves répondants pour l'ensemble des régions. Le tableau 1.4 présente le nombre d'écoles et de classes sélectionnées et admissibles à l'enquête[17] par région, ainsi que le nombre attendu d'élèves répondants correspondant[18]. Le nombre de classes est détaillé en ses trois composantes : le nombre de classes sélectionnées pour l'échantillon de base, le nombre de classes ajoutées pour les régions ayant des objectifs sous-régionaux et le nombre de classes ajoutées ou retranchées pour pallier la non-réponse ou pour faire face aux écarts entre le nombre visé de classes et le nombre réel de classes dans les écoles recensées.

15. Les tailles d'échantillon de ces régions n'ont permis de prévoir que la part de non-réponse des élèves au sein des classes participantes.
16. L'ajout de classes n'a été nécessaire que lorsque la perte de classes répondantes était telle qu'elle compromettait l'atteinte des objectifs de précision visés.
17. Dans la région des Laurentides, des classes ont été sélectionnées initialement dans 32 écoles, pour un total de 484 écoles pour l'ensemble des régions. L'une de ces 32 écoles, pour laquelle une seule classe avait été sélectionnée, a fermé ses portes. Une classe de remplacement a été sélectionnée dans l'une des écoles de l'échantillon de cette région, portant ainsi le nombre d'écoles admissibles à 31 pour cette région et à 483 pour l'ensemble des régions.
18. Ces nombres sont fournis pour l'ensemble des niveaux scolaires.

Tableau 1.4

Nombre total d'écoles et de classes sélectionnées admissibles à l'enquête et nombre attendu d'élèves répondants selon la région sociosanitaire, Québec, 2010-2011

Région sociosanitaire	Nombre d'écoles sélectionnées	Nombre de classes sélectionnées				Nombre attendu d'élèves répondants
		Échant. de base	Échant. supplém. régional	Ajustement non-réponse ou écoles recensées	Échant. total	
01 Bas-Saint-Laurent	21	150	0		150	3 550
02 Saguenay–Lac-Saint-Jean	19	150	0	+1	151	3 550
03 Capitale-Nationale	41	150	50		200	5 050
04 Mauricie et Centre-du-Québec	32	150	49	+1	200	5 050
05 Estrie	25	150	90	+16	256	6 100
06 Montréal	83	150	0		150	3 550
07 Outaouais	23	150	0		150	3 550
08 Abitibi-Témiscamingue	19	150	52	+4	206	5 100
09 Côte-Nord	12	152	0	-4	148	3 550
10 Nord-du-Québec	5	37	0	+3	40	800
11Gaspésie–Îles-de-la-Madeleine	17	150	26	+17	193	4 150
12 Chaudière-Appalaches	33	150	10		160	4 050
13 Laval	18	150	0	+11	161	3 550
14 Lanaudière	28	150	25	+5	180	4 450
15 Laurentides	31	150	0	+3	153	3 550
16 Montérégie	76	150	75		225	5 700
Total	**483**	**2 289**	**377**	**+57**	**2 723**	**65 300**

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

1.5 Méthode de collecte

1.5.1 Instruments de collecte

Le choix des thèmes abordés dans l'EQSJS 2010-2011 résulte des travaux du COP[19]. Une consultation sur les indicateurs du PCS à suivre en priorité dans l'enquête a été préalablement menée par le MSSS auprès de ses partenaires du réseau au printemps 2007. Cet exercice a alimenté, dès le départ, les travaux du COP.

L'EQSJS 2010-2011 a innové en ayant recours à un questionnaire et un système de transmission des données entièrement informatisés. Les élèves ont ainsi rempli un questionnaire autoadministré sur miniportable durant une période de cours. En raison du grand nombre de thèmes à couvrir dans l'EQSJS et du temps limité d'une période de cours[20], deux versions du questionnaire d'une durée moyenne attendue estimée à 32 minutes ont été élaborées. En tenant compte du temps requis pour l'installation des miniportables, la présentation des objectifs de l'enquête et la formulation des consignes aux élèves, ainsi que du fait que certains élèves connaissent des difficultés d'apprentissage, une plage-horaire d'une durée minimale de 45 minutes devait être réservée pour l'obtention des questionnaires complétés. La majorité des questions apparaissent aux deux versions, alors que certaines ne figurent que dans l'une des deux. Chacune des versions a été assignée aléatoirement à environ la moitié des élèves de chaque classe, suivant une approche de sous-échantillonnage. Pour ce faire, l'attribution des questionnaires a été faite à l'aide d'un carton remis aux élèves en même temps qu'un miniportable. Sur ce carton apparaissait un code secret associé à l'une des deux versions. Il suffisait à l'élève d'entrer le code secret pour obtenir son questionnaire informatisé. L'élève pouvait choisir de le remplir en français ou en anglais[21]. Les thèmes

19. Le mandat du COP est de conseiller l'ISQ et le MSSS dans la réalisation de l'EQSJS 2010-2011 afin que celle-ci réponde adéquatement aux besoins d'information du Ministère, du réseau de la santé et des services sociaux et de ses partenaires.
20. La durée d'une période de cours peut être aussi courte que 45 minutes dans certaines écoles.
21. Les questionnaires de l'EQSJS sont disponibles sur le site Web de l'ISQ à l'adresse http://www.stat.gouv.qc.ca.

abordés dans chacune des versions ainsi que le nombre de questions s'y rapportant sont présentés au tableau 1.5.

Tableau 1.5
Nombre de questions posées[1] selon le thème et la version du questionnaire, Québec, 2010-2011

Thème	Version 1	Version 2
Perception de l'état de santé	1	1
Santé respiratoire (sifflements et crises d'asthme)	24	0
Habitudes alimentaires	24	0
Comportements sexuels chez les élèves de 14 ans et plus	13	13
Poids et taille autodéclarés	2	2
Image corporelle	11	0
Santé buccodentaire	2	0
Tabagisme	6	6
Consommation d'alcool	8	8
Consommation de drogues	24	24
Expérience de travail	10	8
Activités physiques de transport	0	4
Activités physiques de loisir	5	5
Santé mentale et problèmes d'adaptation	52	69
Compétences sociales	21	36
Environnement scolaire et intimidation	16	32
Environnement familial	12	12
Caractéristiques du groupe de pairs	6	6
Conditions sociodémographiques	8	12
Caractéristiques sociodémographiques	5	5
Ensemble des sections	**250**	**243**

1. Chacun des items des questions comptant plusieurs items est dénombré.

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

1.5.2 Stratégie de collecte

♦ **Groupes de discussion**

Des groupes de discussion ont été formés afin de vérifier si la formulation de certaines questions était bien comprise par les élèves. Une attention particulière a été apportée aux questions délicates, complexes ou nouvelles, ou encore aux questions adaptées ou reformulées à partir d'un questionnaire existant. Cinq groupes d'environ huit élèves francophones de tous les niveaux scolaires ont été invités au printemps 2009 à remplir un questionnaire en format papier, puis à discuter de leur compréhension des questions.

♦ **Enquête-pilote**

Une enquête-pilote a été réalisée à l'automne 2009 afin de valider non seulement le questionnaire, sa durée et les taux de réponse attendus, mais aussi la procédure de listage des classes et la procédure de collecte informatisée en classe. Un total de 961 élèves répartis dans 43 classes de différents types d'école (privée ou publique, de langue d'enseignement française ou anglaise, de petite ou grande taille, d'un secteur plus ou moins défavorisé) des régions de Montréal, de l'Outaouais et de Laval y ont participé. Les principaux ajustements apportés à l'enquête consistent dans un meilleur arrimage de la procédure de listage des groupes en fonction des données disponibles dans les écoles ainsi que dans le recours à un serveur local pour le stockage temporaire des données dans la classe et leur transmission subséquente à l'ISQ.

♦ **Formation des intervieweurs**

Une formation a été donnée aux cinq intervieweurs téléphoniques assignés à l'enquête afin de bien expliquer les tenants et les aboutissants de la collecte des renseignements auprès du responsable de l'école. La dénomination des types de parcours n'étant pas uniforme dans toutes les écoles, il arrivait que le responsable ait de la difficulté à fournir les renseignements demandés. Connaissant les enjeux de cette collecte, les intervieweurs ont pu contribuer à bonifier la procédure. Par ailleurs, une formation intensive axée sur la compréhension du questionnaire et le fonctionnement de la collecte informatisée a été donnée à 20 équipes d'intervieweurs (composées de deux personnes chacune) qui devaient aller dans les écoles. Celles-ci ont alors pu mettre en pratique toutes les étapes de déploiement des outils informatiques avant de se rendre en classe pour une première fois. Un support téléphonique en temps réel a également été mis à leur disposition lors de leur présence sur le terrain.

♦ **Déroulement de la collecte**

À la suite de l'obtention du consentement passif[22] de toutes les commissions scolaires, les directions des écoles sélectionnées ont été sollicitées pour participer à l'EQSJS 2010-2011. Avec leur approbation, la sélection des classes dans les écoles s'est effectuée par interview téléphonique assistée par ordinateur (ITAO). L'énumération des classes était ainsi supportée par un système informatisé détaillé permettant de minimiser les omissions de classes, ce problème s'étant posé dans l'*Enquête québécoise sur le tabac, l'alcool, la drogue et le jeu chez les élèves du secondaire, 2008* (ETADJES; Dubé et autres, 2009). La sélection aléatoire des classes était également informatisée. Pour faciliter l'échange d'informations avec les écoles, les intervieweurs téléphoniques ont utilisé, dans certains cas plus complexes, des tableaux de compilation pour le listage des classes. La collecte de données a débuté le 13 octobre 2010.

Après la prise de rendez-vous, une équipe composée de deux intervieweurs de l'ISQ se rendait dans chacune des classes sélectionnées pour effectuer la collecte de données en face à face auprès des élèves durant une période de cours. Cette collecte a été réalisée par auto-interview assistée par ordinateur (AIAO). Pour ce faire, les élèves étaient munis de miniportables (« Netbooks »). Chaque élève répondait à un questionnaire autoadministré, anonyme et informatisé. Afin d'optimiser la qualité des données, on demandait à l'enseignant de ne pas circuler dans la classe afin de ne pas influencer les réponses des élèves et de protéger la confidentialité de celles-ci. En moyenne, les élèves ont rempli leur questionnaire en 34 minutes. Les données recueillies étaient ensuite relayées à distance dans une base de données de l'ISQ au moyen d'un serveur et d'un routeur. La quantité et le volume du matériel à apporter sur place ont exigé la présence de deux intervieweurs par classe.

Certaines équipes ont éprouvé des problèmes avec l'équipement informatique et ont dû se rabattre sur l'utilisation des questionnaires en format papier, surtout en début de collecte. Cela s'est produit pour environ 5 % des répondants, cette proportion se situant sous les 10 % dans chacune des régions sociosanitaires. L'impact de l'utilisation des questionnaires papier sur la qualité des données et la comparabilité régionale peut par conséquent être considéré comme négligeable. Par ailleurs, comme pour les questionnaires électroniques, l'attribution de l'une et l'autre des deux versions du questionnaire aux élèves a été faite aléatoirement, de manière à obtenir la moitié des répondants pour chacune d'elles. Les questionnaires en format papier étaient également disponibles en langue française et anglaise.

La collecte dans écoles s'est déroulée du 2 novembre 2010 au 17 mai 2011. La période de collecte a été divisée en trois vagues dans le but d'en optimiser la répartition sur toute la durée dans chacune des régions, ce qui a favorisé la comparabilité régionale. Au moins une école de chacune des commissions scolaires a par ailleurs été sollicitée dès la première vague de collecte, afin de s'assurer de leur participation en début d'enquête.

22. Un consentement passif signifie que l'ISQ a informé, par lettre, la commission scolaire de la tenue de l'enquête dans certaines de ses écoles et que cette dernière ne s'y est pas opposée.

Tout au long de la collecte, un suivi étroit des refus à l'échelle des écoles ou des classes a été effectué pour s'assurer de l'atteinte des objectifs de précision. Lorsque celle-ci était menacée, des tentatives intensives de récupération des écoles perdues ont été entreprises auprès des instances de l'éducation, par l'intermédiaire des directions régionales de santé publique concernées. Ces interventions ont généralement porté fruit et constitué un élément-clé dans l'atteinte des objectifs de l'EQSJS 2010-2011. La démarche de mobilisation des réseaux de la santé et de l'éducation, en amont de la collecte, ainsi que la campagne de promotion de l'enquête ont aussi largement concouru au succès de l'opération.

1.5.3 Validation et saisie

Le recours au questionnaire informatisé permet non seulement une transmission rapide des données à l'ISQ mais également une meilleure qualité des données. D'abord, les erreurs de saisie sont minimisées, la presque totalité des questions étant autocodées et l'élève saisissant lui-même ses réponses. En outre, bien que le questionnaire comporte plusieurs sauts de question lorsque celles-ci ne s'adressent qu'à un sous-groupe d'élèves (par exemple, la section du questionnaire portant sur les relations sexuelles s'adresse uniquement aux élèves de 14 ans et plus), il demeure simple à remplir car ces sauts sont automatisés.

Les données des quelque 3 000 questionnaires remplis en format papier ont été saisies à l'aide du questionnaire informatisé lui-même. Puisqu'il n'était pas prévu de saisir les données d'une aussi grande quantité de questionnaires, une simple saisie a été effectuée. Afin de s'assurer de sa qualité, une double-saisie a été faite de façon indépendante pour 200 questionnaires sélectionnés aléatoirement (en s'assurant de représenter tous les auxiliaires de saisie) et la cohérence entre les deux valeurs saisies a été étudiée. Cette analyse a permis de confirmer la qualité des données saisies; toutefois, quelques variables pour lesquelles la qualité était plus faible (par exemple, le code postal) ont été revérifiées pour tous les questionnaires papier et corrigées au besoin.

Une validation de cohérence entre les différentes réponses de chacun des répondants a été effectuée *a posteriori*. Elle a permis de relever quelques erreurs; les réponses ayant été invalidées ont été corrigées ou remplacées par une valeur manquante selon le cas. 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 donnée a également été vérifiée. Les réponses aux questions ouvertes (par exemple, la catégorie « autre » de certaines questions) ont été reclassées dans les catégories existantes lorsque cela était pertinent.

La validation du niveau scolaire déclaré par les élèves a fait l'objet d'une vérification intensive étant donné son importance tant au sujet des ajustements de la pondération aux effectifs de population (voir la section 2.2) qu'au regard de l'analyse proprement dite. Une attention particulière a été accordée aux élèves dont le niveau déclaré ne concordait pas avec celui attribué à leur classe pour la sélection de l'échantillon (voir la section 1.3.2). Notons qu'il était difficile pour certains élèves de déterminer leur niveau, principalement dans le cas des classes de formation axée sur l'emploi. Lorsque cette difficulté était soulevée par un élève, l'intervieweur avait comme consigne de l'orienter de façon à répondre suivant la nomenclature du MELS[23]. Dans les classes de formation axée sur l'emploi (classes identifiées d'après les renseignements fournis par le responsable), le niveau scolaire attribué à une classe lors de la sélection a été imputé comme niveau déclaré à tous les élèves, peu importe qu'ils aient déclaré ou non un niveau scolaire. Règle générale, dans les classes associées aux autres types de parcours, le niveau scolaire déclaré par l'élève a été conservé tel quel, sauf s'il semblait erroné. Il en résulte que seulement 1 % des élèves ont vu leur niveau déclaré modifié lors de cette opération de validation. C'est la variable validée de niveau scolaire déclaré par l'élève, et non la variable de niveau scolaire attribué aux classes pour leur sélection, qui est utilisée pour l'estimation par niveau scolaire. Il n'en demeure pas moins que 98 % des valeurs prises par ces deux variables sont égales, ce qui signifie que le niveau scolaire déclaré par l'élève correspond presque dans tous les cas au niveau scolaire attribué à sa classe par le responsable de l'école.

23. Les élèves des classes de FMSS devaient indiquer la 3e secondaire, alors que ceux des classes de FPT devaient indiquer la 3e secondaire s'ils en étaient à la 1re année de cette formation, 4e secondaire s'ils en étaient à la 2e année et 5e secondaire s'ils en étaient à la 3e année.

Les données d'âge d'initiation aux relations sexuelles avec consentement, à l'alcool et aux drogues ont été ajustées ou mises à manquantes lorsqu'elles étaient incompatibles avec les autres variables d'âge d'initiation. Le cas échéant, un examen de cohérence détaillé a été effectué individuellement. Pour un élève donné, les principaux critères de validation qui ont été appliqués globalement étaient les suivants :

- lorsque l'âge à la première relation sexuelle avec consentement n'était pas égal à l'âge d'initiation minimum parmi les trois types de relations (orales, vaginales et anales), la valeur de ce dernier a été imputée comme âge d'initiation à la première relation (peu importe le type), sauf lorsque l'écart était trop grand (typiquement quatre ou cinq ans d'écart);
- lorsque l'âge à la première consommation à vie d'alcool ou de drogues était d'un ou deux ans supérieur à l'âge d'initiation à la consommation régulière, la valeur de l'âge à la première consommation à vie a été imputée comme âge d'initiation à la consommation régulière;
- dans les deux situations, les écarts plus grands étaient examinés plus en détail, cas par cas.

Les données d'âge d'initiation aux relations sexuelles consensuelles ont par ailleurs été remplacées par une valeur manquante lorsque l'élève a déclaré un âge inférieur à huit ans, à partir de l'hypothèse que les relations sexuelles avant cet âge ne sont pas consensuelles. Par ailleurs, ce seuil a été fixé en considérant le fait que l'âge d'initiation aux relations sexuelles peut être très précoce dans des milieux défavorisés, comme le suggère une étude américaine récente[24].

1.6 Résultats de collecte

1.6.1 Nombre de répondants

Un nombre total de 63 196 élèves répartis en 2 651 classes de la 1re à la 5e secondaire ont répondu à l'EQSJS 2010-2011, soit 97 % du nombre attendu de 65 300 élèves. Le nombre de répondants atteint plus de 90 % du nombre attendu pour tous les niveaux scolaires et toutes les régions sociosanitaires (voir tableau 1.4), sauf en Abitibi-Témiscamingue où il est légèrement plus faible. Cela s'explique par une non-réponse à l'échelle des classes un peu plus importante pour cette région et par le fait qu'il n'a pas été possible d'ajouter des classes en cours de collecte, puisqu'il n'y en avait pas d'autre dans le RLS où étaient situées les classes non répondantes. Dans tous les autres cas où la non-réponse pouvait compromettre les objectifs de précision, des classes ont été ajoutées. Le tableau 1.6 présente le nombre d'élèves répondants par région sociosanitaire et niveau scolaire déclaré[25].

24. Pour obtenir quelques faits saillants à propos de l'étude de l'Iowa State University, le lecteur est invité à consulter le site Web suivant : http://archive.news.iastate.edu/news/2009/aug/teensex (page consultée le 18 juillet 2010).
25. Il s'agit ici du niveau scolaire déclaré par l'élève et corrigé au besoin lors de la procédure de validation.

Tableau 1.6
Nombre d'élèves répondants, par région sociosanitaire et niveau scolaire déclaré, Québec, 2010-2011

	Région sociosanitaire	1re sec.	2e sec.	3e sec.	4e sec.	5e sec.	Total
01	Bas-Saint-Laurent	631	694	618	673	784	**3 400**
02	Saguenay–Lac-Saint-Jean	690	676	678	768	670	**3 482**
03	Capitale-Nationale	968	1 013	900	894	954	**4 729**
04	Mauricie et Centre-du-Québec	1 084	1 030	882	895	1 001	**4 892**
05	Estrie	1 063	1 240	1 190	1 179	1 262	**5 934**
06	Montréal	695	688	720	751	798	**3 652**
07	Outaouais	767	735	675	695	753	**3 625**
08	Abitibi-Témiscamingue	901	976	765	858	931	**4 431**
09	Côte-Nord	708	597	631	680	605	**3 221**
10	Nord-du-Québec	167	183	170	132	125	**777**
11	Gaspésie–Îles-de-la-Madeleine	718	749	974	734	629	**3 804**
12	Chaudière-Appalaches	833	749	642	780	794	**3 798**
13	Laval	784	727	615	663	638	**3 427**
14	Lanaudière	896	821	854	920	1 009	**4 500**
15	Laurentides	757	726	745	807	831	**3 866**
16	Montérégie	1 080	1 186	1 088	1 100	1 204	**5 658**
Total		**12 742**	**12 790**	**12 147**	**12 529**	**12 988**	**63 196**

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

Bien que l'attribution des questionnaires ait été faite aléatoirement et de manière à obtenir une moitié des répondants à l'une et l'autre des versions pour chacune des classes, il est arrivé que cela ne fonctionne pas dans certaines classes. Néanmoins, environ la moitié des répondants de chacune des régions ont rempli l'une ou l'autre des versions. Globalement, 50,7 % des répondants ont rempli la version 1 du questionnaire.

1.6.2 Taux de réponse

Pour qu'un élève sélectionné soit répondant à l'enquête, il faut d'abord que le directeur de son école accepte d'y participer. Ensuite, il faut que sa classe soit disponible pour l'entrevue et que le professeur accepte de participer. Finalement, l'élève doit lui-même être présent en classe le jour de l'entrevue, accepter de participer et remplir le questionnaire correctement.

Le calcul du taux de réponse global de l'enquête tient compte de la non-réponse observée à chacun de ces niveaux de participation, soit l'école, la classe et l'élève. Ce taux a été évalué par niveau scolaire puisque la sélection a été faite de façon indépendante pour chacun des niveaux et qu'une école n'était pas nécessairement sélectionnée pour tous les niveaux scolaires. Par exemple, une école peut avoir été sélectionnée uniquement pour les 2e et 3e secondaire. Elle doit donc être considérée dans le calcul de la non-réponse des écoles seulement pour ces deux niveaux. À l'instar des taux de réponse généralement diffusés par l'ISQ, les taux de réponse présentés pour l'EQSJS 2010-2011 sont pondérés. Ils tiennent compte, entre autres facteurs, de la non-proportionnalité régionale de l'échantillon de l'enquête.

La non-réponse observée à l'échelle des écoles, évaluée par niveau scolaire, est telle que 96,9 % de l'ensemble des élèves des combinaisons d'école et de niveau scolaire échantillonnées[26] font partie de combinaisons répondantes (ratio pondéré)[27]. La proportion pondérée de classes répondantes, parmi les classes sélectionnées des écoles répondantes, est également très élevée, soit de 99,8 %. C'est au niveau des élèves que la non-réponse est légèrement plus importante : 91,1 % des élèves inscrits des classes répondantes ont rempli le questionnaire[28]. En multipliant les proportions observées à chacune des trois échelles, on obtient un taux de réponse global pondéré de 88,1 %. Le tableau 1.7 présente les taux de réponse globaux pondérés, par niveau scolaire.

26. On fait ici référence à tous les élèves inscrits dans les combinaisons d'école et de niveau échantillonnées, que ces élèves fassent partie ou non de classes sélectionnées pour l'enquête.
27. Cette quantité varie de 95,8 % à 98,1 % d'un niveau scolaire à l'autre.
28. Cette proportion varie peu d'un niveau scolaire à l'autre, soit de 89,2 % à 92,8 %.

Tableau 1.7
Taux de réponse pondéré, par niveau scolaire, Québec, 2010-2011

Niveau scolaire	Taux de réponse (%)
1re secondaire	91,0
2e secondaire	89,2
3e secondaire	85,5
4e secondaire	88,3
5e secondaire	86,8
Total	**88,1**

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

Le tableau 1.8 montre quant à lui les taux de réponse globaux régionaux : ces derniers se situent tous au-delà de 86 %, sauf pour la région de Laval qui présente un taux plus faible (73,5 %) dû à une non-réponse plus importante au niveau des écoles. On a partiellement pallié cette non-réponse par l'ajout de classes. Au niveau des classes, la non-réponse est très faible dans toutes les régions, quoique légèrement plus importante dans celle de l'Abitibi-Témiscamingue. Finalement, au niveau des élèves, les proportions d'élèves répondants au sein des classes répondantes varient peu, soit de 88,6 % à 93,2 % d'une région à l'autre.

Tableau 1.8
Taux de réponse pondéré, par région sociosanitaire, Québec, 2010-2011

	Région sociosanitaire	Taux de réponse (%)
01	Bas-Saint-Laurent	89,6
02	Saguenay–Lac-Saint-Jean	88,7
03	Capitale-Nationale	86,1
04	Mauricie et Centre-du-Québec	88,6
05	Estrie	93,2
06	Montréal	87,0
07	Outaouais	88,6
08	Abitibi-Témiscamingue	87,1
09	Côte-Nord	89,4
10	Nord-du-Québec	92,0
11	Gaspésie–Îles-de-la-Madeleine	90,9
12	Chaudière-Appalaches	89,8
13	Laval	73,5
14	Lanaudière	90,7
15	Laurentides	88,3
16	Montérégie	91,1
Total		**88,1**

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

1.6.3 Période de collecte

En dépit de la répartition de l'échantillon en vagues de collecte, les contraintes de terrain ont fait en sorte que la distribution estimée de l'âge des élèves varie quelque peu entre les régions sociosanitaires. En particulier, le Nord-du-Québec et la Côte-Nord présentent une proportion légèrement plus élevée d'élèves de 12 ans et moins que les autres régions (14,0 % et 13,4 % respectivement par rapport à 11,5 % globalement). Cette situation s'explique en partie par le fait que la collecte a eu lieu plus tôt dans ces régions, surtout dans celle du Nord-du-Québec où 75 % des répondants ont rempli le questionnaire en 2010, comparativement à 36 % pour l'ensemble des régions. Toutefois, puisque les légers écarts observés dans les répartitions d'âge d'une région à l'autre ne suivent pas nécessairement les variations des périodes de collecte, il est possible de croire qu'une part de ces écarts d'âge est inhérente aux régions. La standardisation ou non des statistiques régionales selon l'âge à des fins de comparaison relève par conséquent de l'angle d'analyse souhaité. Quel que soit l'angle retenu, il importe de garder à l'esprit que la région du Nord-du-Québec a connu une collecte sensiblement plus hâtive et que des précautions doivent être prises lors de sa comparaison avec les autres régions. La section 4 abordera la question de la standardisation des proportions selon l'âge de l'élève.

2. Estimation des proportions et des effectifs

2.1 Définition

Deux types d'estimations peuvent être calculés à l'Infocentre de santé publique à partir des données de l'EQSJS 2010-2011 : des effectifs et des proportions. L'estimation d'effectifs est utilisée lorsqu'on désire savoir combien d'élèves de la population étudiée possèdent une caractéristique donnée. Celle-ci est obtenue en faisant la somme des poids populationnels de ces élèves[29]. Si l'on désire plutôt estimer la proportion de la population visée qui possède telle caractéristique, on doit diviser l'estimation de l'effectif la possédant par celui de l'ensemble des élèves de la population, qu'ils possèdent ou non cette caractéristique.

Parce qu'elles sont calculées à partir de données d'enquête, chacune de ces estimations doit être pondérée afin de produire des résultats pouvant être inférés à la population visée par l'EQSJS 2010-2011[30]. Il s'agit, pour ce faire, d'attribuer à chaque répondant à l'enquête une valeur, un poids qui correspond au nombre d'élèves qu'il « représente » au sein de la population visée. Ce poids est appelé « poids populationnel ». Il doit tenir compte tant de la probabilité de sélection de l'élève, prédéterminée par le plan d'échantillonnage, que de la non-réponse à l'échelle des écoles, des classes et des élèves. Le poids n'inclut toutefois aucun ajustement pour la non-réponse partielle. Or, pour un répondant à l'enquête, la présence de non-réponse à la question analysée, appelée non-réponse partielle, peut entraîner un biais dans l'estimation. L'ampleur de ce biais potentiel 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

Dans l'EQSJS 2010-2011, les objectifs de diffusion à l'échelle régionale entraînent une variation importante des probabilités de sélection et font en sorte que la répartition de l'échantillon selon les régions sociosanitaires ne représente pas bien celle de la population, d'où la nécessité de tenir compte de la pondération dans toutes les estimations. Ce faisant, on corrige en quelque sorte la distorsion créée par le plan d'échantillonnage utilisé pour la sélection de l'échantillon. De plus, en l'absence d'une liste exhaustive de toutes les classes des écoles composant la population lors de la planification de l'enquête, le plan d'échantillonnage a inévitablement résulté en probabilités de sélection variables, et ce, bien qu'il tendait à une probabilité de sélection des élèves le plus uniforme possible pour une région et un niveau scolaire donnés. Finalement, la non-réponse observée au niveau des écoles, des classes et des élèves a touché davantage certains groupes de répondants, ce qui a nécessité un ajustement de la pondération visant à contrer un biais de non-réponse potentiel. Il est donc nécessaire de tenir compte de tous ces éléments, qui doivent être intégrés à la pondération, lors de la production de statistiques tirées de l'EQSJS 2010-2011. L'annexe 1 peut être consultée pour savoir comment utiliser la pondération.

En raison de l'existence de deux versions du questionnaire, trois pondérations sont requises pour l'analyse :

- une pondération principale, portant sur l'ensemble des répondants, aux fins de l'analyse des questions communes aux deux versions;
- une pondération pour les sous-groupes de répondants à chacune des deux versions du questionnaire, aux fins de l'analyse des questions qui n'apparaissent que dans une version, ou d'un croisement de celles-ci avec une question commune aux deux versions.

29. Cette estimation est effectuée en supposant qu'il y a absence de non-réponse partielle. Dans le cas contraire, un ajustement est fait pour redistribuer la non-réponse partielle. Pour plus de détails, consulter la section 2.3.
30. Il est à noter que l'estimation d'un effectif ou d'une proportion pourra varier légèrement pour une variable commune aux deux versions du questionnaire selon que cette dernière est croisée ou non avec une variable apparaissant dans une seule version, et ce, en raison de l'application de l'une ou de l'autre des pondérations pour l'estimation (voir section 2.2).

L'établissement de chacune des pondérations comporte plusieurs étapes. Ces étapes ont été réalisées indépendamment par niveau scolaire. Le poids de départ d'un élève est un poids au niveau de l'école égal à l'inverse de la probabilité de sélection de son école. Afin de tenir compte de la non-réponse de certaines écoles, ce poids est par la suite divisé par le ratio du nombre pondéré d'élèves inscrits des écoles répondantes sur le nombre pondéré d'élèves inscrits des écoles échantillonnées. Cette opération tient compte de la non-réponse plus élevée dans la région de Laval pour les années du 2e cycle d'enseignement (ajustement pour la non-réponse par le ratio; Haziza et autres, 2010). On obtient ainsi un « poids école », c'est-à-dire qui correspond approximativement au nombre d'écoles de la population enquêtée représentées par une école répondante, pour un niveau scolaire donné.

En second lieu, le poids école est divisé par la probabilité de sélection de la classe du répondant, puis par la proportion pondérée de classes répondantes, en tenant compte de la non-réponse plus élevée observée en Abitibi-Témiscamingue. On obtient alors un poids au niveau de la classe.

La modélisation de la non-réponse des élèves permet par la suite de former des sous-groupes d'élèves sur la base du type de parcours scolaire de la classe tel que déclaré par le responsable de l'école, du réseau et de la langue d'enseignement de l'école, en considérant la non-réponse plus élevée observée dans une école ayant exigé un consentement actif de la part des parents en début de collecte. Conséquemment, la troisième étape consiste à diviser le poids du répondant par le ratio du nombre pondéré d'élèves répondants sur le nombre pondéré d'élèves inscrits des classes répondantes, au sein du sous-groupe auquel il appartient (ajustement par le ratio). On obtient ainsi finalement un poids au niveau de l'élève, uniforme pour tous les élèves d'une même classe.

Or il appert que les poids des élèves de certaines classes sont très élevés relativement aux poids des élèves des classes du même niveau scolaire et de la même région ou du même RLS, selon le cas. Certains poids ont par conséquent été abaissés afin d'éviter qu'une classe puisse avoir une influence indue sur les statistiques. Les poids ont ainsi dû être modifiés pour seulement 0,3 % des répondants. La dernière étape de pondération consiste à ajuster les poids aux effectifs scolaires du MELS de l'année 2010-2011. Cet ajustement a été réalisé par un calage aux marges[31] visant à obtenir une somme de poids égale aux effectifs scolaires de la population visée par région, niveau scolaire déclaré et sexe, d'une part, et par RLS ou CS et cycle d'enseignement secondaire, d'autre part, pour les régions ayant des objectifs de diffusion à l'échelle sous-régionale[32].

Le même type de traitement est appliqué pour les trois pondérations. Toutefois, pour les pondérations propres à l'analyse des questions qui n'apparaissent que dans l'une des deux versions du questionnaire, une étape est ajoutée pour tenir compte de la probabilité de sélection des élèves des classes répondantes. Cette probabilité est égale à un pour la pondération principale mais elle varie autour d'une demie pour les pondérations propres à l'analyse des questions qui n'apparaissent que dans une version. En l'absence de répondants pour une des deux versions dans certaines classes[33], des ajustements ont aussi dû être apportés aux probabilités de sélection des écoles ou des classes.

Le tableau 2.1 présente un aperçu des étapes ayant mené à la création des pondérations.

31. Le *calage aux marges* est une procédure qu'on peut appliquer pour incorporer des données auxiliaires. Il rajuste les poids d'échantillonnage au moyen de multiplicateurs appelés les *facteurs de calage*, lesquels font correspondre les estimations aux totaux connus. Dans cette enquête-ci, les données auxiliaires sont les effectifs scolaires du MELS pour l'année scolaire 2010-2011.
32. Contrairement aux étapes de pondération antérieures, le niveau scolaire considéré pour l'étape de calage est celui déclaré par l'élève et validé; il ne correspond pas dans tous les cas au niveau scolaire pour lequel la classe a été sélectionnée selon le plan d'échantillonnage.
33. En dépit de la procédure d'attribution aléatoire de chacune des deux versions du questionnaire à la moitié de la classe, il est arrivé dans quelques classes qu'une seule version du questionnaire soit utilisée.

Tableau 2.1
Stratégie de pondération de l'EQSJS 2010-2011

Niveau scolaire		
1re secondaire	**…**	**5e secondaire**
Poids initial (sélection école)		Poids initial (sélection école)
Ajustement non-réponse école		Ajustement non-réponse école
Création poids classe (sélection classe)		Création poids classe (sélection classe)
Ajustement non-réponse classe		Ajustement non-réponse classe
Création poids élève (sélection élève)[1]		Création poids élève (sélection élève)[1]
Ajustement non-réponse élève		Ajustement non-réponse élève
Ajustement poids extrêmes		Ajustement poids extrêmes
Calage aux marges (effectifs scolaires 1re sec.)		Calage aux marges (effectifs scolaires 5e sec.)

1. Cette étape est nécessaire seulement pour l'analyse des questions apparaissant dans une seule version.

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

2.3 Non-réponse partielle

2.3.1 Définition

La pondération tient compte de la non-réponse totale mais n'a pas été ajustée pour la non-réponse partielle. Celle-ci se produit lorsqu'une variable comporte des données manquantes pour les répondants à l'enquête. Il est connu qu'une non-réponse partielle importante peut entraîner certains biais dans les estimations, au même titre que la non-réponse totale, s'il s'avère que les non-répondants présentent des caractéristiques différentes de celles des répondants et que ces caractéristiques sont de surcroît liées au thème abordé. Le taux de non-réponse partielle pondéré est défini comme le rapport du nombre pondéré de non-répondants sur le nombre pondéré d'élèves admissibles à répondre.

Plus ce taux est élevé, plus le risque de biais induits par la non-réponse partielle est grand. Pour l'EQSJS 2010-2011, on fait l'hypothèse qu'une non-réponse partielle inférieure à 5 % a un impact négligeable, en général, sur les estimations à l'échelle nationale, sauf dans la situation où elle est fortement caractérisée. C'est dire que l'on suppose en général que les risques de biais occasionnés par un débalancement des caractéristiques des non-répondants par rapport à celles des répondants sont faibles. Pour l'estimation d'une proportion, il est par conséquent acceptable dans ce cas de redistribuer les non-répondants partiels dans les différentes catégories d'une variable, proportionnellement à la répartition observée chez les répondants, sans tenir compte des caractéristiques différenciant les répondants des non-répondants (voir section 2.3.4). Cette approche est valable 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.

2.3.2 Taux de non-réponse partielle

Les variables 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-estimation ou surestimation). Pour savoir comment procéder au diagnostic et à l'analyse qualitative de la non-réponse partielle, le lecteur est invité à consulter l'annexe 2.

En raison des faibles taux de non-réponse partielle observés pour la grande majorité des variables de l'enquête, les risques de biais, relativement à la précision des estimations, sont assez faibles. Malgré tout, les estimations d'effectifs doivent prendre en compte l'effet de la non-réponse partielle, s'il y a lieu, et ce, peu importe le taux observé[34].

34. L'ajustement pour la non-réponse partielle dans les estimations d'effectifs produites par l'ISQ a été fait à partir de l'hypothèse que le taux de non-réponse partielle est uniforme pour toutes les catégories de la variable étudiée.

Les variables d'analyse qui présentent une non-réponse partielle supérieure à 5 % sont la taille autorapportée (10 %), l'indice de masse corporelle et le statut pondéral (11 % dans les deux cas), la confirmation de l'asthme par un médecin (13 %), le nombre de facteurs déclenchant les crises d'asthme (11 %), de même que le nombre de partenaires à vie et l'utilisation du condom lors des dernières relations sexuelles anales consensuelles chez les 14 ans et plus (10 % dans les deux cas). Quelques variables de croisement affichent également une non-réponse partielle élevée : les variables du plus haut niveau de scolarité de la mère (11 %) et du père (12 %) considérées séparément, l'indicateur du plus haut niveau de scolarité des parents[35] (16 %), l'occupation principale de la mère (5,1 %) et du père (6,0 %), le statut d'emploi des parents (8,2 %), le lieu de naissance du père biologique (5,9 %), le lieu de naissance des parents (6,8 %), de même que les indices de défavorisation matérielle et/ou sociale (15 % dans les trois cas). La non-réponse importante observée pour ces derniers indices est due aux données manquantes relatives au code postal de l'élève, puisque ce dernier entre dans leur composition (voir ISQ, 2013). Les variables d'analyse ont fait l'objet d'un examen visant à déceler des biais potentiels induits par cette non-réponse. Les deux variables de croisement présentant les taux les plus élevés ont également été étudiées. Ces travaux seront présentés à la section 2.3.3.

Par ailleurs, quelque 2 000 élèves de 14 ans et plus ont sauté les questions de la section sur la sexualité HV7.1 à HV7.11, ainsi qu'il leur était suggéré de faire s'ils n'étaient pas à l'aise d'y répondre. Bien que les taux de non-réponse partielle des variables concernées se situent sous la barre des 5 % parmi l'ensemble des élèves de 14 ans et plus, cette non-réponse pose problème puisque les questions en cause s'adressaient exclusivement aux élèves ayant déclaré avoir déjà eu des relations sexuelles. Si l'on n'en tient pas compte, la non-réponse a pour effet d'induire des biais non négligeables pour les statistiques portant sur l'ensemble des élèves de 14 ans et plus qui ont déjà eu ou non des relations sexuelles[36]. Cette dernière variable a par conséquent fait l'objet d'une imputation, décrite à la section 2.3.3.

Une situation similaire se produit pour l'indicateur « Proportion des élèves du secondaire ayant consommé de la drogue au cours des 12 derniers mois ». La construction de cet indicateur est basée sur la réponse à neuf questions, soit la fréquence de consommation de différents types de drogues au cours de cette période (voir ISQ, 2013 pour la définition de cet indicateur). Seuls les élèves ayant répondu au préalable avoir déjà consommé de la drogue au cours de leur vie avaient à répondre à ces questions. Bien qu'il n'y ait qu'environ 700 élèves qui n'ont pas répondu à toutes les questions, la non-réponse pourrait induire un biais dans l'estimation si leur statut de consommation était simplement laissé manquant, puisqu'ils avaient presque tous déclaré avoir consommé au moins un type de drogue au cours des 12 derniers mois. Une imputation a donc été réalisée pour minimiser l'impact de la non-réponse.

Pour les situations décrites précédemment, différentes actions ont été prises afin de minimiser l'impact d'une non-réponse partielle non négligeable ou d'en évaluer l'importance. Elles sont présentées à la section 2.3.3.

Il est à noter qu'à l'échelle régionale ou locale, en raison de la moins bonne précision des estimations, on considère que les risques de biais sont faibles pour un taux de non-réponse partielle de moins de 10 %.

35. On fait ici référence au plus élevé des deux niveaux de scolarité atteints par les parents de l'élève ou au niveau de scolarité du parent seul.

36. Il est à noter que pour le sous-groupe des élèves de 14 ans et plus ayant déjà eu des relations sexuelles, la non-réponse partielle atteint 12 %. Bien que l'impact de la non-réponse sur les statistiques portant sur ce sous-groupe soit moins important, il reste que les non-répondants sont proportionnellement plus nombreux parmi ceux chez qui la première relation sexuelle consensuelle est plus récente dans leur vie. Cela pourrait également induire de légers biais.

2.3.3 Analyse de la non-réponse partielle et imputation

L'imputation des données manquantes est une solution souvent mise de l'avant pour minimiser l'impact de la non-réponse partielle dans les enquêtes. Pour une variable présentant de la non-réponse partielle, l'imputation consiste à attribuer une valeur à un élève en remplacement d'une donnée manquante, en se basant sur les renseignements disponibles au sujet de cet élève[37]. Une imputation valable requiert toutefois la présence de variables corrélées aux variables à imputer et qui ne présentent pas elles-mêmes une non-réponse partielle importante.

Pour pallier le problème de non-réponse de l'indicateur du plus haut niveau de scolarité des parents, qui est utilisé comme variable de croisement, les données des non-répondants partiels ont été imputées pour les élèves ayant déclaré le niveau de scolarité d'un des parents, en utilisant cette variable dans le modèle d'imputation. Aucune imputation n'a été faite pour les autres non-répondants partiels puisqu'on ne pouvait faire de lien valable avec aucune autre variable mesurée dans l'enquête. Cela a permis d'abaisser le taux de non-réponse partielle de 16 % à 9 % (après imputation). Toutefois, la comparaison de la répartition de cette variable selon le niveau scolaire après imputation donne à croire qu'il n'a pas été possible de faire une imputation permettant de bien différencier les niveaux de scolarité « études collégiales » et « études universitaires ». Ces deux catégories de l'indicateur imputé ont par conséquent été regroupées pour l'analyse.

La seule variable qui semble corrélée au statut pondéral, soit la perception de son propre poids, apparaît uniquement dans la version 1 du questionnaire et ne peut donc être utilisée judicieusement pour l'imputation. Aussi a-t-il été décidé de réaliser une étude de biais pour cet indicateur. En plus de la perception du poids, plusieurs variables socioéconomiques sont liées à la fois à la non-réponse et au statut pondéral. Lorsque ces variables sont considérées simultanément, le sexe et le niveau scolaire, en plus de la perception du poids, figurent parmi les plus liées au statut pondéral et à la non-réponse. En particulier, on observe que les élèves percevant leur silhouette comme très mince ou obèse, les filles et les élèves de plus bas niveau scolaire ont répondu en moins grande proportion que les autres. En considérant ces trois variables, la non-réponse partielle a été redistribuée au sein des catégories du statut pondéral, en tenant compte des caractéristiques des non-répondants. La distribution estimée ainsi obtenue s'avère très similaire à la distribution estimée en redistribuant les non-répondants partiels sans tenir compte de leurs caractéristiques. Cette analyse suggère que la non-réponse partielle n'a qu'un impact négligeable sur la répartition estimée du statut pondéral.

Lorsque l'on étudie la non-réponse partielle des variables du nombre de partenaires à vie et de l'utilisation du condom lors des dernières relations sexuelles anales consensuelles, on note que le sexe et le niveau scolaire sont liés à la fois à ces variables et à la non-réponse, ce qui entraîne une sous-estimation potentielle de la proportion d'élèves ayant eu plusieurs partenaires ainsi que de la proportion d'élèves ayant porté le condom. Or, en prenant en compte le sexe et le niveau scolaire, la répartition de ces deux variables a été ajustée en redistribuant les non-répondants suivant leurs caractéristiques. Cette analyse suggère que l'impact de la non-réponse partielle sur l'estimation de ces répartitions est négligeable.

L'impact de la non-réponse partielle sur les indicateurs de confirmation de l'asthme par un médecin et du nombre de facteurs déclenchant les crises peut également être considéré comme négligeable. En effet, aucune des caractéristiques socioéconomiques étudiées n'est liée simultanément à la non-réponse et aux indicateurs.

Une analyse de biais a également été réalisée pour la non-réponse partielle observée pour les trois indices de défavorisation. Contrairement au statut pondéral, ces indices sont des variables de croisement. Ils revêtent une importance particulière puisqu'aucune mesure de revenu des parents n'est recueillie dans l'EQSJS 2010-2011. On note en particulier une non-réponse beaucoup plus importante pour la région des Laurentides (46 % c. 15 % globalement) ainsi que chez les élèves de parcours autres que celui de

37. Ces informations sur les non-répondants partiels peuvent provenir du questionnaire d'enquête lui-même, ou encore de la base de sondage.

formation générale (25 %). La non-réponse et les indices de défavorisation sont par ailleurs tous deux liés au nombre de pièces dans la maison ou l'appartement où vit l'élève, de même qu'à plusieurs autres variables socioéconomiques, dont le type de parcours et la langue d'enseignement. Lorsque la non-réponse partielle est redistribuée en tenant compte des caractéristiques des non-répondants, la distribution des trois indices de défavorisation est peu modifiée. Il n'en demeure pas moins que cette non-réponse peut affecter la distribution estimée d'une variable d'analyse croisée avec un indice de défavorisation. Le cas échéant, il est recommandé de comparer la distribution de la variable d'analyse estimée avec ou sans les non-répondants partiels à l'indice de défavorisation. Un écart non négligeable indique un biais dû à la non-réponse partielle dans l'estimation de la distribution globale, biais qui pourrait se répercuter plus ou moins fortement sur les distributions estimées par catégorie de défavorisation. La prudence est de mise dans l'interprétation d'un tel croisement. En présence d'un résultat significatif pour lequel le seuil observé est près de 5 %, il est recommandé de nuancer davantage l'interprétation, puisque ce seuil aurait pu être supérieur à 5 % en l'absence de non-réponse partielle. Par ailleurs, en présence d'un résultat significatif et d'un biais modéré, il est recommandé de ne pas rapporter les quantités estimées, mais de discuter uniquement des différences observées de manière qualitative.

Par ailleurs, les quelque 2 000 élèves ayant sauté les questions de la section sur la sexualité ont vu leurs valeurs manquantes imputées. Un donneur a été choisi aléatoirement parmi les élèves répondants ayant des caractéristiques similaires et toutes ses réponses aux questions HV7.1 à HV7.11 ont été imputées au non-répondant partiel. Cette opération a été répétée pour chacun des élèves concernés. Cette imputation a été réalisée en dépit du fait que les taux de non-réponse partielle des variables avant imputation se situaient sous la barre des 5 % afin de minimiser les biais potentiels sur les statistiques portant sur l'ensemble des élèves de 14 ans et plus.

L'indicateur « Proportion des élèves du secondaire ayant consommé de la drogue au cours des 12 derniers mois » a également fait l'objet d'une imputation malgré une non-réponse partielle faible globalement. En effet, puisque l'on pouvait confirmer que plus de 90 % des non-répondants partiels sont des consommateurs sur 12 mois, il a été décidé de leur imputer cette valeur dans la construction de l'indicateur. La valeur des quelques non-répondants partiels pour lesquels ce statut ne pouvait être confirmé a quant à elle été laissée manquante.

Pour les estimations régionales ou locales tirées de l'EQSJS 2010-2011, la non-réponse partielle n'a pas fait l'objet d'un examen détaillé. Cette tâche revient donc à l'utilisateur. Ce dernier devra évaluer lui-même si l'impact de la non-réponse sur les estimations de proportions peut être considéré comme négligeable. Ce n'est pas le cas lorsqu'un écart important est observé dans les taux de non-réponse partielle entre des sous-groupes et que les écarts observés dans les estimations de proportions au sein de ces sous-groupes sont également importants. Aux fins de l'analyse, l'usager peut faire l'hypothèse que l'impact de la non-réponse partielle à l'échelle régionale est similaire à celui observé à l'échelle nationale, ou encore établir un diagnostic qualitatif de l'impact de la non-réponse suivant la procédure décrite à l'annexe 2.

2.3.4 Traitement pour l'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 guide[38]) comme une réponse possible ou comme une absence de réponse. S'il la considère comme une réponse en soi, il pourra estimer adéquatement la proportion d'élèves qui « ne savent pas » ou qui « 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 devra considérer la non-réponse comme une absence de réponse et ne pas estimer la proportion d'individus de cette catégorie. La non-réponse partielle sera ainsi répartie dans les autres catégories de réponse, ce qui permettra d'estimer correctement la proportion du phénomène étudié.

38. Pour les besoins du guide, ont été regroupées sous l'appellation « NRP » les catégories de non-réponse suivantes : « ne sait pas », « ne répond pas », « inconnu » ou « refus global ».

En général, cela a peu d'impact sur les estimations des proportions d'intérêt puisque la non-réponse partielle est plutôt faible dans l'EQSJS 2010-2011. Malgré tout, pour les estimations de proportions, 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 2.2 et 2.3 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. Cela 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 de l'Infocentre de santé publique à partir des données de l'EQSJS 2010-2011 sont calculées en appliquant la méthode décrite ci-dessus, les proportions du tableau 2.3 (mais pas les effectifs – voir la note au bas du tableau) correspondent aux valeurs que l'on retrouve sur ce portail, comme en témoigne le visuel de la sortie produite à l'Infocentre de santé publique présentée au tableau 2.4.

Tableau 2.2
Effectifs et proportions pondérés des élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

HV7.0 (A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP	1 478	0,5	1 478	0,5
Oui	112 136	36,9	113 614	37,4
Non	190 472	62,6	304 086	100,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 2.3
Effectifs et proportions pondérés des élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

HV7.0 (A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement)	Effectif*	Proportion	Effectif cumulatif	Proportion cumulative
NRP	.	.	.	.
Oui	112 136	37,1	112 136	37,1
Non	190 472	62,9	302 607	100,0

* Les valeurs présentées dans cette colonne ne sont pas des estimations sans biais de l'effectif des élèves, car la non-réponse partielle n'a pas été intégrée aux calculs des effectifs; elle ne l'a été que pour le calcul de la proportion. 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 2.3.5 (tableau 2.4).

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Il a été dit que pour la grande majorité des variables de l'EQSJS 2010-2011, le taux de non-réponse partielle au niveau de l'ensemble du Québec est inférieur à 5 % et que, par conséquent, les risques de biais y étant associés relativement à la précision des estimations sont assez faibles. Ainsi, de façon générale, aucune attention particulière n'est nécessaire au moment d'analyser les données de l'ensemble du Québec, sauf pour les variables analysées à la section 2.3.3.

2.3.5 Traitement pour l'estimation des effectifs

La somme des poids populationnels des élèves qui possèdent une caractéristique donnée fournit une estimation du nombre d'élèves de la population visée qui possèdent cette caractéristique. Or, les poids populationnels ont préalablement été ajustés pour considérer les élèves qui n'ont pas participé à l'enquête (non-réponse totale, voir section 2.2), mais non pour tenir compte du fait que certains élèves n'ont pas répondu à une question donnée (non-réponse partielle, voir section 2.3). S'il y a lieu, la non-réponse partielle entraîne une sous-estimation des effectifs de population d'autant plus importante que le taux de non-réponse partielle est élevé. Des corrections devraient donc être apportées à cet égard, de façon à produire des estimations pour lesquelles les biais sont minimums, quel que soit le taux de non-réponse partielle observé.

La méthode recommandée à la section 2.3.4 pour l'estimation de proportions ne peut donc être appliquée pour l'estimation d'effectifs. En effet, si l'on produisait un tableau sans la catégorie « NRP », la somme des poids s'en trouverait diminuée proportionnellement à la représentativité des poids des « NRP ». Cela mènerait à une sous-estimation des effectifs de population pour ce phénomène. Les poids des non-répondants partiels doivent être redistribués au sein des différentes catégories de la variable d'analyse, proportionnellement à la répartition observée chez les répondants.

Les tableaux 2.2 et 2.3 montrent 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'élèves de la population qui possèdent la caractéristique étudiée. L'explication tient au fait qu'un certain nombre d'élèves n'ayant pas fourni de réponse présentent peut-être cette caractéristique. Ainsi, il n'est pas approprié d'utiliser uniquement la somme des poids populationnels des unités possédant la caractéristique étudiée 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 consiste à multiplier l'effectif estimé en présence de non-réponse partielle par l'inverse du taux de réponse à la question (parmi l'ensemble des élèves qui avaient à répondre à la question). On redistribue ainsi l'effectif de la non-réponse dans les autres catégories de réponse en tenant compte de la distribution de l'effectif des répondants[39]. Les résultats diffusés sur le portail de l'Infocentre de santé publique ont été produits en utilisant cette méthode et correspondent, pour notre exemple, aux estimations du tableau 2.4 (voir le détail des calculs à l'annexe 3)[40].

Tableau 2.4
Proportion des élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

Statistiques					
Nombre	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
112 700	37,1	1,3	0,50	(36,2 – 38,1)	0,5

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

39. Cette façon de faire présuppose que la répartition de l'indicateur chez les non-répondants partiels est égale à celle estimée à partir de l'ensemble des répondants, peu importe leurs caractéristiques respectives.
40. Il est à noter que les estimations d'effectifs produites sur le portail de l'Infocentre de santé publique sont toutes arrondies à la centaine près.

2.3.6 Traitement de la non-réponse en cascade

Lorsque l'on analyse la réponse à une question pour laquelle une sous-population (domaine d'étude) est définie par la réponse à une autre question, il faut considérer l'effet de cascade émanant de la non-réponse partielle à cette question filtre[41]. Par exemple, si l'on souhaite estimer le nombre d'élèves ayant eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, on doit tenir compte non seulement de la non-réponse à la question analysée (tableau 2.6), mais également de la non-réponse à la question filtre, qui définit la sous-population des élèves ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois (tableau 2.5)[42].

Tableau 2.5
Effectifs et proportions pondérés de la question HV6.1 portant sur les sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, pour l'ensemble du Québec, EQSJS 2010-2011

HV6.1 (Au cours des 12 derniers mois, as-tu déjà eu des sifflements dans la poitrine (bruits anormaux ou « râlements » lorsque tu respires), à un moment quelconque?)	Effectif*	Proportion	Effectif* cumulatif	Proportion cumulative
NRP	1 654	.	1 654	.
Oui	78 735	18,3	80 388	18,3
Non	350 712	81,7	431 100	100,0

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 2.6
Effectifs et proportions pondérés de la question HV6.4 portant sur les crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, pour l'ensemble du Québec, EQSJS 2010-2011

HV6.4 (Au cours des 12 derniers mois, ces crises de sifflements ont-elles été si graves qu'elles ont limité ton discours à un ou deux mots maximum entre deux respirations?)	Effectif*	Proportion	Effectif* cumulatif	Proportion cumulative
NRP	218**	.	218	.
Oui	10 345	13,2	10 563	13,2
Non	68 172	86,8	78 735	100,0

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

** Effectifs pondérés des élèves qui ont répondu « NRP »à la question HV6.4 en enlevant les effectifs pondérés des élèves qui ont répondu « NRP » à la question HV6.1.

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Les schémas 1 et 2 présentent la façon utilisée pour traiter la non-réponse partielle pour les données de l'EQSJS 2010-2011, 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 affiche 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

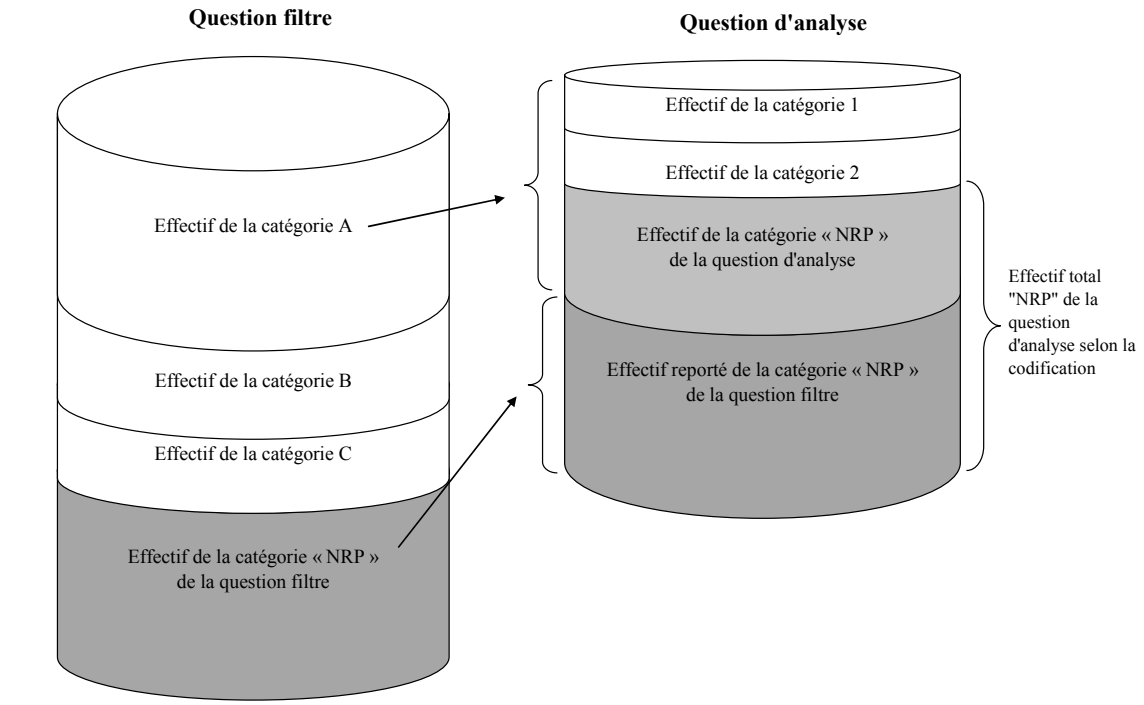
41. Il existe d'autres formes de non-réponse en cascade pour les indicateurs de l'EQSJS qui ne sont pas liés à une question filtre, comme celle pour les indicateurs construits à partir de plus d'une question, mais dont aucune n'est une question filtre. C'est le cas, par exemple, de l'indicateur *Proportion des élèves du secondaire ayant consommé de la drogue au cours des 12 derniers mois*, qui est construit à partir des questions HV_J-9.1 à HV_J-9.9 portant sur la consommation de différents types de drogue au cours des 12 derniers mois. 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.

42. La question HV6.4, qui permet de mesurer ce phénomène au cours des 12 derniers mois, n'a été posée qu'aux élèves ayant d'abord déclaré avoir déjà eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois (question HV6.1).

de chaque catégorie de réponse peut être estimé adéquatement, c'est-à-dire sans biais[43]. Voici d'abord quelques explications sur la codification illustrée par les deux schémas.

Le schéma 1 présente les catégories de réponse originales de la question filtre et de la question d'analyse, alors que le schéma 2 expose ces mêmes catégories une fois le traitement de la non-réponse effectué. Dans cet exemple, la sous-population visée pour la question d'analyse correspond à l'ensemble des unités de la catégorie A. Or, contrairement à ce que la codification peut porter à croire, bien qu'un certain nombre d'élèves n'ayant pas fourni de réponse à la question filtre puissent appartenir à cette sous-population, ces unités n'en font pas toutes nécessairement partie et n'avaient donc par conséquent pas toutes à répondre à la question d'analyse. Ainsi, on ne peut analyser telles quelles les réponses à la question d'analyse, car cela présupposerait que toutes les unités qui sont classées « NRP » à la question filtre font partie de la catégorie A et avaient à répondre à la question d'analyse.

Schéma 1
Catégories de réponse selon la codification des données de l'EQSJS 2010-2011 dans le cas d'une sous-population définie à partir d'une question filtre

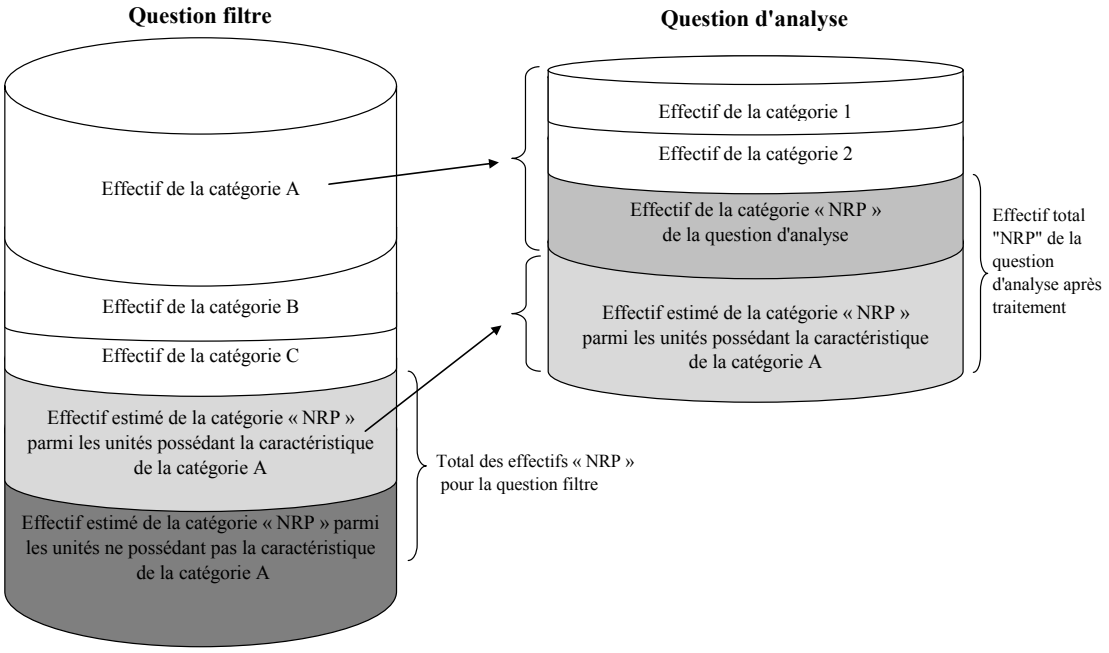


Note: La catégorie « NRP » représente l'ensemble de la non-réponse partielle, ce qui veut dire qu'elle inclut les « ne sait pas », « ne répond pas », « inconnu » et « refus global ».

43. Cette estimation est faite à partir de l'hypothèse d'une non-réponse partielle distribuée aléatoirement.

Schéma 2

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



Note: La catégorie « NRP » de la question filtre est scindée en deux parties, celle représentée en gris pâle étant l'effectif estimé des personnes possédant la caractéristique de la catégorie A, mais qui n'ont pas répondu à la question filtre.

Pour estimer adéquatement un effectif de population pour une sous-population définie à partir d'une question filtre, on doit d'abord décomposer la non-réponse partielle à la question filtre en deux parties, soit la non-réponse des unités qui appartiennent à la catégorie A et la non-réponse de celles qui n'y appartiennent pas. Le schéma 2 illustre cette décomposition. Dans la boîte de gauche (question filtre), la section en gris pâle représente une estimation des unités qui, bien que n'ayant pas fourni de réponse à la question filtre, possèdent la caractéristique de la catégorie A (sous-population étudiée). La section en gris foncé représente quant à elle une estimation des unités n'ayant pas fourni de réponse à la question filtre et qui n'avaient pas à répondre à la question d'analyse car elles n'ont pas la caractéristique de la catégorie A. Pour l'estimation des effectifs de la question d'analyse, on doit considérer les non-répondants partiels à cette question parmi les répondants de la catégorie A de la question filtre (section en gris moyen de la boîte de droite), en plus des non-répondants à la question filtre qui appartiennent à la catégorie A (section en gris pâle de la boîte de gauche et de droite). Ces deux types d'unités non répondantes doivent être répartis dans les différentes catégories de la question d'analyse, dans les mêmes proportions que les répondants.

L'annexe 3 présente les détails du calcul de l'estimation en cascade d'un effectif pour l'exemple exposé aux tableaux 2.5 et 2.6. Les données disponibles sur le portail de l'Infocentre de santé publique ont été recodées afin de permettre l'application de la méthode proposée, de sorte que les résultats offerts sur ce portail sont conformes (tableau 2.7). Il est à noter que les estimations d'effectifs présentées sur le portail sont toutes arrondies à la centaine près.

Tableau 2.7

Proportion des élèves du secondaire ayant eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, pour l'ensemble du Québec, EQSJS 2010-2011

Statistiques					
Nombre	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
10 400	13,2	4,9	0,65	(11,9 – 14,5)	0,7

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Étant donné qu'elle repose sur la même hypothèse que celle portant sur l'estimation de proportions, à savoir que les répondants sont représentatifs des non-répondants partiels, cette solution 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 répondants. Une telle analyse a été faite pour les indicateurs présentant les taux de non-réponse partielle les plus élevés; les résultats sont présentés aux sections 2.3.2 et 2.3.3. Le lecteur est également invité à consulter l'annexe 2, qui explique comment procéder au diagnostic et à l'analyse qualitative de la non-réponse partielle.

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 qu'elles puissent être inférées à l'ensemble de la population visée par l'enquête. Dans cette section, une autre forme de poids sera abordée, soit les poids « bootstrap ». Ceux-ci ne remplacent pas les variables de pondération utilisées pour l'estimation des proportions et des effectifs, mais servent plutôt à estimer la précision des quantités estimées. Il ressort que l'approche utilisant les poids « bootstrap » permet de construire des intervalles de confiance selon une méthode empirique (section 3.3.3), en vue d'effectuer des tests de comparaison deux à deux tout en s'affranchissant de l'hypothèse habituelle de normalité des observations (section 5.1.2).

3.1.1 Définition et méthodes d'estimation

Comme toutes les estimations d'enquête par échantillon, celles tirées de l'EQSJS 2010-2011 sont entachées d'erreurs puisque les données proviennent d'un échantillon plutôt que de l'ensemble de la population. Lorsque l'on infère les résultats obtenus à partir d'un échantillon à la population visée, 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 estimation de proportion issue de l'échantillon.

La complexité du plan d'échantillonnage de l'EQSJS 2010-2011 influence grandement la précision des estimations : la répartition régionale non proportionnelle, l'absence d'une liste exhaustive des classes de la population pour la planification de l'enquête ainsi que la taille variable des classes ont pour effet de réduire la précision, par rapport à celle que l'on obtiendrait à partir d'un échantillon aléatoire simple d'un même nombre d'élèves. L'impact n'est pas négligeable et doit être pris en compte dans l'estimation de la précision.

Certains logiciels d'analyse statistique, tels que SPSS et SAS, incluent des procédures[44] qui permettent d'obtenir des estimations sans biais de la précision pour certains plans de sondage complexes, comme celui utilisé dans l'EQSJS 2010-2011. La variance peut être estimée à l'aide d'une méthode de linéarisation, aussi appelée approximation par séries de Taylor.

D'autres logiciels spécialisés, comme SUDAAN et WESVAR, peuvent aussi être utilisés aux mêmes fins. Ces logiciels ont l'avantage de permettre l'utilisation de plusieurs méthodes d'estimation de la variance qui tiennent compte de la complexité du plan de sondage : méthode de linéarisation et méthode de ré-échantillonnage (ex. : méthode du « bootstrap »)[45].

Peu importe le logiciel utilisé, l'emploi de l'une ou l'autre des méthodes susmentionnées nécessite d'avoir accès à l'information sur les paramètres du plan de sondage. Pour utiliser la méthode de linéarisation, les indicateurs de strates et d'unités primaires d'échantillonnage doivent être disponibles avec les microdonnées. Pour utiliser la méthode du « bootstrap », il suffit d'avoir accès aux ensembles de poids « bootstrap », puisque ce sont ceux-là mêmes qui reflètent la façon dont l'échantillon a été sélectionné. Par ailleurs, contrairement à la méthode de linéarisation, la méthode du « bootstrap » permet de tenir compte de tous les ajustements apportés à la pondération. En ce sens, la première doit être considérée comme une méthode plus approximative pour l'estimation de la variance.

44. Dans la version 9 de SAS, ces procédures sont incluses 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 ré-échantillonnage dans les ouvrages de Lohr (1999), Rust et Rao (1996) et Wolter (1985).

Pour l'EQSJS 2010-2011, il a par conséquent été convenu d'estimer la précision des estimations produites à partir des données de l'enquête à l'aide de la méthode du « bootstrap ». Ainsi, l'ISQ a produit des ensembles de poids « bootstrap », qui accompagne 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 de santé publique. Par contre, les poids « bootstrap » ne sont pas inclus dans le fichier masqué (FMII), puisque la presque totalité des usagers ont recours au logiciel SPSS pour exploiter les données de ce fichier. Ce dernier contient plutôt des variables décrivant les strates et les unités primaires d'échantillonnage, de sorte que la précision des statistiques produites peut être estimée de façon approximative[46].

Voici, brièvement, comment fonctionne la méthode du « bootstrap » (plus de détails sont fournis à l'annexe 4). Pour obtenir une estimation de la variance associée à l'estimation d'une proportion (statistique calculée à partir du poids de sondage), il suffit de recalculer cette même estimation de la proportion un grand nombre de fois en utilisant chacun des ensembles de poids « bootstrap » fournis avec les données de l'enquête. Cette opération donne une distribution de valeurs possibles pour l'estimateur de la proportion. La variabilité observée entre les différentes proportions estimées permet d'estimer la variance ou autres mesures de précision. Dans l'EQSJS 2010-2011, 2 000 ensembles de poids « bootstrap » ont été produits. Ceux-ci peuvent être utilisés en totalité ou en partie, selon l'analyse qui est faite. La section 3.3.3 donne des explications à ce sujet.

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. Celle-ci a tendance à croître avec le taux de valeurs imputées et peut être considérable si ce taux est élevé.

Il a été mentionné à la section 2.3.3 que l'imputation a été utilisée pour minimiser l'impact de la non-réponse partielle dans trois situations précises : le refus ou l'incapacité pour les élèves de fournir l'information requise pour la variable de croisement du plus haut niveau de scolarité des parents, la non-réponse aux questions de la section portant sur la sexualité dans les cas où l'élève n'était pas à l'aise d'y répondre et la non-réponse à l'indicateur de la « Proportion des élèves du secondaire ayant consommé de la drogue au cours des 12 derniers mois », attribuable à certains items manquants dans la construction de l'indicateur. Dans les trois cas, la non-réponse est caractérisée et son impact sur la variance ne peut donc être considéré comme négligeable. Quoi qu'il en soit, puisqu'il est plus important de minimiser le biais d'une estimation ponctuelle que de minimiser le biais de l'estimation de sa précision, l'imputation demeure une bonne option.

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'enquêtes sociosanitaires est le coefficient de variation.

46. Les estimations de précision obtenues par le biais de la méthode de linéarisation et de la méthode « bootstrap » sont en général légèrement différentes.

Le coefficient de variation (C.V.) est une mesure de précision relative. Exprimé en pourcentage, il s'obtient 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 la prudence.

3.2.2 Recommandations concernant les seuils de diffusion

Certaines estimations ont une excellente précision, alors que d'autres sont moins précises. Le potentiel d'analyse des premières est élevé tandis que celui de la deuxième catégorie soulève plus de questions. 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 dont le C.V. se situe 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 statistiques disponibles par l'intermédiaire d'autres sources. L'ISQ recommande également, dans le cas de statistiques peu précises, d'émettre des indications claires sur les limites d'utilisation 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'usager. Cependant, dans un souci d'uniformisation des normes appliquées pour toutes les enquêtes sociosanitaires au Québec, le Comité tripartite MSSS-INSPQ-ISQ a décidé que les normes de l'ISQ seront utilisées pour la présentation des résultats des indicateurs de l'EQSJS 2010-2011 diffusés sur le portail de l'Infocentre de santé publique.

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. ≤ 15 %[1]	(aucun signe)	Estimation diffusée sans mention
15 % < C.V. ≤ 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. ≤ 5%); B : (5% < C.V. ≤ 10%) et C : (10% < C.V. ≤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 l'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 calculait son intervalle de confiance, 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. Certaines se distinguent par le choix de la loi servant à l'approximation de la loi de l'estimateur. Les méthodes standard utilisent habituellement la loi Normale ou la loi Binomiale. Ces méthodes seront discutées à la section 3.3.2.

Il existe également d'autres méthodes qui permettent de s'affranchir de l'hypothèse selon laquelle les données suivent une loi connue; la méthode basée sur le calcul des centiles 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 3.3.3). Par contre, ils constituent une option supplémentaire lorsque ces poids sont utilisés.

En somme, le choix de la méthode pour construire un intervalle de confiance est indépendant du choix de la méthode utilisée pour estimer la variance. Les prochaines sections présentent deux méthodes différentes pour construire un intervalle de confiance à partir des données de l'EQSJS 2010-2011. D'abord, il sera question des intervalles de confiance standard (section 3.3.2), et ensuite des intervalles de confiance « bootstrap » (section 3.3.3).

3.3.2 Méthode de calcul pour les intervalles de confiance standard

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 :

$$\text{marge d'erreur} = z_{(1-\alpha/2)} \times \text{erreur-type de l'estimation}$$

où $z_{(1-\alpha/2)}$ est le 100(1-a/2)^{e} centile de la distribution Normale centrée réduite. Par exemple, si le seuil de signification α est fixé à 0,05, alors $z_{(1-\alpha/2)}$ = 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 symétrique[47] 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.

Pour les 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 le nombre de répondants est petit et que la proportion étudiée est faible ou, au contraire, très élevée, 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 se trouvent à l'annexe 5. Dans le cas de l'intervalle de confiance Binomial, la marge d'erreur perd son sens en raison de l'asymétrie de celui-ci. 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, appelées ici « méthodes standard ».

47. Cela signifie 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.

3.3.3 Méthode de calcul pour les intervalles de confiance « bootstrap »

Pour l'EQSJS 2010-2011, une autre approche a été utilisée en remplacement des méthodes standard. Il s'agit de la méthode des intervalles de confiance « bootstrap ». Elle 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 se répercute principalement au niveau des tests statistiques. Plus de détails sont fournis à la section 5.1.2.

La méthode des intervalles de confiance « bootstrap » est valide pour les données de l'EQSJS 2010-2011. 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'EQSJS 2010-2011 est la méthode basée sur le calcul des centiles de la distribution empirique « bootstrap », à laquelle une correction a été appliquée pour la continuité (Pires et autres, 2008). De façon très succincte, on peut dire qu'un intervalle de confiance basé sur les centiles « bootstrap » s'obtient de la façon suivante : on se sert de chacun des 2 000 ensembles de poids « bootstrap », discutés à la section 3.1.1, pour calculer 2 000 estimations ponctuelles, comme les estimations d'une proportion. On place ensuite ces 2 000 valeurs en ordre croissant, puis on attribue comme bornes de l'intervalle de confiance, de niveau de confiance de 95 %, les 2,5e et 97,5e centiles de la distribution empirique ordonnée[48].

Comme il a été mentionné à la section 3.1.1, l'estimation de la précision des estimations issues de l'EQSJS 2010-2011 repose sur la méthode du « bootstrap ». Selon la littérature, un nombre de 500 répliques (ou ensembles de poids « bootstrap ») est jugé suffisant pour estimer adéquatement la précision des estimations obtenues à partir de données d'enquête. Toutefois, dans le cas de l'estimation de centiles (nécessaires à la construction des intervalles de confiance « bootstrap »), ce nombre peut s'avérer largement insuffisant. En effet, un nombre trop faible de répliques engendre une variabilité des bornes de l'intervalle qui n'est pas souhaitée. Par conséquent, une évaluation détaillée de l'utilisation des intervalles de confiance « bootstrap » a été réalisée par l'ISQ au début de l'année 2010, à l'occasion des travaux associés à l'*Enquête québécoise sur la santé de la population 2008* (EQSP). 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 standard, et 2) dans la négative, déterminer le nombre d'ensembles de poids « bootstrap » requis.

L'étude de l'ISQ conclut que 500 ensembles de poids « bootstrap », bien que suffisants pour estimer la variance des estimations, ne suffisent pas pour estimer adéquatement les centiles de la distribution empirique « bootstrap ». Elle recommande un minimum de 2 000 répliques pour répondre à cet objectif, afin de construire des intervalles de confiance « bootstrap » de niveau de confiance de 95 %.

On a vérifié que la méthode de calcul pour les intervalles de confiance « bootstrap » possède de bonnes propriétés pour les données de l'EQSJS 2010-2011, tout comme pour celles de l'EQSP. Il a finalement 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 produites et réaliser des tests globaux de comparaison (tests du khi-deux), mais que 2 000 ensembles de poids « bootstrap »[49] seraient utilisés pour la construction des intervalles de confiance « bootstrap », desquels sont issus les tests d'égalité de deux proportions (voir section 5.1). Ces 2 000 ensembles de poids « bootstrap » ont donc été créés pour l'EQSJS 2010-2011 et sont disponibles avec les données du fichier maître.

48. Une correction de continuité est appliquée à ces centiles.
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.

3.3.4 Recommandations relatives au choix du niveau de confiance ou au seuil de signification

En raison du choix de l'intervalle de confiance « bootstrap » pour les données de l'EQSJS 2010-2011 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]. Lorsque ceux-ci sont utilisés pour faire des tests statistiques, le seuil de signification correspondant au niveau de confiance de 95 % est de 5 %.

Par contre, l'usager qui souhaiterait produire des intervalles de confiance standard à partir du FMII, ou en ayant recours à l'erreur-type fournie dans les tableaux produits à l'Infocentre de santé publique, est libre d'utiliser le seuil de signification qui convient à ses analyses. Pour plus d'information concernant le niveau de confiance d'un intervalle, le lecteur est invité à consulter l'annexe 6.

50. C'est ce qui est produit sur le portail de l'Infocentre de santé publique.

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 termes de structure d'âge?

Bien que l'ajustement des proportions selon l'âge soit une pratique peu courante pour des données d'enquête, il est préférable de produire des proportions ajustées lorsque le phénomène qu'on veut étudier est fortement lié à l'âge et qu'on désire en éliminer l'effet. C'est le cas, notamment, de certains indicateurs de l'état de santé et des habitudes de vie, pour lesquels l'âge peut être un déterminant majeur. Dans le cas de l'EQSJS 2010-2011, les différences dans les structures d'âge inhérentes à chacune des régions et les variations dans les périodes de collecte au cours de l'année scolaire font en sorte que les comparaisons entre deux régions peuvent être altérées par un effet d'âge. Il serait normal, par exemple, d'observer plus souvent certains comportements liés à l'âge au sein d'une région dont la structure d'âge indique une plus grande proportion d'élèves plus âgés. L'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

Il est recommandé 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 7 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 autres, 2005). La population de référence est généralement celle de l'ensemble du Québec, pour les comparaisons régionales, ou du Canada pour les comparaisons provinciales (population au plus récent recensement, sexes réunis, corrigée pour le sous-dénombrement). Toute autre population de référence peut également être considérée en fonction du type et de la période de comparaison. Enfin, il est possible d'utiliser différentes structures d'âge de la population de référence. Il est cependant important de sélectionner des groupes d'âge pour lesquels il est possible d'obtenir des estimations de proportions avec une assez bonne précision.

Pour l'EQSJS 2010-2011, la structure par âge suivante est préconisée pour les indicateurs se rapportant à l'ensemble de la population visée par l'enquête : 12 ans et moins, 13 ans, 14 ans, 15 ans, 16 ans et 17 ans et plus. Cependant, pour certains indicateurs, d'autres structures d'âge doivent être considérées; elles sont présentées dans le tableau 4.1. La population de référence est celle visée par l'enquête, et la proportion servant à l'ajustement, pour un groupe d'âge donné, est obtenue en prenant la somme des poids de sondage des élèves appartenant à ce groupe d'âge divisée par la somme des poids de sondage de tous les élèves appartenant à l'ensemble des groupes d'âge retenus pour cet ajustement.

Tableau 4.1
Structure d'âge retenue pour la standardisation selon la population couverte par les indicateurs de l'EQSJS 2010-2011

Population couverte	Structure d'âge					
	12 ans et moins	13 ans	14 ans	15 ans	16 ans	17 ans et plus
Toute la population	X	X	X	X	X	X
12 ans et plus	12 ans seul.	X	X	X	X	X
13 ans et plus		X	X	X	X	X
14 ans et plus			X	X	X	X
15 ans et plus				X	X	X
16 ans et plus					X	X

4.3 Variance et intervalle de confiance pour les proportions ajustées

L'utilisation des poids « bootstrap » pour calculer la variance de proportions ajustées est la méthode la plus précise qui est recommandée, car elle tient pleinement compte du plan de sondage complexe et de la non-réponse. Il s'agit simplement de calculer pour chaque ensemble de poids « bootstrap » la proportion ajustée selon l'âge et de considérer la variance de ces estimations de proportions ajustées. Cette méthode est également conseillée pour l'obtention des intervalles de confiance, comme il est expliqué à la section 3.3.3.

Toutefois, si les poids « bootstrap » ne sont pas disponibles dans le fichier de données (par exemple, le 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 paramètres du plan de sondage dans les logiciels statistiques (voir section 3.1.1). L'approximation de la variance est obtenue en prenant la somme pondérée de ces variances précédemment calculées. Le poids retenu est celui de la proportion au carré que représente le groupe d'âge dans la population de référence (voir à ce sujet un exemple à l'annexe 7). On parle ici d'une approximation car cette façon de faire ne tient pas compte des covariances qui existent, en raison de la complexité du plan de sondage, entre les proportions estimées de deux différents groupes d'âge, contrairement à la méthode utilisant les poids « bootstrap ».

Pour le calcul des proportions ajustées sur le portail de l'Infocentre de santé publique, la population de référence qui a été retenue est celle visée par l'EQSJS 2010-2011. Pour retrouver les chiffres de population de chaque âge ou groupe d'âge, il suffit de prendre la somme des poids de sondage 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 de la méthode du « bootstrap ».

Cette section présente les tests statistiques qui doivent être utilisés pour analyser adéquatement les données de l'EQSJS 2010-2011. L'ensemble des renseignements et des recommandations présentés dans les précédentes sections du guide convergent vers une utilisation de qualité des données. Les tests statistiques, abordés ci-après, visent le même objectif. On y traite notamment de la théorie des tests statistiques ainsi que de leurs avantages, leurs limites et leurs conditions d'application.

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, comme le niveau scolaire ou le sexe. Un test global d'indépendance (section 5.1.1) permet de répondre à cette question. Le cas échéant, il importe de savoir quelles sont les catégories de la variable de croisement qui diffèrent entre elles. Pour ce faire, on compare ces catégories entre elles deux à deux (par exemple, deux niveaux scolaires) à l'aide d'un test d'égalité 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 8.

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, tenir compte du plan de sondage de l'enquête. Ils doivent notamment intégrer la pondération de l'enquête.

5.1.1 Test global d'indépendance

Lorsqu'on étudie la relation entre des variables à plus de deux catégories (indicateur ou variable de croisement)[51], la première chose à faire est la production d'un test global d'indépendance entre l'indicateur analysé et la variable de croisement. Le test du khi-deux peut être utilisé pour confirmer ou infirmer l'hypothèse selon laquelle les valeurs prises par l'indicateur ne dépendent pas des valeurs prises par la variable de croisement[52]. Un test global permet de contrôler le seuil de signification du test, et ce n'est qu'en présence d'un résultat significatif qu'il est justifié d'examiner plus attentivement les proportions présentant les écarts les plus importants. Cette démarche est discutée à la section 5.1.2.

Le test du khi-deux doit tenir compte correctement du plan de sondage complexe de l'enquête pour être adéquat, au même titre que les estimations de variance produites à partir des données de l'EQSJS 2010-2011, qui doivent intégrer la pondération et les poids « bootstrap » pour ce faire. Par ailleurs, il est important de s'assurer que les conditions de validité du test du khi-deux sont satisfaites. Les résultats produits sur le portail sont vérifiés afin de s'assurer que ces conditions sont remplies. Dans le cas contraire, une mise en garde s'affiche au-dessous du tableau, à l'endroit où est 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, l'annexe 9 peut être consultée.

5.1.2 Test d'égalité de deux proportions

Une fois la relation entre un indicateur à plusieurs catégories et une variable de croisement déclarée significative (test global d'indépendance significatif), il faut examiner plus en détail les proportions qui définissent cette relation. Pour ce faire, on utilise un test qui permet de comparer les proportions deux à deux. Il est à noter que dans le cas d'un indicateur à deux catégories seulement, analysé selon une variable de croisement qui ne possède également que deux catégories, telle que le sexe, le test décrit ici fournit un seuil observé approximativement égal à celui du test global du khi-deux présenté à la section précédente. Il est donc inutile d'effectuer au préalable un test global du khi-deux.

51. Dans le cas d'un indicateur à deux catégories et d'une variable de croisement à deux catégories, se référer à la section 5.1.2.
52. Par exemple, on peut faire un test de l'hypothèse selon laquelle le niveau d'activité physique n'est pas relié au niveau scolaire.

Pour les besoins du portail de l'Infocentre de santé publique, le test retenu par le Comité tripartite MSSS-INSPQ-ISQ pour évaluer si la différence entre deux proportions est significative est le test empirique basé sur l'intervalle de confiance de la différence de deux proportions (méthode 1). Cette méthode s'appuie sur l'utilisation des 2 000 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). Les détails de la construction de ce test sont présentés à l'annexe 10, mais il est important de mentionner 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 il a été 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 les 2,5e et 97,5e centiles de cette distribution. Le résultat du test s'obtient ainsi : on peut conclure que deux proportions sont différentes au seuil de signification de 5 % lorsque l'intervalle de confiance de la différence ne contient pas la valeur nulle. La puissance du test de la méthode 1 est du même ordre de grandeur que celle d'un test de Student (test *t*). Notons que cette méthode a été conçue et mise en place initialement pour l'analyse des données de l'EQSP 2008 à l'Infocentre de santé publique.

Toutes les comparaisons deux à deux produites à l'Infocentre de santé publique à partir des données de l'EQSJS 2010-2011 sont le résultat de tests basés sur la méthode 1. Cependant, dans certaines circonstances, les comparaisons deux à deux ne peuvent être obtenues directement de la requête de l'usager sur le portail (voir la section 6.3). Dans ces cas, l'utilisateur pourra employer une méthode alternative pour comparer les proportions. Une méthode suggérée est 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, au seuil de signification de 5 %, lorsque leurs intervalles de confiance à 95 % sont disjoints. Dans le cas contraire, toutefois, on ne peut conclure. Un écart significatif selon la méthode 2 implique nécessairement un écart significatif selon la méthode 1, mais non l'inverse. 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. Des précisions sur le test basé sur la méthode 2 sont disponibles à l'annexe 10[53].

Notons qu'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 révèle d'écart significatif. 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 quand on l'étudie du point de vue des comparaisons deux à deux, en particulier si la méthode 2 est utilisée pour effectuer ces comparaisons. Dans le cas de l'EQSJS 2010-2011, 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 d'égalité de deux proportions, non seulement la méthode utilisée permet de s'affranchir de l'hypothèse selon laquelle les observations suivent une loi connue, mais elle requiert aussi un nombre supérieur d'ensembles de poids « bootstrap » (2 000).

Par ailleurs, lors de l'interprétation des résultats d'un test statistique, il importe de prendre certaines précautions. La partie III présente un certain nombre de recommandations pour l'interprétation de résultats provenant de données d'enquête dans un contexte général.

53. Dans les situations non concluantes, l'usager peut choisir de calculer lui-même la statistique approximative d'un test d'égalité des proportions, en supposant l'indépendance entre les deux proportions. Cette statistique, appelée *statistique de Wald*, est le rapport de la valeur absolue de la différence entre les proportions sur la racine carrée de la somme des variances des proportions. Elle suit approximativement une distribution Normale. Contrairement au test empirique de la méthode 1, le test de Wald sera en général conservateur lorsque l'hypothèse d'indépendance n'est pas vérifiée, mais il le sera moins que le test basé sur la comparaison des intervalles de confiance. C'est pourquoi il pourrait permettre de détecter un écart en dépit d'intervalles de confiance qui se recoupent légèrement.

6. Exploitation des données de l'EQSJS 2010-2011

La présente section aborde des éléments à considérer pour faire une saine exploitation des données de l'EQSJS 2010-2011. 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 de santé publique pour les données de l'EQSJS. Finalement, la section 6.3 rassemble les principales particularités des indicateurs de l'EQSJS et des résultats diffusés par l'Infocentre de santé publique. On y présente notamment les choix faits pour certains indicateurs et les impacts sur leur utilisation et leur interprétation. Il est à noter que cette section n'a pas pour objet de couvrir l'ensemble des choix faits pour l'EQSJS, 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'EQSJS 2010-2011 avec d'autres sources de données

L'EQSJS 2010-2011 a été élaborée d'abord et avant tout dans un but de comparaison avec les futures éditions de l'enquête. La comparaison avec d'autres enquêtes devrait donc en général se limiter à se situer par rapport à elles en termes d'ordre de grandeur. Les différences méthodologiques commandent la prudence.

La prudence est généralement de mise lorsque les populations visées par deux enquêtes sont différentes (ex. : provinces différentes, groupes d'âge différents). Toutefois, pour des populations similaires étudiées à des périodes différentes mais selon une méthodologie variable, il pourrait être tentant d'interpréter les écarts observés en termes de changements dans la population. Or, il est particulièrement délicat de faire des comparaisons temporelles avec d'autres enquêtes, notamment pour de courtes périodes. En effet, les écarts entre deux périodes rapprochées sont faibles en général, de sorte qu'il devient hasardeux d'attribuer des écarts observés à des changements réels plutôt qu'à des différences dans la méthodologie, surtout s'ils sont petits. À l'opposé, si les écarts sont importants, on pourra croire à un impact méthodologique puisqu'il est peu probable que de grands écarts soient observés dans la population sur de courtes périodes. Plusieurs éléments méthodologiques peuvent avoir un impact sur la comparabilité. La base de sondage, la population visée, la méthode de collecte, la formulation des questions, le mode d'administration des questions et l'univers couvert par l'indicateur étudié comptent parmi les plus importants.

Par exemple, l'EQSJS 2010-2011 présente plusieurs similarités méthodologiques avec l'ETADJES. Dans les deux cas, il s'agit d'une sélection de classes dans les écoles secondaires et d'un questionnaire autoadministré à l'école. Toutefois, comparativement à l'ETADJES, le contenu de l'EQSJS 2010-2011 est plus général, sa période de collecte plus longue, et un questionnaire électronique a été utilisé plutôt qu'un questionnaire papier. De plus, pour l'EQSJS, un effort plus grand a été investi pour inclure le plus de classes possible dans la base de sondage des classes, en particulier celles qui ne sont pas composées d'élèves suivant un parcours de formation générale. Par conséquent, le cas échéant, il pourrait être intéressant de mentionner qu'un phénomène mesuré dans chacune des enquêtes présente une prévalence similaire, mais on devrait se garder d'interpréter un léger écart comme un changement observé dans la population. À l'opposé, un écart important, puisqu'il est peu probable, ne devrait pas être mentionné si l'on n'en a pas déterminé la raison, car il indiquerait vraisemblablement un problème au niveau de la comparabilité.

Lorsque l'univers couvert par un indicateur diffère entre deux enquêtes, il est parfois possible de définir un domaine d'étude commun afin de favoriser la comparaison. Par exemple, si l'on souhaite comparer les jeunes de l'*Enquête sociale et de santé auprès des enfants et des adolescents québécois 1999* (ESSEA; Aubin et autres, 2002) avec ceux de l'EQSJS 2010-2011 quant à leurs comportements sexuels, l'analyse pourrait se limiter au domaine commun aux deux enquêtes, soit les jeunes de 16 ans. Toutefois, contrairement à l'EQSJS, l'ESSEA couvre l'ensemble des jeunes de 16 ans, incluant ceux qui ne fréquentent pas l'école secondaire (environ 5 % des jeunes de 16 ans en 1999). L'interprétation des résultats doit tenir compte de cette différence, notamment.

De même, la comparaison des données de l'EQSJS 2010-2011 avec celles d'enquêtes portant sur la population en général, comme l'ESCC, devrait tenir compte de toutes les limites inhérentes à la méthodologie sous-jacente aux deux enquêtes. Il importe de mentionner que l'EQSJS vise les élèves inscrits à l'école secondaire au secteur des jeunes et exclut donc, en particulier, les élèves des centres de formation professionnelle. Par ailleurs, l'enquête ne couvre pas l'ensemble des jeunes de 12 à 17 ans. Les groupes d'âge extrêmes de l'échantillon des répondants, soit ceux des 12 ans et moins et des 17 ans et plus, sont les moins bien représentés car une part non négligeable des jeunes de cet âge ne sont pas inscrits au secondaire.

En somme, bien qu'aucune étude détaillée n'ait été effectuée sur la comparabilité des données de l'EQSJS 2010-2011 avec celles 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 de santé publique

Le portail de l'Infocentre de santé publique propose plusieurs options d'analyse pour les indicateurs du PCS dont la source est l'EQSJS 2010-2011. La première partie de cette section (6.2.1) présente les différents paramètres d'une requête type ainsi que les valeurs que peuvent prendre ces paramètres, tandis que la partie suivante (6.2.2) précise les contraintes imposées à ceux-ci pour l'analyse de certains indicateurs. La partie 6.2.3 expose les caractéristiques des niveaux de production possibles pour une requête : le niveau géographique, les variables de croisement autorisées, les tests effectués, etc. La dernière partie de la section (6.2.4) fournit quelques exemples de requêtes et de résultats pour chacun des niveaux géographiques permis (ensemble du Québec, régional et local). Des recommandations importantes à l'attention de 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. Il est d'ailleurs à souligner que ces recommandations ont toutes été retenues et appliquées sur le portail.

6.2.1 Paramètres d'une requête sur le portail de l'Infocentre de santé publique

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

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

Bien entendu, la première étape dans l'exécution d'une requête 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 retenu, l'usager a accès à la requête; les paramètres suivants doivent alors êtres définis.

♦ **Paramètre « Niveau géographique de l'école »**

Ce paramètre permet de définir l'unité territoriale d'analyse pour la production du tableau. Les niveaux géographiques possibles sont : « Ensemble du Québec », « Régional », « Local (RLS) » et « Local (Commission scolaire) ».

♦ **Paramètre « Région sociosanitaire de l'école »**

Ce paramètre permet de choisir la région sociosanitaire de l'école à analyser lorsque le niveau géographique « Régional » est sélectionné. Il y a seize régions potentielles.

♦ **Paramètre « Région sociosanitaire de l'école avec suréchantillon[54] »**

Ce paramètre permet de choisir la région sociosanitaire de l'école à analyser lorsque le niveau géographique « Local (RLS) » ou « Local (Commission scolaire) » est sélectionné. Cela est possible pour les régions avec suréchantillon. Pour une analyse par RLS, sept régions se sont prévalues de cette option pour l'EQSJS 2010-2011. Quant à l'analyse par commission scolaire, elle est possible pour deux régions. Seules ces régions peuvent être sélectionnées pour ce paramètre.

♦ **Paramètre « Sous-population »**

Ce paramètre permet de définir une seule ou plusieurs sous-populations d'analyse. Les choix offerts sont : « Sexe », « Niveau de scolarité de l'élève », « Sexe et niveau de scolarité de l'élève » 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 garçons et pour les filles, en plus du résultat pour les sexes combinés. À l'échelle locale, les niveaux de scolarité de l'élève sont regroupés par cycle : 1er cycle (1re et 2e secondaire) et 2e cycle (3e, 4e et 5e secondaire), et le choix « Sexe et niveau de scolarité de l'élève » n'est pas disponible. Lorsqu'aucune sous-population spécifique n'est visée, 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 « Groupe d'âge » est sélectionné, la répartition de l'indicateur sera produite selon la catégorisation des groupes d'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 variables de sous-population, et celles listées dans la fiche de l'indicateur.

♦ **Paramètre « Comparaison régionale »**

Ce paramètre permet à l'utilisateur de définir quel type d'analyse régionale il souhaite faire. Les choix possibles sont : « Aucune », « Région sociosanitaire de l'école comparée avec le reste du Québec » et « Région sociosanitaire de l'école comparée avec une autre région sociosanitaire ». Lors de la sélection de ce dernier type d'analyse, le paramètre « Région de comparaison » apparaît. Quinze régions sociosanitaires sont alors proposées. Ce paramètre est affiché seulement lorsque le niveau géographique spécifié est « Régional ».

♦ **Paramètre « Comparaison locale »**

Ce paramètre permet à l'utilisateur de définir quel type d'analyse au niveau des RLS ou des commissions scolaires il souhaite effectuer et il est affiché seulement lorsque le niveau géographique spécifié est « Local (RLS) » ou « Local (Commission scolaire) ». Les choix possibles sont alors : « Aucune » et « RLS comparés avec le reste de la région sociosanitaire de l'école » ou « Commissions scolaires comparées avec le reste des commissions scolaires de la région sociosanitaire de l'école ».

54. Les huit régions avec suréchantillon ont une taille d'échantillon accrue de manière à obtenir des estimations précises à l'échelle locale (par RLS ou CS, selon le cas).

♦ **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 de l'école 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 produit[55]. Ensuite, si le résultat est significatif, toutes les comparaisons deux à deux sont réalisées[56]. 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 de santé publique remarquera rapidement que les tests statistiques effectués sur les données de l'EQSJS 2010-2011 accessibles sur le portail prennent toujours la même forme. 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 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 étudiée[57]. Dans le cas où le test global montre une relation significative (valeur *p* du test inférieure au seuil de signification a fixé à 5 %[58]), des tests de comparaison de proportions deux à deux sont effectués[56] et un tableau présentant les résultats est affiché à la suite du tableau des estimations.

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 d'égalité de deux proportions est utilisé comme test global (section 5.1.2). Le test global du khi-deux (section 5.1.1) pourrait aussi s'appliquer et donnerait en général un résultat similaire. C'est pour cette raison que seul le test d'égalité de deux proportions, qui correspond à celui de la méthode 1 décrit à la section 5.1.2, est effectué (comme illustré au diagramme 6.1). Le tableau présentant le résultat de ce test est donc affiché directement à la suite de celui qui rapporte les estimations.

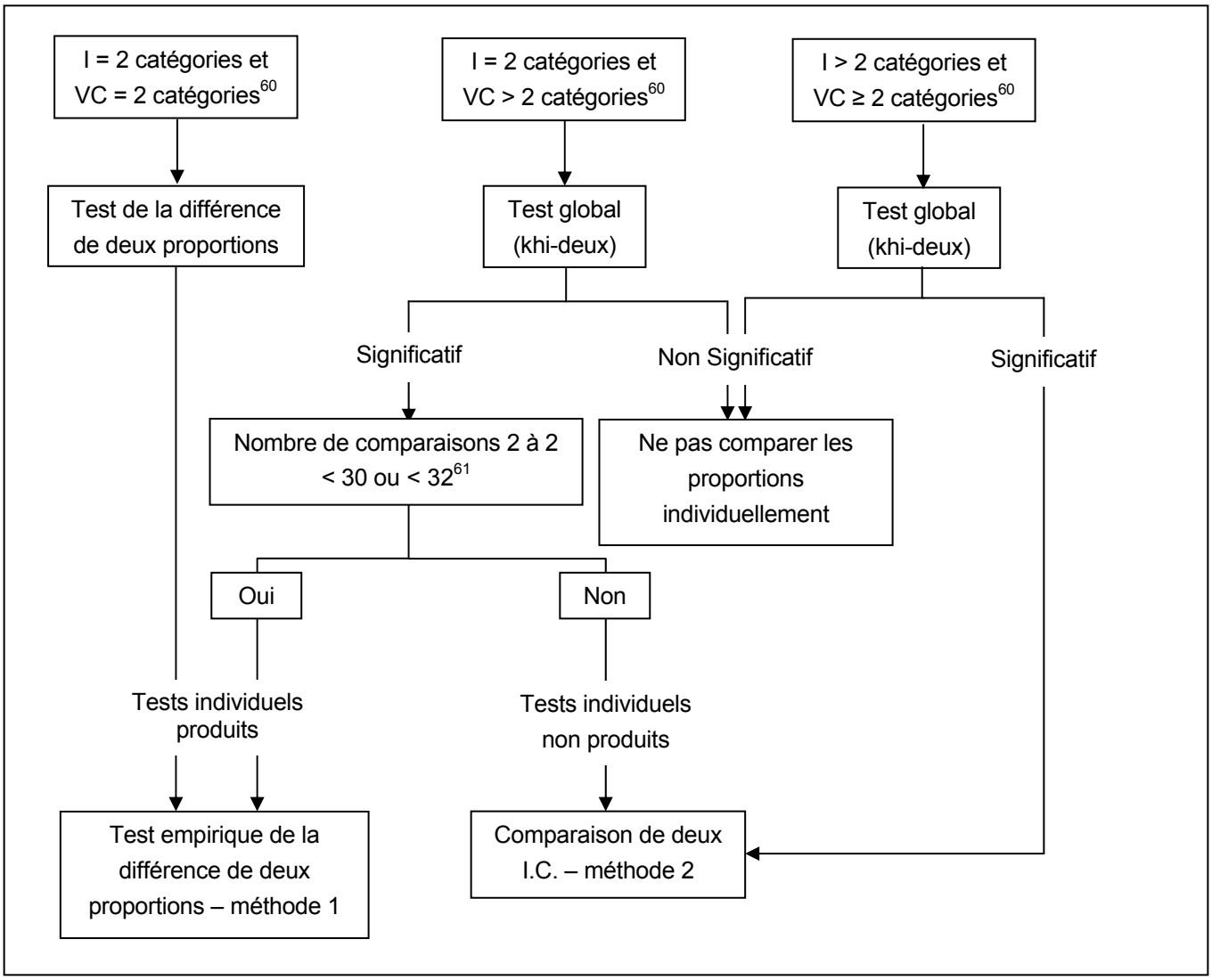
Dans certaines circonstances, qui sont décrites aux sections 6.2.2, 6.3.1 et 6.3.2, seuls les tests globaux du khi-deux seront produits par la requête. En présence d'un résultat significatif, l'utilisateur est invité à utiliser les intervalles de confiance de chacune des proportions qui sont présentés dans les tableaux afin de déterminer la source de la différence observée (méthode 2 – section 5.1.2). L'utilisateur doit donc lui-même comparer les intervalles de confiance fournis dans le tableau pour savoir si l'écart observé est significatif, ou encore réduire la taille de la requête[59].

---

55. Dans le cas d'un indicateur à deux catégories et d'une variable de croisement aussi à deux catégories, seules les comparaisons deux à deux sont produites. Voir la section 5.1.2 pour plus de détails.
56. Il est possible que malgré un test global significatif, les comparaisons deux à deux ne soient pas effectuées. Pour plus de détails, consulter la section 6.3.
57. Il y aura autant de tests globaux du khi-deux qu'il y a de sous-populations formées par la requête.
58. Pour l'EQSJS 2010-2011, une seule valeur pour le seuil de signification est permise, soit 5 %. Pour plus de détails, la section 3.3.4 peut être consultée.
59. Dans les situations non concluantes, l'usager peut choisir de calculer lui-même la statistique approximative d'un test d'égalité des proportions, en supposant l'indépendance entre les deux proportions. Cette statistique, appelée *statistique de Wald*, est le rapport de la valeur absolue de la différence entre les proportions sur la racine carrée de la somme des variances des proportions. Elle suit approximativement une distribution Normale. Contrairement au test empirique de la méthode 1, le test de Wald sera en général conservateur lorsque l'hypothèse d'indépendance n'est pas vérifiée, mais il le sera moins que le test basé sur la comparaison des intervalles de confiance. C'est pourquoi il pourrait permettre de détecter un écart en dépit d'intervalles de confiance qui se recoupent légèrement.

Diagramme 6.1

Tests statistiques recommandés et produits à l'Infocentre de santé publique pour la comparaison des proportions d'un indicateur de l'EQSJS 2010-2011 selon une variable de croisement



60. I = Indicateur et VC = Variable de croisement
61. Les tests individuels sont produits seulement si l'indicateur (I) a 2 catégories et que le nombre total de comparaisons 2 à 2 est inférieur à 30, ou encore à 32 lorsque la région sociosanitaire de l'école est sélectionnée comme variable de croisement. Le nombre total de comparaisons deux à deux correspond à la sommation des comparaisons deux à deux (entre la variable de croisement et l'indicateur) des modalités de la sous-population pour lesquelles le test du khi-deux est significatif.

♦ **Paramètre « Type de proportion »**

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

Toutefois ce paramètre n'est pas offert lorsque le groupe d'âge est choisi comme variable de croisement ou que le niveau de scolarité de l'élève est choisi comme sous-population ou variable de croisement. De plus, pour certains indicateurs qui incluent l'âge dans leur construction, la standardisation selon l'âge est possible ou non, selon le groupe d'âge visé. Ainsi, elle ne l'est pas pour les indicateurs ne concernant que les 17 ans et plus, mais elle l'est pour les 12 ans et plus, les 13 ans et plus, les 14 ans et plus, les 15 ans et plus et les 16 ans et plus, et le système de poids adéquat est alors utilisé.

♦ **Paramètre « Diffusion »**

Pour les utilisateurs ayant un niveau d'accès 20 ou 30, ce paramètre permet d'obtenir des tableaux avec masquage, tels que vus par les utilisateurs ayant un niveau d'accès 40 ou 50, ou des tableaux sans masquage. Dans le premier cas (« Diffusion » = Avec masquage), la diffusion des tableaux ne cause pas de problème de confidentialité, impliquant donc une diffusion sans autre restriction. Dans l'autre cas (« Diffusion » = Sans masquage), les tableaux complets ne pourront être diffusés tels quels si des avertissements liés à la confidentialité les accompagnent.

6.2.2 Particularités des paramètres pour certains indicateurs

♦ **Particularités des paramètres pour certains indicateurs dont le titre mentionne l'âge**

Pour les indicateurs qui incluent l'âge dans leur construction, le croisement de l'âge et du niveau de scolarité de l'élève n'est pas disponible. De plus, le niveau de scolarité de l'élève, comme variable de croisement ou comme sous-population, est disponible seulement pour les indicateurs concernant les 12 ans et plus, les 13 ans et plus et les 14 ans et plus, avec les catégories suivantes :

- Pour les 12 ans et plus et les 13 ans et plus :
  - 1re secondaire, 2e secondaire, 3e secondaire, 4e secondaire, 5e secondaire (analyse nationale et analyse régionale)
  - 1re et 2e secondaire, 3e secondaire, 4e et 5e secondaire (analyse nationale et analyse régionale)
  - 1er cycle (1re et 2e secondaire), 2e cycle (3e, 4e et 5e secondaire);

- Pour les 14 ans et plus :
  - 1re et 2e secondaire, 3e secondaire, 4e secondaire, 5e secondaire (analyse nationale et analyse régionale)
  - 1re et 2e secondaire, 3e secondaire, 4e et 5e secondaire (analyse nationale et analyse régionale)
  - 1er cycle (1re et 2e secondaire), 2e cycle (3e, 4e et 5e secondaire).

♦ **Indicateurs dont le titre mentionne l'âge et qui sont utilisés comme variables de croisement**

Les indicateurs sur les relations sexuelles (pour les 14 ans et plus) de même que les indicateurs d'âge d'initiation peuvent parfois être utilisés comme variables de croisement. Lorsque l'indicateur d'intérêt porte lui-même sur un groupe d'âge différent de celui de l'indicateur utilisé comme variable de croisement, la population est réduite :

- à l'âge couvert par l'indicateur d'intérêt lorsque celui-ci couvre un groupe d'âge plus restreint que celui de la variable de croisement. Ainsi, les résultats pour l'indicateur « Proportion des élèves du secondaire de 15 ans et plus ayant eu un épisode de consommation régulière de drogues pour la première fois avant l'âge de 15 ans » croisé avec la variable « Relation sexuelle consensuelle au cours de la vie (14 ans et plus) » sont restreints à la population de 15 ans et plus ;
- à l'âge couvert par la variable de croisement lorsque l'indicateur d'intérêt couvre toute la population visée. Par exemple, les résultats pour l'indicateur « Proportion des élèves du secondaire ayant eu un épisode de consommation régulière de drogues au cours de leur vie » croisé avec la variable « Relation sexuelle consensuelle au cours de la vie (14 ans et plus) » sont restreints à la population de 14 ans et plus.

Cependant, dans les situations suivantes, ce type de variables de croisement n'est pas disponible sur le portail de l'Infocentre de santé publique :

- lorsque le groupe d'âge ou le niveau de scolarité de l'élève est choisi comme variable de croisement ou comme sous-population ;
- lorsque l'indicateur d'intérêt ne couvre pas toute la population visée par l'enquête mais un groupe d'âge plus large que celui de la variable de croisement. Ainsi, le croisement de l'indicateur « Proportion des élèves du secondaire de 12 ans et plus ayant eu un épisode de consommation régulière de drogues pour la première fois avant l'âge de 12 ans » avec la variable de croisement « Relation sexuelle consensuelle au cours de la vie (14 ans et plus) » n'est pas disponible sur le portail.

♦ **Indicateur ayant une fréquence inférieure à 1 % au niveau de l'ensemble du Québec**

L'indicateur « Proportion des élèves du secondaire ayant fait usage de drogues par injection au cours de leur vie » a une fréquence inférieure à 1 % au niveau de l'ensemble du Québec. C'est pourquoi le niveau géographique de l'école ne permet pas la sélection des niveaux régional, local (RLS) et local (Commission scolaire). De plus, une seule variable de sous-population ou une seule variable de croisement (dans le cas où la sous-population totale est sélectionnée) peut être choisie.

6.2.3 Algorithme d'une requête paramétrable sur le portail pour les données de l'EQSJS 2010-2011

Cette section présente l'algorithme d'une requête faite sur le portail de l'Infocentre de santé publique; elle permet donc de connaître en détail les analyses permises lors d'une telle requête. Le lecteur est invité à consulter la section 6.2.4 pour avoir des exemples précis de requêtes et de sorties de tableaux ou l'annexe 11 pour visualiser des tableaux récapitulatifs de cet algorithme.

♦ **Estimations portant sur les élèves du secondaire**

1. Pour le paramètre « Niveau géographique de l'école » = « Ensemble du Québec » :

   i. Les choix du paramètre « Sous-population » sont : « Sexe », « Niveau de scolarité de l'élève », « Sexe et niveau de scolarité de l'élève », « Totale ».

      - Si « Sexe » est choisi, un paramètre « Sexe » permet de sélectionner : masculin, féminin, par sexe.
      - Si « Niveau de scolarité de l'élève » est choisi, un paramètre « Niveau de scolarité de l'élève » permet de sélectionner une des catégorisations du niveau de scolarité de l'élève mentionnées dans la fiche de l'indicateur.
      - Si « Sexe et niveau de scolarité de l'élève » est choisi, un paramètre « Sexe » permet de choisir entre masculin, féminin et par sexe, et un paramètre « Niveau de scolarité de l'élève » permet de choisir une des catégorisations du niveau de scolarité de l'élève mentionnées dans la fiche de l'indicateur.

   ii. Le paramètre « Première variable de croisement » permet de choisir :

      - « Aucune » :
        - Il n'y a pas d'autre paramètre à choisir.
        - Il n'y a aucun test statistique offert.

      - La variable « Région sociosanitaire de l'école » :
        - Si cette variable 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 », 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 régions sociosanitaires avec le reste du Québec et ainsi de comparer l'indicateur sous un angle d'analyse national, selon les régions sociosanitaires.
        - Il n'y a pas d'autres variables de croisement à choisir.

      - La variable de sous-population non sélectionnée entre le sexe et le niveau de scolarité de l'élève (les deux si le paramètre sous-population = « Totale » et aucune si la valeur du paramètre est « Sexe et niveau de scolarité de l'élève »), le groupe d'âge (si le paramètre sous-population n'est pas « Niveau de scolarité de l'élève » ni « Sexe et niveau de scolarité de l'élève ») 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 niveau de scolarité de l'élève ») et permet de choisir :
          - « Aucune » :
            - Affiche 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 », le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement ».
          - La variable « Région sociosanitaire de l'école » :
            - Aucun test statistique.
          - Les variables de croisement mentionnées dans la fiche de l'indicateur, à l'exception du sexe et du niveau de scolarité de l'élève :
            - Aucun test statistique.

iii. Si le « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population » ou le « Test de comparaison de l'indicateur selon la variable de croisement » est sélectionné :

1. Test avec l'intervalle de confiance de la différence, si l'indicateur et la variable de croisement ont deux modalités.
2. Test du khi-deux, autrement. Si le résultat du test du khi-deux est significatif :
   a. Si l'indicateur a deux modalités, l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour déterminer la source de la différence, jusqu'à un maximum de 30 comparaisons 2 à 2 (32 si la variable de croisement est la région de l'école), sinon une note est affichée.
   b. Si l'indicateur a plus de deux modalités, donc s'il est une répartition, il n'y a pas de test effectué pour déceler la source de la différence. Il est alors suggéré de comparer les intervalles de confiance.
3. S'il y a deux variables de croisement : aucun test statistique.

iv. Le choix d'obtenir une proportion ajustée selon l'âge n'est offert que si un test statistique est demandé, que l'âge n'est pas sélectionné comme variable de croisement et que le niveau de scolarité de l'élève n'est pas sélectionné comme variable de croisement ni comme sous-population.

v. Pour les utilisateurs ayant un niveau d'accès 20 ou 30, le paramètre « Diffusion » permet d'obtenir des tableaux dont la diffusion est permise (« Diffusion » = Avec masquage) ou non (« Diffusion » = Sans masquage).

2. Pour le paramètre « Niveau géographique de l'école » = « Régional » :

i. Choisir la région à l'aide du paramètre « Région sociosanitaire de l'école ».

ii. Les choix du paramètre « Sous-population » sont : « Sexe », « Niveau de scolarité de l'élève », « Sexe et niveau de scolarité de l'élève », « Totale ».

- Si « Sexe » est choisi, un paramètre « Sexe » permet de sélectionner : masculin, féminin, par sexe.
- Si « Niveau de scolarité de l'élève » est retenu, un paramètre « Niveau de scolarité de l'élève » permet de choisir une des catégorisations du niveau de scolarité de l'élève mentionnées dans la fiche de l'indicateur.
- Si « Sexe et niveau de scolarité de l'élève » est choisi, un paramètre « Sexe » permet de choisir entre masculin, féminin et par sexe, et un paramètre « Niveau de scolarité de l'élève » permet de sélectionner une des catégorisations du niveau de scolarité de l'élève mentionnées dans la fiche de l'indicateur.

iii. Les choix du paramètre « Comparaison régionale » sont : « Aucune », « Région sociosanitaire de l'école comparée avec le reste du Québec », « Région sociosanitaire de l'école comparée avec une autre région sociosanitaire ».

- Si « Aucune » et que la sous-population est « Sexe et niveau de scolarité de l'élève », on ne peut choisir de variable de croisement.
- Si « Aucune » et que la sous-population est « Totale » ou « Sexe » ou « Niveau de scolarité de l'élève » :
  - 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 le niveau de scolarité de l'élève (les deux si le paramètre sous-population = « Totale »), le groupe d'âge (si le paramètre sous-population n'est pas « Niveau de scolarité de l'élève » ou « Sexe et niveau de scolarité de l'élève ») et les autres variables de croisement mentionnées dans la fiche de l'indicateur :
  - Affiche 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 », le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement ».

- Si « Région sociosanitaire de l'école comparée avec le reste du Québec » :
  - On ne peut choisir d'autres variables de croisement, la région en deux 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 sociosanitaire de l'école comparée avec une autre région sociosanitaire » :
  - On ne peut choisir d'autres variables de croisement, la région en deux catégories (région sélectionnée et 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 deux 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 sociosanitaire de l'école comparée avec le reste du Québec » ou « Région sociosanitaire de l'école comparée avec une autre région sociosanitaire » est sélectionné :

- Test avec l'intervalle de confiance de la différence si l'indicateur et la variable de croisement ont deux modalités.
- Test du khi-deux autrement. Si le résultat du test du khi-deux est significatif :
  - Si l'indicateur a deux modalités, l'ensemble des tests deux à deux (intervalle de confiance sur la différence) sont faits pour déterminer la source de la différence, jusqu'à un maximum de 30 comparaisons deux à deux, sinon une note est affichée.
  - Si l'indicateur a plus de deux modalités, donc s'il est une répartition, il n'y a pas de test effectué pour déceler la source de la différence. Il est alors suggéré de comparer les intervalles de confiance.

v. Le choix d'obtenir une proportion ajustée selon l'âge n'est offert que si un test statistique est demandé, que l'âge n'est pas sélectionné comme variable de croisement et que le niveau de scolarité de l'élève n'est pas sélectionné comme variable de croisement ni comme sous-population.

vi. Pour les utilisateurs ayant un niveau d'accès 20 ou 30, le paramètre « Diffusion » permet d'obtenir des tableaux dont la diffusion est permise (« Diffusion » = Avec masquage) ou non (« Diffusion » = Sans masquage).

3. Pour le paramètre « Niveau géographique de l'école » = « Local (RLS) » ou « Local (Commission scolaire) » :

   i. Choisir la région pour laquelle on veut les RLS ou les commissions scolaires à l'aide du paramètre « Région sociosanitaire de l'école avec suréchantillon ».

   ii. Les choix du paramètre « Sous-population » sont : « Sexe », « Niveau de scolarité de l'élève », « Totale ».
      - Si « Sexe » est choisi, un paramètre « Sexe » permet de sélectionner : masculin, féminin, par sexe.
      - Si « Niveau de scolarité de l'élève » est choisi, un paramètre « Niveau de scolarité de l'élève » est fixé à « 1er cycle (1re et 2e secondaire), 2e cycle (3e, 4e et 5e secondaire) ».

   iii. Les choix du paramètre « Comparaison locale » sont : « Aucune », « RLS comparés avec le reste de la région sociosanitaire de l'école » ou « Commissions scolaires comparées avec le reste des commissions scolaires de la région sociosanitaire de l'école ».

      - Si « Aucune » et que la sous-population est « Totale », « Sexe » ou « Niveau de scolarité de l'élève » :
        - 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 le niveau de scolarité de l'élève (les deux si le paramètre sous-population = « Totale »), le groupe d'âge (si le paramètre sous-population n'est pas « Niveau de scolarité de l'élève ») et les autres variables de croisement mentionnées dans la fiche de l'indicateur :
            - Affiche 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 et le RLS (ou la commission scolaire) ». Si le paramètre sous-population = « Totale », le dernier choix du paramètre « Test statistique » est plutôt « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon le RLS (ou la commission scolaire) ».
      - Si « RLS comparés avec le reste de la région sociosanitaire de l'école » ou « Commissions scolaires comparées avec le reste des commissions scolaires de la région sociosanitaire de l'école » :
        - On ne peut choisir d'autres variables de croisement, le RLS ou la commission scolaire étant considéré comme une variable de croisement.
        - Le test statistique correspondant est effectué.

   iv. Si « RLS comparés avec le reste de la région sociosanitaire de l'école » ou « Commissions scolaires comparées avec le reste des commissions scolaires de la région sociosanitaire de l'école » 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 deux modalités.
      - Test du khi-deux, autrement. Si le résultat du test du khi-deux est significatif :
        - Si l'indicateur a deux modalités, l'ensemble des tests 2 à 2 (intervalle de confiance sur la différence) sont faits pour déterminer 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, il n'y a pas de test effectué pour déceler la source de la différence. Il est alors suggéré de comparer les intervalles de confiance.
      - Il n'est pas possible de comparer deux RLS ou commissions scolaires entre eux.

v. Le choix d'obtenir une proportion ajustée selon l'âge n'est offert que si un test statistique est demandé, que l'âge n'est pas sélectionné comme variable de croisement et que le niveau de scolarité de l'élève n'est pas sélectionné comme variable de croisement ni comme sous-population.
vi. Pour les utilisateurs ayant un niveau d'accès 20 ou 30, le paramètre « Diffusion » permet d'obtenir des tableaux dont la diffusion est permise (« Diffusion » = Avec masquage) ou non (« Diffusion » = Sans masquage).

6.2.4 Exemples de résultats de requêtes sur le portail de l'Infocentre de santé publique et recommandations

Pour les indicateurs du PCS dont la source est l'EQSJS 2010-2011, l'Infocentre de santé publique offre trois niveaux géographiques d'analyse : une production pour l'ensemble du Québec, une pour l'échelle régionale et une pour l'échelle locale. C'est le paramètre « Niveau géographique de l'école » 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 il a été 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 montre les différentes options pour obtenir des estimations de proportions portant sur l'ensemble des élèves du secondaire du Québec visés par l'enquête. Le second groupe concerne les analyses pour une région sociosanitaire de l'école, ce qui permet de se comparer soit au reste du Québec, soit à une autre région. Le troisième vise des analyses encore plus fines, au niveau des RLS ou des commissions scolaires, ce qui permet de comparer les résultats obtenus à celui du reste de la région sociosanitaire de l'école ou à celui du reste des commissions scolaires de la région sociosanitaire de l'école.

♦ **Premier groupe d'exemples : ensemble du Québec**

Ce premier exemple analyse l'évaluation de l'apparence des élèves québécois du secondaire selon leur niveau de scolarité et leur sexe. Deux options s'offrent : la première est de produire les proportions selon le sexe comme variable de sous-population, et le niveau de scolarité de l'élève comme variable de croisement (tableau 6.1); la seconde est de choisir la combinaison sexe-niveau de scolarité de l'élève comme variable de sous-population (tableau 6.2). La différence entre les deux options réside dans les analyses et les tests disponibles. Il incombe donc à l'utilisateur de produire le tableau qui correspond à ses besoins. Par exemple, au tableau 6.1, il est possible d'obtenir un test statistique de comparaison de l'indicateur selon la variable de croisement (niveau de scolarité de l'élève), stratifié selon la sous-population (sexe), tandis qu'au tableau 6.2, puisqu'il n'y a aucune variable de croisement sélectionnée, il n'est pas possible d'effectuer un tel test.

Pour des tableaux portant sur l'ensemble du Québec, il est aussi possible de choisir la valeur « Région sociosanitaire de l'école » comme première variable de croisement. Ce choix permet de faire une analyse des régions sous un angle national. De cette façon, 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 régions sociosanitaires. Le tableau 6.3 reproduit le visuel de la sortie de l'Infocentre de santé publique pour une telle requête sans les tests.

Seulement pour le niveau géographique national[62], il est possible dans certaines situations (consulter l'algorithme à la section 6.2.3) 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. Le tableau 6.4 donne un exemple de sortie de l'Infocentre de santé publique.

Tableau 6.1

Répartition des élèves du secondaire selon l'évaluation de son apparence, pour l'ensemble du Québec, EQSJS 2010-2011 (extrait du tableau complet)

Sous-population	Variable de croisement	Indicateur	Statistiques					
Sexe	Niveau de scolarité de l'élève	Évaluation de son apparence	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	1re secondaire	Poids insuffisant	6 600	14,8	6,7	1,00	(12,8 - 16,7)	0,5
		Poids normal	36 900	82,3	1,3	1,06	(80,2 - 84,4)	0,5
		Excès de poids	1 300	* 2,9	15,1	0,43	(2,0 - 3,8)	0,5
	2e secondaire	Poids insuffisant	5 200	11,1	7,2	0,80	(9,4 - 12,6)	0,2
		Poids normal	39 700	85,2	1,1	0,92	(83,4 - 87,2)	0,2
		Excès de poids	1 700	3,7	14,5	0,54	(2,6 - 4,8)	0,2
	3e secondaire	Poids insuffisant	5 000	10,0	8,3	0,83	(8,5 - 11,6)	0,3
		Poids normal	43 400	86,7	1,0	0,86	(85,0 - 88,3)	0,3
		Excès de poids	1 700	* 3,3	16,0	0,53	(2,4 - 4,4)	0,3
	4e secondaire	Poids insuffisant	3 900	9,5	8,8	0,83	(7,9 - 11,1)	0,2
		Poids normal	36 000	87,7	1,0	0,91	(85,9 - 89,5)	0,2
		Excès de poids	1 100	2,7	13,7	0,37	(2,0 - 3,6)	0,2
	5e secondaire	Poids insuffisant	3 100	8,7	9,6	0,83	(7,2 - 10,4)	0,1
		Poids normal	31 500	88,0	1,0	0,92	(86,2 - 89,7)	0,1
		Excès de poids	1 200	3,3	13,7	0,45	(2,4 - 4,3)	0,1
Féminin	1re secondaire	Poids insuffisant	10 400	25,0	4,5	1,13	(22,9 - 27,3)	0,4
		Poids normal	30 700	74,2	1,5	1,09	(72,0 - 76,3)	0,4

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

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 6.2

Répartition des élèves du secondaire selon l'évaluation de son apparence, pour l'ensemble du Québec, EQSJS 2010-2011 (extrait du tableau complet)

Sous-population		Indicateur	Statistiques					
Sexe	Niveau de scolarité de l'élève	Évaluation de son apparence	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	1re secondaire	Poids insuffisant	6 600	14,8	6,7	1,00	(12,8 - 16,7)	0,5
		Poids normal	36 900	82,3	1,3	1,06	(80,2 - 84,4)	0,5
		Excès de poids	1 300	* 2,9	15,1	0,43	(2,0 - 3,8)	0,5
	2e secondaire	Poids insuffisant	5 200	11,1	7,2	0,80	(9,4 - 12,6)	0,2
		Poids normal	39 700	85,2	1,1	0,92	(83,4 - 87,2)	0,2
		Excès de poids	1 700	3,7	14,5	0,54	(2,6 - 4,8)	0,2
	3e secondaire	Poids insuffisant	5 000	10,0	8,3	0,83	(8,5 - 11,6)	0,3
		Poids normal	43 400	86,7	1,0	0,86	(85,0 - 88,3)	0,3
		Excès de poids	1 700	* 3,3	16,0	0,53	(2,4 - 4,4)	0,3
	4e secondaire	Poids insuffisant	3 900	9,5	8,8	0,83	(7,9 - 11,1)	0,2
		Poids normal	36 000	87,7	1,0	0,91	(85,9 - 89,5)	0,2
		Excès de poids	1 100	2,7	13,7	0,37	(2,0 - 3,6)	0,2
	5e secondaire	Poids insuffisant	3 100	8,7	9,6	0,83	(7,2 - 10,4)	0,1
		Poids normal	31 500	88,0	1,0	0,92	(86,2 - 89,7)	0,1
		Excès de poids	1 200	3,3	13,7	0,45	(2,4 - 4,3)	0,1
Féminin	1re secondaire	Poids insuffisant	10 400	25,0	4,5	1,13	(22,9 - 27,3)	0,4
		Poids normal	30 700	74,2	1,5	1,09	(72,0 - 76,3)	0,4

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 6.3

Répartition des élèves du secondaire selon l'évaluation de son apparence, pour l'ensemble du Québec, EQSJS 2010-2011 (extrait du tableau complet)

Variable de croisement	Indicateur	Statistiques					
Région de l'école	Évaluation de son apparence	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
01 Bas-Saint-Laurent	Poids insuffisant	1 700	16,8	6,1	1,01	(14,8 - 18,7)	0,0
	Poids normal	8 400	80,9	1,4	1,11	(78,7 - 83,1)	0,0
	Excès de poids	200	* 2,3	18,8	0,44	(1,5 - 3,2)	0,0
02 Saguenay—Lac-Saint-Jean	Poids insuffisant	2 400	15,9	5,0	0,79	(14,4 - 17,5)	0,4
	Poids normal	12 200	81,3	1,1	0,91	(79,4 - 83,0)	0,4
	Excès de poids	400	* 2,8	15,5	0,43	(1,9 - 3,7)	0,4
03 Capitale-Nationale	Poids insuffisant	6 000	17,9	4,9	0,88	(16,2 - 19,6)	0,2
	Poids normal	27 200	80,8	1,0	0,84	(79,1 - 82,4)	0,2
	Excès de poids	500	* 1,3	17,9	0,24	(0,9 - 1,9)	0,2
04 Mauricie et Centre-du-Québec	Poids insuffisant	3 900	14,7	5,2	0,76	(13,3 - 16,2)	0,2
	Poids normal	22 100	82,9	1,0	0,81	(81,2 - 84,4)	0,2
	Excès de poids	600	2,4	14,8	0,36	(1,7 - 3,2)	0,2
05 Estrie	Poids insuffisant	2 700	15,3	5,6	0,86	(13,7 - 17,1)	0,3
	Poids normal	14 400	82,6	1,1	0,87	(80,8 - 84,3)	0,3
	Excès de poids	400	2,1	13,7	0,29	(1,5 - 2,7)	0,3
06 Montréal	Poids insuffisant	15 000	15,1	5,9	0,90	(13,3 - 16,9)	0,4
	Poids normal	82 000	82,5	1,1	0,91	(80,7 - 84,3)	0,4
	Excès de poids	2 400	* 2,4	16,7	0,40	(1,6 - 3,3)	0,4
07 Outaouais	Poids insuffisant	3 100	15,4	5,8	0,89	(13,6 - 17,1)	0,3
	Poids normal	16 500	81,6	1,1	0,89	(79,9 - 83,3)	0,3
	Excès de poids	600	3,0	13,0	0,39	(2,2 - 3,8)	0,3
08 Abitibi-Témiscamingue	Poids insuffisant	1 200	13,7	5,8	0,80	(12,2 - 15,3)	0,1
	Poids normal	7 200	83,8	1,1	0,89	(81,9 - 85,4)	0,1
	Excès de poids	200	* 2,5	17,7	0,44	(1,7 - 3,5)	0,1

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 6.4

Répartition des élèves du secondaire selon l'évaluation de son apparence, pour l'ensemble du Québec, EQSJS 2010-2011

Variables de croisement		Indicateur	Statistiques					
Sexe	Perception de sa santé	Évaluation de son apparence	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	Passable ou mauvaise	Poids insuffisant	1 100	* 14,8	15,2	2,25	(10,3 - 19,4)	0,4
		Poids normal	5 000	65,0	4,8	3,10	(59,3 - 70,8)	0,4
		Excès de poids	1 600	20,1	12,2	2,46	(15,3 - 25,3)	0,4
	Excellente, très bonne ou bonne	Poids insuffisant	22 600	10,8	3,5	0,38	(10,0 - 11,5)	0,3
		Poids normal	182 100	86,7	0,5	0,39	(85,9 - 87,5)	0,3
		Excès de poids	5 400	2,6	7,5	0,19	(2,2 - 3,0)	0,3
Féminin	Passable ou mauvaise	Poids insuffisant	1 700	16,0	12,2	1,95	(12,1 - 20,0)	0,0
		Poids normal	8 000	73,6	3,2	2,39	(68,7 - 78,4)	0,0
		Excès de poids	1 100	* 10,4	15,8	1,64	(7,2 - 13,8)	0,0
	Excellente, très bonne ou bonne	Poids insuffisant	40 500	20,1	2,3	0,46	(19,2 - 21,0)	0,3
		Poids normal	159 200	78,9	0,6	0,47	(78,0 - 79,9)	0,3
		Excès de poids	1 900	1,0	11,5	0,11	(0,7 - 1,2)	0,3
Total	Passable ou mauvaise	Poids insuffisant	2 900	15,5	9,6	1,49	(12,6 - 18,3)	0,2
		Poids normal	13 000	70,1	2,8	1,97	(66,1 - 73,8)	0,2
		Excès de poids	2 700	14,4	9,8	1,42	(11,6 - 17,4)	0,2
	Excellente, très bonne ou bonne	Poids insuffisant	63 100	15,3	2,0	0,31	(14,7 - 16,0)	0,3
		Poids normal	341 400	82,9	0,4	0,30	(82,3 - 83,5)	0,3
		Excès de poids	7 400	1,8	6,3	0,11	(1,6 - 2,0)	0,3

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

♦ Deuxième groupe d'exemples : régional (Région sociosanitaire)

Pour les analyses à l'échelle régionale, le paramètre « Comparaison régionale » permet de définir le type d'analyse souhaitée. Les choix sont : « Aucune », « Région sociosanitaire de l'école comparée avec le reste du Québec » et « Région sociosanitaire de l'école comparée avec une autre région sociosanitaire ». Selon la situation, la sélection d'une seule variable de croisement est permise. L'algorithme de la section 6.2.3 peut être consulté pour identifier ces situations. Le tableau 6.5, obtenu à l'aide du choix « Aucune » pour le paramètre « Comparaison régionale », rapporte la consommation d'alcool au cours de la vie selon le niveau de scolarité de l'élève pour la région de l'Abitibi-Témiscamingue, en plus des résultats de l'ensemble du Québec. Si l'utilisateur souhaite vérifier si cette proportion d'élèves du secondaire ayant consommé de l'alcool au cours de leur vie est différente de celle du reste du Québec, il obtient l'information dans la colonne « Résultat du test » au tableau 6.6, à l'aide du choix « Région sociosanitaire de l'école comparée avec le reste du Québec » pour le paramètre « Comparaison régionale », de même que la proportion pour le reste du Québec. S'il souhaite vérifier si cette proportion d'élèves du secondaire ayant consommé de l'alcool au cours de leur vie est différente de celle d'une autre région, il obtient l'information dans la colonne « Résultat du test » au tableau 6.7, à l'aide du choix « Région sociosanitaire de l'école comparée avec une autre région sociosanitaire » pour le paramètre « Comparaison régionale ».

Tableau 6.5

Proportion des élèves du secondaire ayant consommé de l'alcool au cours de leur vie, EQSJS 2010-2011

Sous-population	Territoire	Statistiques					
Niveau de scolarité de l'élève	Région de l'école	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
1re secondaire	08 Abitibi-Témiscamingue	700	39,8	5,9	2,36	(35,2 - 44,4)	0,0
	Ensemble du Québec	24 900	28,9	2,7	0,79	(27,4 - 30,7)	0,0
2e secondaire	08 Abitibi-Témiscamingue	1 100	65,0	3,0	1,96	(61,1 - 69,3)	0,0
	Ensemble du Québec	47 000	52,1	2,3	1,19	(50,1 - 54,5)	0,0
3e secondaire	08 Abitibi-Témiscamingue	1 800	80,8	1,9	1,53	(77,8 - 83,9)	0,0
	Ensemble du Québec	66 800	69,5	1,3	0,93	(67,6 - 71,3)	0,0
4e secondaire	08 Abitibi-Témiscamingue	1 300	88,7	1,5	1,34	(86,1 - 91,1)	0,0
	Ensemble du Québec	65 900	79,3	1,2	0,91	(77,8 - 81,1)	0,0
5e secondaire	08 Abitibi-Témiscamingue	1 200	93,4	0,9	0,82	(91,8 - 95,1)	0,0
	Ensemble du Québec	65 400	86,7	0,8	0,70	(85,3 - 88,1)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 6.6

Proportion des élèves du secondaire ayant consommé de l'alcool au cours de leur vie, EQSJS 2010-2011

Territoire	Statistiques						
Région de l'école	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	6 200	72,4	1,2	0,85	(70,9 - 74,0)	0,0	sig
Reste du Québec	263 800	62,4	0,7	0,43	(61,7 - 63,4)	0,0	

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau 6.7

Proportion des élèves du secondaire ayant consommé de l'alcool au cours de leur vie, EQSJS 2010-2011

Territoire	Statistiques						
Région de l'école	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
08 Abitibi-Témiscamingue	6 200	72,4	1,2	0,85	(70,9 - 74,0)	0,0	sig
03 Capitale-Nationale	21 400	63,7	1,7	1,06	(61,7 - 65,8)	0,0	

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

♦ **Troisième groupe d'exemples : local (RLS) ou local (Commission scolaire)**

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 « Aucune » et « RLS comparés avec le reste de la région sociosanitaire de l'école » ou « Commissions scolaires comparées avec le reste des commissions scolaires de la région sociosanitaire de l'école » sont possibles. Encore une fois, selon la situation, la sélection d'une seule variable de croisement peut être permise. Le lecteur est invité à se référer à la section 6.2.3 pour identifier ces situations à l'aide des algorithmes. Le tableau 6.8 présente la sortie d'une requête à l'échelle locale, où aucune sous-population ni variable de croisement ne sont demandées, et pour laquelle la comparaison de chacun des RLS avec le reste de la région est requise (à l'aide du choix « RLS comparés avec le reste de la région sociosanitaire de l'école » pour le paramètre « Comparaison locale »), le territoire de RLS étant considéré comme une variable de croisement. L'utilisateur de l'Infocentre de santé publique 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 des élèves du secondaire selon le niveau d'activité physique de loisir et de transport, région sociosanitaire de l'école 03, Capitale-Nationale, EQSJS 2010-2011

Territoire (variable de croisement)	Indicateur	Statistiques				
RLS (CSSS)	Niveau d'activité physique	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
0301 - RLS de Portneuf (CSSS de Portneuf)	Actif	29,3	13,6	3,96	(21,4 - 36,9)	0,0
	Moyennement actif	21,9	13,4	2,93	(15,6 - 27,7)	0,0
	Un peu actif	* 10,9	23,9	2,61	(6,0 - 16,0)	0,0
	Très peu actif	18,3	14,3	2,61	(13,2 - 23,9)	0,0
	Sédentaire	* 19,6	15,9	3,13	(14,0 - 26,4)	0,0
0302 - RLS de Québec-Sud (CSSS de la Vieille-Capitale)	Actif	37,9	7,0	2,65	(32,8 - 43,3)	0,2
	Moyennement actif	21,2	7,0	1,49	(18,3 - 23,9)	0,2
	Un peu actif	* 6,3	17,0	1,08	(4,4 - 8,6)	0,2
	Très peu actif	14,4	9,0	1,30	(11,8 - 17,0)	0,2
	Sédentaire	20,2	8,9	1,80	(16,7 - 23,8)	0,0
0303 - RLS de Québec-Nord (CSSS de Québec-Nord)	Actif	27,7	7,1	1,98	(24,0 - 31,9)	0,4
	Moyennement actif	22,3	6,8	1,52	(19,2 - 25,2)	0,4
	Un peu actif	7,4	11,9	0,88	(5,6 - 9,2)	0,4
	Très peu actif	22,2	7,8	1,74	(19,1 - 25,8)	0,4
	Sédentaire	20,3	8,0	1,62	(17,1 - 23,4)	0,2
0304 - RLS de Charlevoix (CSSS de Charlevoix)	Actif	22,5	10,8	2,43	(16,8 - 26,9)	0,3
	Moyennement actif	* 22,2	16,8	3,72	(14,4 - 28,8)	0,3
	Un peu actif	* 6,8	23,6	1,60	(3,7 - 10,5)	0,3
	Très peu actif	16,2	12,6	2,05	(12,0 - 20,1)	0,3
	Sédentaire	* 32,3	17,7	5,74	(22,2 - 44,7)	0,0
Total de la région 03 Capitale-Nationale	Actif	32,2	4,7	1,50	(29,2 - 35,2)	0,3
	Moyennement actif	21,8	4,6	1,00	(19,7 - 23,5)	0,3
	Un peu actif	7,2	9,2	0,66	(5,9 - 8,5)	0,3
	Très peu actif	18,3	5,2	0,95	(16,3 - 20,3)	0,3
	Sédentaire	20,7	5,3	1,10	(18,4 - 23,0)	0,1

Test global d'indépendance entre l'indicateur et les RLS, région de l'école 03 Capitale-Nationale

Statistique
Valeur p[1]
0,0009

Notes

1 : Valeur p obtenue par le test du khi-deux.

· 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 à l'Institut national de santé publique du Québec, juin 2012.

6.3 Particularités des analyses produites à l'Infocentre de santé publique

Les prochaines sections décrivent certaines particularités des indicateurs qui ont un impact significatif pour l'analyse des données disponibles sur le portail de l'Infocentre de santé publique. Chaque sous-section présente un cas précis d'indicateur pour lequel des décisions ont été prises par le Comité tripartite MSSS-INSPQ-ISQ, et précise l'impact ou les mises en garde qui en découlent pour les analyses.

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, même si un test global est significatif, aucun test de comparaison des proportions (comparaisons deux à deux) 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 est affichée mentionnant que les comparaisons deux à deux 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 présentés dans les tableaux, afin de déterminer la source de la différence observée (méthode 2). Pour plus de détails sur la méthode 2, la section 5.1.2 peut être consultée.

Tableau 6.9
Répartition des élèves du secondaire de 14 ans et plus ayant déjà eu des relations sexuelles orales consensuelles, selon le nombre de partenaires à vie pour ce type de relation, pour l'ensemble du Québec, EQSJS 2010-2011

Variable de croisement	Indicateur	Statistiques					
Sexe	Nombre de partenaires	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	1 partenaire	21 700	44,4	1,9	0,84	(42,7 - 46,0)	1,3
	2 partenaires	11 100	22,7	2,9	0,66	(21,3 - 24,0)	1,3
	3 partenaires	6 000	12,2	5,4	0,66	(11,0 - 13,5)	1,3
	4 partenaires ou plus	10 100	20,7	4,0	0,84	(19,2 - 22,4)	1,3
Féminin	1 partenaire	23 900	46,9	1,8	0,85	(45,2 - 48,5)	0,6
	2 partenaires	11 700	22,9	2,8	0,63	(21,7 - 24,3)	0,6
	3 partenaires	6 400	12,5	4,5	0,57	(11,5 - 13,6)	0,6
	4 partenaires ou plus	8 900	17,6	3,5	0,62	(16,5 - 18,8)	0,6
Total	1 partenaire	45 600	45,7	1,4	0,66	(44,4 - 46,8)	1,0
	2 partenaires	22 800	22,8	1,9	0,44	(22,0 - 23,7)	1,0
	3 partenaires	12 300	12,4	3,5	0,43	(11,6 - 13,2)	1,0
	4 partenaires ou plus	19 100	19,1	2,9	0,56	(18,1 - 20,2)	1,0

Test global d'indépendance entre l'indicateur et le sexe

Statistique
Valeur p[1]
0,0106

Notes

1 : Valeur p obtenue par le test du khi-deux.

· 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 à l'Institut national de santé publique du Québec, juin 2012.

6.3.2 *Nombre trop important de tests demandés pour une requête*

Lorsque le nombre de comparaisons 2 à 2 est supérieur à 30, celles-ci ne sont pas produites lors de l'exécution de la requête et seuls les tests globaux du khi-deux le sont, un pour chacune des sous-populations définies par le paramètre « Sous-population », et ce, même si un test du khi-deux est significatif.

Lorsqu'une telle situation se présente, une note au bas du tableau est affichée mentionnant que le nombre de comparaisons deux à deux est trop élevé et que celles-ci ne sont 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 de déterminer la source de la différence observée (méthode 2). Pour plus de détails sur la méthode 2, la section 5.1.2 peut être consultée.

Tableau 6.10

Proportion des élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011 (extrait du tableau complet)

Sous-population		Variable de croisement	Statistiques					
Sexe	Niveau de scolarité de l'élève	Indice provincial de défavorisation matérielle	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	1re et 2e secondaire	Quintile 1 (très favorisé)	1 000	25,8	14,1	3,64	(18,6 - 34,2)	2,7
		Quintile 2	1 000	22,8	11,9	2,72	(17,5 - 28,2)	1,2
		Quintile 3	1 000	22,6	11,3	2,55	(17,7 - 27,9)	0,9
		Quintile 4	1 300	27,4	9,6	2,63	(22,0 - 33,0)	0,3
		Quintile 5 (très défavorisé)	1 300	25,4	9,6	2,43	(20,8 - 30,4)	2,0
	3e secondaire	Quintile 1 (très favorisé)	2 100	24,2	9,5	2,31	(20,1 - 29,3)	1,2
		Quintile 2	2 400	27,1	7,4	2,01	(23,2 - 31,1)	1,0
		Quintile 3	2 600	32,9	6,5	2,13	(28,7 - 37,3)	0,7
		Quintile 4	2 600	32,6	7,0	2,27	(28,4 - 37,6)	0,4
		Quintile 5 (très défavorisé)	2 600	32,9	7,1	2,33	(28,9 - 37,8)	0,6
	4e secondaire	Quintile 1 (très favorisé)	2 900	31,8	8,5	2,70	(26,8 - 37,0)	0,3
		Quintile 2	3 000	37,4	5,7	2,15	(33,3 - 41,7)	0,2
		Quintile 3	2 600	40,8	6,4	2,62	(35,6 - 46,0)	0,4
		Quintile 4	2 600	42,0	5,2	2,20	(37,6 - 46,4)	0,4
		Quintile 5 (très défavorisé)	2 300	39,5	6,6	2,59	(34,2 - 44,6)	0,3
	5e secondaire	Quintile 1 (très favorisé)	3 600	47,3	4,4	2,10	(43,3 - 51,6)	0,3
		Quintile 2	3 600	47,8	4,7	2,26	(43,4 - 52,3)	0,1

Test global d'indépendance entre l'indicateur et l'indice provincial de défavorisation matérielle, stratifié selon la sous-population

Sous-population		Statistique
Sexe	Niveau de scolarité de l'élève	Valeur p[1]
Masculin	1re et 2e secondaire	0,6643
	3e secondaire	0,0054
	4e secondaire	0,0104
	5e secondaire	0,9369
Féminin	1re et 2e secondaire	0,0632
	3e secondaire	<0,0001
	4e secondaire	<0,0001
	5e secondaire	0,0184

Notes

1 : Valeur p obtenue par le test du khi-deux.

· 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 à l'Institut national de santé publique du Québec, juin 2012.

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

♦ **Proportion de non-réponse partielle supérieure au seuil critique**

Comme il a 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-estimation ou surestimation)[63].

Chacun des tableaux produits sur le portail de l'Infocentre de santé publique affiche les pourcentages de non-réponse partielle de l'indicateur analysé. Afin d'aider les utilisateurs, une note est placée au bas du tableau lorsque cette non-réponse dépasse le seuil critique correspondant au niveau géographique considéré. Les tableaux 6.11 et 6.12 présentent des exemples de tableaux sans croisement affichant chacune de ces notes.

Tableau 6.11
Répartition des élèves du secondaire selon le statut pondéral, pour l'ensemble du Québec, EQSJS 2010-2011

Sous-population	Indicateur	Statistiques					
Sexe	Statut pondéral	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	Poids insuffisant	17 700	8,1	3,2	0,26	(7,6 - 8,6)	[a] 10,6
	Poids normal	145 600	66,7	0,7	0,46	(65,7 - 67,5)	[a] 10,6
	Embonpoint	38 200	17,5	1,9	0,33	(16,8 - 18,2)	[a] 10,6
	Obésité	16 900	7,7	3,2	0,25	(7,3 - 8,3)	[a] 10,6
Féminin	Poids insuffisant	26 400	12,4	2,6	0,32	(11,8 - 13,0)	[a] 11,9
	Poids normal	151 000	71,0	0,5	0,39	(70,1 - 71,7)	[a] 11,9
	Embonpoint	23 200	10,9	2,5	0,27	(10,4 - 11,5)	[a] 11,9
	Obésité	12 300	5,8	3,6	0,21	(5,4 - 6,2)	[a] 11,9
Total	Poids insuffisant	44 000	10,2	2,0	0,21	(9,8 - 10,6)	[a] 11,2
	Poids normal	296 500	68,8	0,4	0,30	(68,1 - 69,3)	[a] 11,2
	Embonpoint	61 400	14,2	1,6	0,22	(13,8 - 14,7)	[a] 11,2
	Obésité	29 200	6,8	2,5	0,17	(6,4 - 7,1)	[a] 11,2

Notes

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

63. Pour savoir comment procéder au diagnostic et à l'analyse qualitative de la non-réponse partielle, le lecteur est invité à consulter l'annexe 2.

Tableau 6.12

Répartition des élèves du secondaire selon le statut pondéral, EQSJS 2010-2011 (extrait du tableau complet)

Sous-population	Territoire	Indicateur	Statistiques					
Niveau de scolarité de l'élève	Région de l'école	Statut pondéral	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
1re secondaire	16 Montérégie	Poids insuffisant	1 900	11,1	10,7	1,19	(8,9 - 13,7)	** 24,8
		Poids normal	11 400	68,0	3,2	2,15	(63,6 - 71,9)	** 24,8
		Embonpoint	2 200	13,1	9,6	1,25	(10,5 - 15,5)	** 24,8
		Obésité	1 300	7,8	14,9	1,16	(5,7 - 10,2)	** 24,8
	Ensemble du Québec	Poids insuffisant	11 100	12,8	4,1	0,52	(11,9 - 13,9)	* 17,1
		Poids normal	57 500	66,7	1,1	0,72	(65,2 - 68,1)	* 17,1
		Embonpoint	11 000	12,7	3,9	0,50	(11,8 - 13,7)	* 17,1
		Obésité	6 600	7,7	5,0	0,38	(6,9 - 8,5)	* 17,1
2e secondaire	16 Montérégie	Poids insuffisant	1 700	* 9,5	15,6	1,49	(6,7 - 12,6)	** 15,7
		Poids normal	12 000	68,9	2,6	1,76	(65,1 - 72,1)	** 15,7
		Embonpoint	2 500	14,6	10,4	1,52	(11,5 - 17,7)	** 15,7
		Obésité	1 200	7,0	13,8	0,97	(5,2 - 9,2)	** 15,7
	Ensemble du Québec	Poids insuffisant	10 300	11,4	4,8	0,55	(10,4 - 12,4)	* 13,3
		Poids normal	60 200	66,8	1,0	0,68	(65,3 - 68,1)	* 13,3
		Embonpoint	12 900	14,3	3,3	0,47	(13,4 - 15,3)	* 13,3
		Obésité	6 800	7,5	6,0	0,45	(6,6 - 8,5)	* 13,3
3e secondaire	16 Montérégie	Poids insuffisant	1 600	8,7	8,7	0,76	(7,1 - 10,1)	** 12,0
		Poids normal	13 600	72,6	2,6	1,92	(68,5 - 76,0)	** 12,0

Notes

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

** : 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.

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

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 en général beaucoup plus faible que celle de l'indicateur sans croisement puisque chacun des domaines est un sous-ensemble de la population globale sur laquelle porte l'indicateur. 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

Répartition des élèves du secondaire selon le statut pondéral, pour l'ensemble du Québec, EQSJS 2010-2011

Sous-population	Variable de croisement	Indicateur	Statistiques					
Sexe	Perception de sa santé	Statut pondéral	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	Passable ou mauvaise	Poids insuffisant	900	11,1	14,7	1,64	(7,9 - 14,4)	‡ 13,8
		Poids normal	2 900	36,3	6,3	2,30	(31,8 - 40,7)	‡ 13,8
		Embonpoint	2 000	25,2	7,2	1,82	(21,4 - 29,0)	‡ 13,8
		Obésité	2 200	27,4	7,6	2,09	(23,3 - 31,3)	‡ 13,8
	Excellente, très bonne ou bonne	Poids insuffisant	16 800	8,0	3,3	0,26	(7,5 - 8,5)	‡ 10,4
		Poids normal	142 400	67,8	0,7	0,45	(66,9 - 68,6)	‡ 10,4
		Embonpoint	36 200	17,2	1,9	0,33	(16,5 - 17,9)	‡ 10,4
		Obésité	14 700	7,0	3,4	0,24	(6,6 - 7,5)	‡ 10,4
Féminin	Passable ou mauvaise	Poids insuffisant	1 100	10,7	11,8	1,27	(8,2 - 13,3)	‡ 15,9
		Poids normal	5 700	53,7	3,8	2,01	(49,7 - 57,8)	‡ 15,9
		Embonpoint	2 000	19,1	7,7	1,47	(16,3 - 22,3)	‡ 15,9
		Obésité	1 700	16,5	9,0	1,49	(13,7 - 19,5)	‡ 15,9
	Excellente, très bonne ou bonne	Poids insuffisant	25 200	12,5	2,7	0,34	(11,8 - 13,1)	‡ 11,7
		Poids normal	145 100	71,8	0,6	0,41	(71,0 - 72,6)	‡ 11,7
		Embonpoint	21 200	10,5	2,6	0,27	(9,9 - 11,0)	‡ 11,7
		Obésité	10 600	5,2	3,8	0,20	(4,9 - 5,7)	‡ 11,7
Total	Passable ou mauvaise	Poids insuffisant	2 000	10,9	9,0	0,98	(9,0 - 12,9)	‡ 15,0
		Poids normal	8 500	46,0	3,3	1,51	(42,9 - 49,1)	‡ 15,0
		Embonpoint	4 000	21,8	5,0	1,08	(19,5 - 24,3)	‡ 15,0
		Obésité	3 900	21,3	5,7	1,21	(18,7 - 23,7)	‡ 15,0
	Excellente, très bonne ou bonne	Poids insuffisant	41 900	10,2	2,1	0,21	(9,8 - 10,6)	‡ 11,0
		Poids normal	287 400	69,8	0,4	0,31	(69,1 - 70,3)	‡ 11,0
		Embonpoint	57 400	13,9	1,6	0,23	(13,5 - 14,4)	‡ 11,0
		Obésité	25 300	6,1	2,6	0,16	(5,8 - 6,5)	‡ 11,0

Notes

‡ : 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 « Perception de sa santé » est de 0,1 %.

· 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 à l'Institut national de santé publique du Québec, mai 2012.

♦ Sous-estimation des effectifs de population dans les tableaux avec croisement

Il a été mentionné à la section 2.3.5 que les estimations d'effectifs diffusés sur le portail de l'Infocentre de santé publique sont ajustées afin de tenir compte de la non-réponse partielle. Aucun traitement de la non-réponse partielle de la variable de croisement n'est toutefois effectué, c'est-à-dire que la non-réponse partielle de la variable de croisement n'est pas répartie au sein de ses différentes catégories. Seule la non-réponse de l'indicateur l'est pour le sous-groupe des répondants à la variable de croisement, de sorte que les estimations des effectifs qui apparaissent dans les tableaux avec croisement sous-estiment les effectifs réels de la population[64]. Afin de faire état de cette limite, une note est affichée, le cas échéant, au bas du tableau. Le tableau 6.14 est un exemple de tableau affichant cette note.

64. 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 d'élèves qui ont consommé de la drogue au cours des 12 derniers mois parmi ceux qui ont un emploi durant l'année scolaire, il faut prendre la proportion des élèves qui ont consommé de la drogue au cours des 12 derniers mois parmi ceux qui ont un emploi durant l'année scolaire dans le tableau croisé et multiplier cette proportion par l'effectif des élèves qui ont consommé de la drogue au cours des 12 derniers mois du tableau sans croisement.

Tableau 6.14

Proportion des élèves du secondaire ayant consommé de la drogue au cours des 12 derniers mois, pour l'ensemble du Québec, EQSJS 2010-2011

Variable de croisement	Statistiques					
Emploi durant l'année scolaire	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Oui	56 000	30,3	1,7	0,51	(29,5 - 31,4)	0,4
Non	54 900	22,2	2,3	0,51	(21,3 - 23,3)	0,2

Notes

· Pour ce tableau, la non-réponse partielle pour la variable « Emploi durant l'année scolaire » est de 0,1 %.

· 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 à l'Institut national de santé publique du Québec, mai 2012.

6.3.4 Diffusion par réseau local de services (RLS) ou par commission scolaire

♦ **Disponibilité pour huit régions sociosanitaires de l'école seulement**

Comme il a été 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 avec suréchantillon (pour la diffusion de statistiques par RLS : Abitibi-Témiscamingue, Gaspésie–Îles-de-la-Madeleine, Lanaudière, Mauricie et Centre-du-Québec, Estrie, Capitale-Nationale et Montérégie, et pour la diffusion de statistiques par commission scolaire : Estrie et Chaudière-Appalaches) et l'échelle régionale pour les autres régions sociosanitaires. 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 mises au point pour l'EQSJS 2010-2011 n'avaient 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 de santé publique.

Pour la diffusion par commission scolaire, les écoles privées sont exclues. Le total affiché est celui des commissions scolaires de la région, à l'exclusion des écoles privées. Le tableau 6.15 présente un exemple fictif de sortie de l'Infocentre de santé publique.

Tableau 6.15
Exemple fictif de diffusion par commission scolaire à l'Infocentre de santé publique

Territoire	Statistiques				
Commission scolaire	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
271000 - Commission scolaire A	4,3	11,8	0,50	(3,4 - 5,3)	0,0
272000 - Commission scolaire B	4,6	7,8	0,36	(3,9 - 5,4)	0,2
273000 - Commission scolaire C	3,8	11,9	0,45	(3,0 - 4,8)	0,0
274000 - Commission scolaire E	3,0	8,9	0,27	(2,5 - 3,5)	0,1
Total des commissions scolaires de la région 27 Nord	4,6	7,8	0,36	(3,9 - 5,4)	0,1

Source : Données fictives.

6.3.5 Analyse des données par niveau scolaire

En raison de la façon dont a été planifiée l'enquête, il est recommandé d'analyser les données par niveau scolaire plutôt que par groupe d'âge. En effet, le plan d'échantillonnage de l'EQSJS 2010-2011 garantit une représentativité par niveau scolaire et non par âge, puisque la population visée n'assure pas une couverture complète des jeunes de 12 à 17 ans. De plus, comme la stratification du plan d'échantillonnage tient compte du niveau scolaire plutôt que de l'âge, les estimations selon l'âge souffrent d'une perte de précision par rapport à celles relatives au niveau scolaire. Quoi qu'il en soit, une analyse par groupe d'âge peut tout de même être réalisée lorsqu'il n'est pas pertinent d'étudier un indicateur par niveau scolaire. Son interprétation doit cependant tenir compte des aspects de représentativité énoncés précédemment. Les deux possibilités d'analyse sont offertes à l'Infocentre de santé publique.

Il est par ailleurs important de noter que la répartition des élèves selon le parcours de formation déclaré (après validation) varie d'un niveau scolaire à l'autre, comme il est illustré au tableau 6.16. En particulier, la proportion d'élèves des parcours de formation générale est plus faible en 3e secondaire et, dans une moindre mesure, en 1re secondaire. La classification de tous les élèves de FMSS en 3e secondaire, peu importe leur âge, contribue à abaisser cette proportion. Étant donné que les parcours de formation autre que générale s'adressent pour la plupart à des élèves en difficulté, ces variations peuvent entraîner une cassure dans l'évolution de certains indicateurs de la 1re à la 5e secondaire. Cela se produit dans le cas des indicateurs selon lesquels les élèves des parcours de formation générale diffèrent des autres de façon importante. L'interprétation des résultats doit tenir compte de cette possibilité. Une variable de croisement indiquant le type de parcours (formation générale ou autre) a été créée aux fins d'analyse.

Tableau 6.16
Répartition des élèves selon le parcours de formation déclaré, par niveau scolaire

Niveau scolaire	Formation générale	Formation axée sur l'emploi		Autre
	(%)	FPT (%)	FMSS (%)	(%)
1re secondaire	91,8			8,2
2e secondaire	95,1			4,9
3e secondaire	87,1	2,5	5,1	5,3
4e secondaire	95,0	1,7		3,3
5e secondaire	98,2	0,9		0,9

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

6.3.6 *Région sociosanitaire de l'école*

Rappelons que le plan d'échantillonnage a été élaboré de manière à produire des estimations régionales de bonne précision, selon la région sociosanitaire de l'école fréquentée par l'élève et non selon la région de résidence de celui-ci, et que pour la grande majorité des régions, on observe une bonne concordance entre ces deux variables (voir section 1.3.1). De plus, le nombre de valeurs manquantes pour le code postal de l'élève, qui sert à construire la variable de région de résidence, est élevé. Par conséquent, à l'Infocentre de santé publique, les analyses par région sociosanitaire ne sont effectuées que pour la région où est située l'école de l'élève.

6.3.7 *Écarts d'âge régionaux*

La distribution estimée de l'âge des élèves varie quelque peu entre les régions sociosanitaires. En particulier, les régions du Nord-du-Québec et de la Côte-Nord présentent une proportion légèrement plus élevée d'élèves de 12 ans et moins que les autres régions. Cette situation s'explique en partie par le fait que la collecte a eu lieu plus tôt dans ces régions, surtout dans le Nord-du-Québec. Toutefois, puisque les légers écarts observés dans les répartitions d'âge d'une région à l'autre ne suivent pas nécessairement les variations des périodes de collecte, il est permis de croire qu'une part de ces écarts d'âge demeure inhérente aux régions. La standardisation ou non des statistiques régionales selon l'âge, aux fins de comparaison, relève par conséquent de l'angle d'analyse souhaité. Quelle que soit l'option retenue, il importe de garder à l'esprit que la région Nord-du-Québec a eu une collecte sensiblement plus hâtive et que des précautions doivent être prises dans sa comparaison avec les autres régions. La section 1.6.3 fournit quelques détails additionnels sur cette question. La section 4 aborde quant à elle la question de la standardisation des proportions selon l'âge de l'élève.

6.3.8 *Indicateurs d'âge d'initiation*

L'EQSJS 2010-2011 comporte plusieurs mesures de l'âge d'initiation à certains comportements, soit aux relations sexuelles avec consentement, à la cigarette, à l'alcool et aux drogues. Ces mesures sont liées à l'âge réel de l'élève au moment de l'enquête. Un élève enquêté à l'âge de 12 ans ne pourrait pas déclarer un âge d'initiation supérieur à 12 ans, et ainsi de suite. Il est malaisé d'interpréter des statistiques telles que la proportion d'élèves ayant amorcé un comportement ou encore l'âge moyen d'initiation à un comportement lorsqu'elles portent sur des élèves d'âge variable. Par exemple, s'il était statué que l'âge moyen d'initiation à la consommation d'alcool est de 14 ans pour l'ensemble des élèves du secondaire, que cela signifierait-il pour les élèves de moins de 14 ans? Pour les indicateurs disponibles à l'Infocentre de santé publique, il a donc été décidé de considérer l'initiation des comportements uniquement pour la période antérieure à l'âge au moment de l'enquête de l'ensemble des élèves pris en compte pour l'estimation. Par exemple, l'indicateur de la proportion des élèves de 13 ans et plus ayant amorcé un comportement avant l'âge de 13 ans exclut de son dénominateur tous les élèves n'ayant pas atteint l'âge de 13 ans. D'autres angles d'analyse sont également envisageables.

Deux indicateurs d'âge d'initiation peuvent être croisés entre eux seulement s'ils portent sur un même groupe d'âge. Ainsi, il est possible de croiser la proportion des élèves du secondaire de 13 ans et plus ayant consommé de la drogue pour la première fois avant l'âge de 13 ans avec le fait d'avoir consommé ou non de l'alcool avant l'âge de 13 ans pour les élèves de 13 ans et plus. Toutefois, il n'est pas possible de croiser la proportion des élèves du secondaire de 12 ans et plus ayant consommé de la drogue pour la première fois avant l'âge de 12 ans avec le fait d'avoir consommé ou non de l'alcool avant l'âge de 13 ans pour les élèves de 13 ans et plus.

6.3.9 Emploi durant l'année scolaire

Bien que cette enquête vise à mesurer le travail des élèves du secondaire durant l'année scolaire, la question utilisée fait référence à un emploi actuel. Comme la collecte des données s'est déroulée de novembre 2010 à mai 2011, il est possible que des jeunes ayant eu un emploi durant l'année scolaire n'aient pas travaillé au moment de la collecte des données. Ainsi, la proportion des élèves occupant un emploi durant l'année scolaire pourrait être sous-estimée.

6.3.10 Point de césure des indicateurs selon les centiles

Pour certains indicateurs de l'EQSJS, il a été convenu de retenir les centiles (quintile, quartile ou tercile) afin de décrire les sous-groupes les plus vulnérables :

- Quintile
  - Proportion des élèves du secondaire se situant à un niveau élevé à l'indice de détresse psychologique
  - Proportion des élèves du secondaire se situant à un niveau faible à l'échelle d'estime de soi
  - Proportion des élèves du secondaire se situant à un niveau élevé à l'indice d'autocontrôle
  - Proportion des élèves du secondaire se situant à un niveau élevé à l'indice de risque de décrochage scolaire

- Quartile
  - Proportion des élèves du secondaire se situant à un niveau élevé aux indices d'inattention et d'hyperactivité

- Tercile

  - Proportion des élèves du secondaire dont les parents ont un niveau élevé de supervision parentale

Le choix d'utiliser les centiles pour décrire les sous-groupes plus vulnérables a toutefois comme limite que l'on ne peut interpréter les indicateurs ainsi définis à titre de proportions estimées, puisque les seuils sont obtenus par définition. En conséquence, ces indicateurs ne peuvent être utilisés que dans un but comparatif, en les croisant avec d'autres variables. Par exemple, bien que 20,8 % de la population visée se situe au niveau élevé à l'indice de détresse psychologique, il serait incorrect de rapporter cette proportion comme un résultat. Les proportions présentées doivent seulement servir pour la comparaison de groupes d'individus et pour cibler des groupes à risque, soit des groupes comprenant de plus grandes proportions d'élèves se situant au niveau élevé de l'indice. Notons que bien qu'elle représente la proportion supérieure au quintile supérieur, la proportion de 20,8 % n'est pas exactement égale à 20 %, uniquement en raison des scores égaux à ce quintile obtenus par plusieurs élèves. Sur le portail de l'Infocentre de santé publique, une note au bas du tableau de résultats est automatiquement produite pour ces indicateurs (voir le tableau 6.17). Plus de détails sont fournis dans la fiche des indicateurs concernés.

La construction de ce type d'indicateurs implique que les seuils utilisés pour les définir ne sont pas toujours les mêmes d'une enquête à l'autre, en raison notamment des différences dans les populations étudiées. Les comparaisons entre différentes populations, voire entre différentes enquêtes, telles que l'*Enquête québécoise sur la santé de la population* (EQSP) de l'ISQ et les cycles généraux de l'*Enquête sur la santé dans les collectivités canadiennes* (ESCC) de Statistique Canada, pour lesquelles des seuils différents sont appliqués, ne sont donc pas pertinentes.

Tableau 6.17

Proportion des élèves du secondaire se situant à un niveau élevé à l'indice de détresse psychologique pour l'ensemble du Québec, EQSJS 2010-2011

Statistiques					
Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
89 700	20,8	1,9	0,39	(20,1 - 21,6)	0,3

Notes

1 : Population estimée (arrondie à la centaine) des élèves du secondaire se situant à un niveau élevé à l'indice de détresse psychologique. Puisque la non-réponse partielle a été répartie selon la proportion observée pour chaque catégorie de la variable de croisement (ou de la sous-population ou du territoire), la somme des estimations de population obtenues ne correspond pas nécessairement à l'estimation de population pour les catégories regroupées.

- I.C. : Intervalle de confiance. C.V. : Coefficient de variation.

- Les résultats présentés sont basés uniquement sur le questionnaire 2 puisqu'au moins une des questions utilisées pour produire la requête est spécifique au questionnaire 2.

- Les seuils de cet indicateur sont basés sur les centiles, il n'est donc pas possible de le comparer avec d'autres enquêtes.

6.3.11 Biais de désirabilité sociale

Les mesures autorapportées sont sujettes au biais de désirabilité sociale, c'est-à-dire que pour des questions portant sur un sujet délicat, le répondant peut avoir tendance à répondre de façon à projeter une image de lui favorable aux autres (Fisher, 1993). Bien que ce biais puisse être présent pour tous les indicateurs de l'EQSJS, certains sujets pourraient être plus délicats que d'autres.

7. Protection des renseignements personnels[65]

La protection des renseignements personnels (PRP) se fait à deux niveaux : 1) les données individuelles des répondants (masquage du fichier de microdonnées - section 7.1) et 2) les données agrégées (masquage des tableaux – sections 7.2 et 7.3).

7.1 Confidentialité des fichiers de microdonnées

Le traitement de confidentialité appliqué au fichier disponible en accès à distance (FMII) est léger et vise à éviter l'identification involontaire des élèves, ce qui lui confère un bon potentiel analytique. Le masquage est de trois types : la suppression de valeurs, le regroupement de catégories et l'élimination de variables. Les utilisateurs qui ont accès à ce fichier peuvent consulter la documentation afférente pour en savoir davantage.

7.2 Confidentialité des tableaux destinés à la diffusion

Bien que le fichier de microdonnées de l'EQSJS 2010-2011 disponible à l'Infocentre de santé publique soit un fichier non masqué, en raison de l'accord du répondant à communiquer des renseignements à cet organisme, il existe, pour celui qui diffuse des données sous forme de tableaux, une obligation contractuelle de respecter leur caractère confidentiel. 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*[66] », donne une définition assez complète du risque de divulgation et de nombreux exemples. L'encadré présenté à la fin de cette section-ci est un extrait tiré de ce document. Il fournit de l'information supplémentaire sur la divulgation d'information lors de la diffusion de tableaux de résultats, dans le contexte général d'une enquête.

Dans l'EQSJS 2010-2011, l'utilisateur doit non seulement veiller à protéger la confidentialité des données individuelles, mais également de celles associées aux écoles. Il doit, pour ce faire, s'assurer que les résultats diffusés ne puissent être attribués à une école en particulier. Autrement dit, une statistique ne peut être diffusée si une école donnée y contribue de façon trop importante. Ces précautions s'appliquent tant pour la diffusion de tableaux issus du fichier maître que de ceux issus du FMII.

L'ISQ s'est doté d'un ensemble de règles de confidentialité pour parer à la divulgation de renseignements confidentiels résultant de la diffusion de tableaux. Leur application permet de déterminer quels tableaux ou résultats comportent un risque de divulgation et quels traitements de confidentialité doivent leur être appliqués. L'Infocentre de santé publique est soumis aux mêmes obligations que l'ISQ pour les données de l'EQSJS 2010-2011. C'est pourquoi les utilisateurs de données de cette enquête doivent appliquer des règles équivalant à celles établies par l'ISQ pour la diffusion[67] de tableaux. Ces règles sont présentées à la section 7.3 dans le contexte de la diffusion de résultats obtenus à partir de l'Infocentre de santé publique, mais pourraient tout aussi bien s'appliquer au contexte de la diffusion de résultats obtenus de l'ISQ par l'accès à distance.

65. Certains éléments de cette section ont été extraits intégralement de documents de l'ISQ sur la confidentialité des microdonnées et des tableaux.
66. 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.
67. Le terme diffusion comprend, entre autres choses, la publication, la sortie d'un résultat hors des locaux sécurisés de l'organisme, la transmission du résultat, qu'elle soit sous forme papier, informatisée ou verbale, etc.

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 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 tous les élèves 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.

7.3 Règles de masquage par suppression en présence de données à risque d'identification

Afin de réduire le risque d'identification des répondants, lors de la production des tableaux statistiques par l'intermédiaire du portail de l'Infocentre de santé publique, les règles de masquage par suppression, présentées dans cette section et insérées en annexe de toutes les fiches des indicateurs, ont été appliquées pour les utilisateurs ayant un droit d'accès 40 ou 50.

Dans cette enquête, deux situations peuvent se présenter où des données sont à risque d'identification, et dans chacune d'elles il y a des obligations à respecter en termes de diffusion. La première a trait à la confidentialité au niveau de l'école, alors que la seconde touche l'élève.

Premièrement, si pour un indicateur, l'utilisateur ayant un droit d'accès 20 ou 30 est en présence d'un tableau incluant une école qui représente une part importante des élèves associés à la population considérée par la proportion, il doit renoncer à la diffusion de cette proportion ou de cette répartition. Dans les tableaux produits à l'Infocentre de santé publique, une telle situation est identifiée par le symbole « † » à la droite du nombre (population estimée). Le tableau 7.1 montre un exemple fictif de sortie de l'Infocentre de santé publique.

Deuxièmement, si pour un indicateur, l'utilisateur ayant un droit d'accès 20 ou 30 est en présence d'un tableau ayant au moins une cellule jugée confidentielle au niveau de l'élève (dans les tableaux produits à l'Infocentre de santé publique, une telle situation est identifiée par le symbole « ^{aaa} » à la gauche du nombre (population estimée)), il doit choisir parmi l'une ou l'autre des solutions suivantes (voir le tableau 7.1 pour un exemple fictif de sortie de l'Infocentre de santé publique) :

- renoncer à la diffusion de ce résultat;
- utiliser un indicateur pour lequel les catégories problématiques sont regroupées avec d'autres catégories;
- utiliser une autre catégorisation de la variable de croisement, le cas échéant;
- masquer la cellule problématique du tableau de même qu'une autre cellule, afin de ne pas pouvoir déduire par une simple soustraction la valeur à masquer. Ce masquage par suppression doit alors suivre les règles suivantes :
    - ne pas présenter les nombres estimés lors de la diffusion des données;
    - ne pas présenter la répartition pour un indicateur en répartition si deux cellules sont composées de moins de cinq répondants;
    - utiliser le masquage propre à l'indicateur, ainsi qu'il est présenté dans la fiche correspondante.

Tableau 7.1

Exemple fictif de sortie révélant qu'un utilisateur est en présence d'un tableau incluant une école qui détient une part importante des élèves associés à la population considérée par la proportion (symbole « † ») et dont au moins une cellule est jugée confidentielle au niveau de l'élève (symbole « ^{aaa} »)

Territoire	Variable de croisement	Statistiques					
RLS (CSSS)	Structure familiale de l'élève	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
7510 - RLS A (CSSS A)	Biparentale	^{aaa}0 †	* 3,4	18,0	0,61	(2,2 - 4,7)	0,0
	Reconstituée	^{aaa}0 †	** 6,0	45,8	2,73	(0,7 - 12,1)	0,0
	Monoparentale	^{aaa}0 †	* 5,7	16,9	0,96	(3,7 - 7,8)	0,0
	Garde partagée	^{aaa}0 †	** 4,8	40,9	1,96	(1,2 - 9,2)	0,0
	Autre	^{aaa}0 †	** 10,1	27,9	2,82	(4,4 - 16,8)	0,0

Notes

^{aaa} : 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'EQSJS et dans la fiche-indicateur.

† : Une école représente une part importante des élèves associés à la population considérée par la proportion. 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'EQSJS.

Source : Données fictives.

Le tableau 7.2 présente les symboles utilisés dans l'application du masquage, et le tableau 7.3 montre un exemple de celui-ci pour un indicateur de l'EQSJS 2010-2011.

Tableau 7.2
Symboles utilisés dans l'application du masquage

Symbole	Signification
< 5	cellule où le nombre de répondants est inférieur à 5, donc qui doit être supprimée
≥ 5	cellule où le nombre de répondants est au moins de 5, mais qui doit être supprimée en raison de la présence d'une cellule complémentaire inférieure à 5
≥ 5	cellule où le nombre de répondants est au moins de 5 et qui peut être présentée

Tableau 7.3
Exemple de masquage spécifique présenté dans la fiche de l'indicateur sur la proportion des élèves du secondaire ne se percevant pas en bonne santé

Situation	Excellente, très bonne ou bonne	Passable ou mauvaise
1	**≥ 5**	**≥ 5**
2	**< 5**	**≥ 5**
3	**≥ 5**	**< 5**

III. Recommandations pour l'interprétation des résultats

Les recommandations présentées dans cette partie 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 catégoriques. 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 donc pas dire :

- « Le statut d'activité et la catégorie professionnelle influencent aussi la qualité de l'alimentation… »
- « Les données confirment l'effet marqué du sexe… »

mais plutôt:

- « 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 forte… »

2. Inférence à la population visée

Il ne faut pas décrire les résultats en faisant référence aux répondants; comme les données sont pondérées, on doit les présenter en les associant à la population visée. Ainsi, on ne doit pas dire :

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

mais plutôt:

- « 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 la situation suivante :

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

3. Référence à un nombre

Lors de la comparaison de proportions, l'interprétation des résultats ne doit pas laisser supposer qu'on fait référence à des nombres, mais bel et bien à des proportions. Ainsi, il ne faut pas utiliser des formulations comme :

- « Un peu plus de Québécois se disent… » ou
- « Moins de femmes que d'hommes… »
- « Les élèves de tel groupe ayant telle caractéristique sont plus nombreux que ceux de tel autre groupe ayant les mêmes caractéristiques. »

mais plutôt:

- « Ces élèves sont plus nombreux, en proportion, à… »
- « Il y a proportionnellement plus de femmes que d'hommes… »
- « Toutes proportions gardées, il y a plus de… »
- « La proportion d'élèves de tel groupe ayant telle caractéristique est plus grande que celle observée dans tel autre groupe. »

4. Estimations et non 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 est en général inadéquate :

- « 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, on estime que 31 % des gens souffrent de la maladie… »

5. Différence non statistiquement significative

Un test statistique effectué dans le cadre d'une enquête par sondage permet de comparer deux sous-groupes ou d'étudier l'association entre deux variables. Puisque le seuil de signification du test est contrôlé, la probabilité de déclarer un résultat significatif alors qu'il ne l'est pas demeure faible (ex. : seuil de 5 %). À l'inverse, en l'absence d'un résultat significatif, on ne peut déclarer qu'il n'y a pas de différence ou d'association entre les caractéristiques, car on ne contrôle pas la puissance du test statistique. On peut toutefois mentionner que l'enquête ne permet pas de déceler de différence ou d'association. Ainsi, on ne doit pas écrire :

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

mais plutôt:

- « 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. »

De même, lorsque l'on compare deux proportions qui ne sont pas statistiquement différentes, on ne peut pas dire qu'elles 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, car ils font référence à des résultats 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 des résultats sous la 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. Par exemple, il se pourrait que ce ratio ne soit pas significativement différent de trois. Ainsi, on ne devrait pas dire :

- « Les proportions enregistrées chez les femmes sont deux fois supérieures à celles des hommes
  (38 % c. 20 %)… »

mais plutôt:

- « La proportion est plus élevée chez les femmes que chez les hommes (38 % c. 20 %)… »

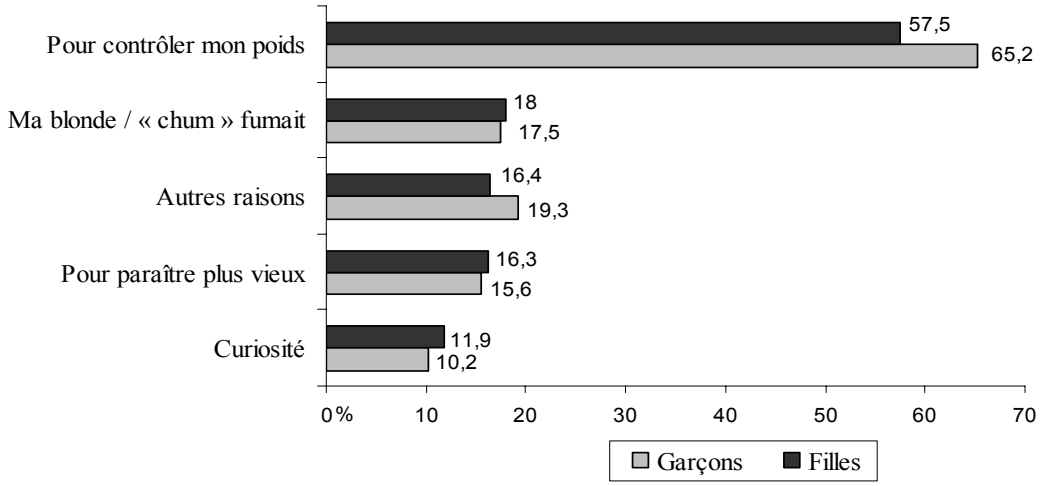
On pourrait toutefois utiliser la borne inférieure de l'intervalle de confiance du ratio pour une interprétation sous forme de ratio, si celle-ci demeure informative. Si cette borne inférieure est de 1,5, par exemple, on pourrait dire :

- « La proportion est au moins 1,5 fois plus élevée chez les femmes que chez les hommes… »

7. Présentation de résultats selon l'ordre des estimations

Lorsqu'on décrit certaines caractéristiques, on peut être tenté de les présenter selon l'ordre (croissant ou décroissant) des estimations. Ordonner ainsi les résultats selon la valeur des estimations peut laisser croire qu'il existe une différence significative entre chacun d'eux. Si l'on souhaite exposer les résultats de cette façon, on devrait s'assurer que des tests de comparaison entre les éléments ordonnés ont été effectués et que leur interprétation en tient compte. Au minimum, une identification des différences qui sont significatives ou un court texte devrait accompagner ces résultats.

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. Interprétation sous forme de 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 légèrement supérieur au seuil de signification fixé et qu'une ou des enquêtes similaires ont déjà permis de détecter un lien significatif entre ces mêmes variables, il est possible de parler de tendance. En effet, même si notre enquête n'a pu déceler un lien significatif (cela peut être dû à une taille d'échantillon trop petite), une tendance dans le sens de la relation détectée par les autres études peut être intéressante à discuter.

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

9. Interprétation des résultats pour les indicateurs basés sur les centiles

En raison de l'utilisation des centiles pour définir les points de césure de certains indicateurs, les proportions estimées pour ceux-ci doivent seulement servir pour comparer des groupes d'individus et pour cibler des groupes à risque. Par exemple, il n'est pas pertinent de présenter la proportion estimée d'élèves se situant au niveau élevé de l'indice de détresse psychologique comme un résultat de l'enquête, puisqu'il s'agit de la proportion d'élèves dont le score est supérieur ou égal au dernier quintile de la répartition des scores sous-jacents (proportion qui est d'environ 20 %, par définition). Il faut plutôt dire que la proportion de tel groupe ayant telle caractéristique est plus grande que celle observée dans tel autre groupe. La proportion globale peut aussi être utilisée comme valeur de référence pour situer un groupe par rapport à l'ensemble des élèves, sans toutefois mettre cette proportion à l'avant-plan. Par exemple, on pourra dire que la proportion d'élèves du secondaire se situant à un niveau élevé à l'indice de détresse psychologique est de tant chez les garçons et de tant chez les filles, comparativement à 21 % globalement.

AUBIN, Jacinthe, et autres (2002). *Enquête sociale et de santé auprès des enfants et des adolescents québécois 1999*, Québec, Institut de la statistique du Québec, 520 p.

COCHRAN, William G. (1977). *Sampling Techniques*, 3rd Ed., New York, John Wiley & Sons, 428 p.

DUBÉ, Gaëtane, et autres (2009). *Enquête québécoise sur le tabac, l'alcool, la drogue et le jeu chez les élèves du secondaire, 2008,* Québec, Institut de la statistique du Québec, 222 p.

FISHER, Robert J. (1993). « Social Desirability Bias and the Validity of Indirect Questioning », *Journal of Consumer Research*, vol. 20, n° 2, septembre, p. 303-315.

FLEISS, Joseph L. (1981). *Statistical Methods for Rates and Proportions*, 2nd Ed., New York, John Wiley & Sons, 321 p.

HAZIZA, David, Katherine J. THOMPSON, et Wesley YUNG (2010). « L'effet des ajustements pour la non-réponse sur l'estimation de la variance », *Techniques d'enquête*, vol. 36, n° 1, juin, p. 39-48.

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 (2009). *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.

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 (2010). *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, 117 p.

INSTITUT DE LA STATISTIQUE DU QUÉBEC (2013). *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011. Cahier technique : livre de codes et définition des indices (Fichier maître),* Québec, Gouvernement du Québec, 857 p.

KORN, Edward L. et Barry I. GRAUBARD (1999). *Analysis of Health Surveys,* New York, John Wiley & Sons, 382 p.

LOHR, Sharon. (1999). *Sampling: Design and Analysis*, 1st Ed., Duxbury Press, USA.

MUECKE, Christine, et autres (2005). *Doit-on utiliser la standardisation directe ou indirecte dans l'analyse géographique de la mortalité à l'échelle des petites régions ?*, Institut national de santé publique du Québec, Montréal, 28 p.

PIRES, Ana M., et Conceição AMADO (2008). « Interval estimators for a binomial proportion: comparison of twenty methods. *REVSTAT – Statistical Journal* », vol. 6, n° 2, juin, p. 165-197.

PLANTE, Nathalie, Robert COURTEMANCHE et Lyne DES GROSEILLIERS, avec la collaboration de Suzanne GINGRAS (2012). « Aspects méthodologiques », dans *L'Enquête québécoise sur la santé des jeunes du secondaire 2010-2011. Le visage des jeunes d'aujourd'hui : leur santé physique et leurs habitudes de vie*, Tome 1, Québec, Institut de la statistique du Québec, p. 31-52.

RUST, Keith F., et Jon N.K. RAO (1996). « Variance estimation for complex surveys using replication techniques », *Statistical Methods in Medical Research,* vol. 5, n° 3, septembre, p. 283-310.

SCHENKER, Nathaniel, et Jane F. GENTLEMAN (2001). « On judging the significance of differences by examining the overlap between confidence intervals », *The American statistician*, vol. 55, n° 3, août, p. 182-186.

STATISTIQUE CANADA (2003). *Méthodes et pratiques d'enquête,* produit n° 12-587-X au catalogue de Statistique Canada, Ottawa.

STATISTIQUE CANADA (2011). *Guide de l'utilisateur – Fichiers de microdonnées de 2010 et 2009-2010 de l'ESCC – composante annuelle*, 106 p.

WOLTER, Kirk M. (1985). *Introduction to Variance Estimation*, New York, Springer-Verlag, 428 p.

Annexe 1

Comment utiliser la pondération lors de la production d'estimations à partir des données de l'EQSJS 2010-2011

Afin de sensibiliser l'utilisateur des données de l'EQSJS 2010-2011 à l'importance d'avoir recours à 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'enquête. Le premier (tableau A1.1) affiche une estimation des fréquences pondérées de l'indicateur sur la consommation d'alcool au cours de la vie, alors que le second (tableau A1.2) en présente la version non pondérée. Les deuxième et troisième parties exposent respectivement la façon de pondérer les données à l'aide du logiciel SAS et la façon de les pondérer avec le logiciel SPSS.

Partie 1 – Exemples de tableaux produits à partir des données du fichier maître de l'EQSJS 2010-2011

Tableau A1.1
Proportions pondérées de l'indicateur sur la consommation d'alcool au cours de la vie, pour l'ensemble du Québec, EQSJS 2010-2011

HV_J_3 (A déjà consommé (bu) de l'alcool au cours de la vie)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP	.	.	.	.
Oui	270 063	62,6	270 063	62,6
Non	161 037	37,4	431 100	100,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau A1.2
Proportions non pondérées de l'indicateur sur la consommation d'alcool au cours de la vie, pour l'ensemble du Québec, EQSJS 2010-2011

HV_J_3 (A déjà consommé (bu) de l'alcool au cours de la vie)	Nombre de répondants	Proportion	Nombre de répondants cumulatif	Proportion cumulative
NRP	.	.	.	.
Oui	42 732	67,6	42 732	67,6
Non	20 464	32,4	63 196	100,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Les tableaux A1.1 et A1.2 montrent bien l'importance d'utiliser la pondération, puisque la proportion des élèves du secondaire ayant consommé de l'alcool au cours de leur vie passe de 67,6 % (non pondéré) à 62,6 % (pondéré), une différence de 5,0 %. Cette différence pourrait avoir un impact sur les conclusions tirées des résultats.

Partie 2 - Utilisation de la pondération à l'aide du logiciel SAS - fichier maître de l'EQSJS 2010-2011

Syntaxe en SAS

```
libname lib "\...répertoire où se trouvent les données...\";
proc freq data= lib. Nom du fichier;
tables HV_J_3;
weight POIDSCOMM[68];
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'EQSJS 2010-2011

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 est « **POIDSCOMM** ».

Les données sont maintenant pondérées pour effectuer les analyses.

Menus et Souris

Ouverture de la base de données :
Dans le menu, choisir « File »;

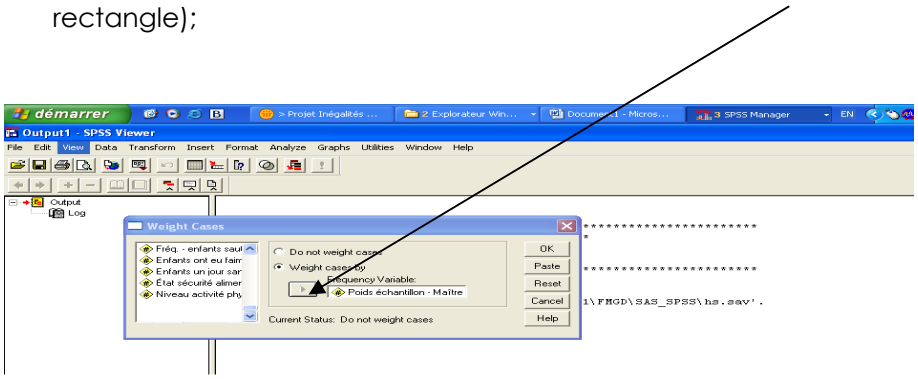
- ensuite, « Open » et « Data… ». Choisir la base de données.

Pour pondérer les données :
Dans le menu, choisir « Data » :

- ensuite, « Weight Cases… »;
- choisir la variable de pondération dans la liste déroulante « Poids commun aux deux questionnaires [**POIDSCOMM**] »;

68. Le choix de l'une ou l'autre des trois variables de pondération relève de l'indicateur et des variables de croisement utilisés pour l'analyse. La variable de pondération POIDSCOMM est utilisée lorsque l'indicateur et la variable de croisement sont tous deux issus de questions apparaissant aux deux versions du questionnaire. La variable de pondération POIDSRESTE_V1 est utilisée lorsque l'indicateur et/ou la variable de croisement sont issus de questions figurant uniquement à la première version, et vice versa pour la variable de pondération POIDSRESTE_V2.

- 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 2

Diagnostic et analyse de la non-réponse partielle pour les données de l'EQSJS 2010-2011

On recommande, lorsque la non-réponse partielle dépasse 5 %[69] pour l'ensemble du Québec et 10 % à l'échelle régionale ou locale, d'effectuer un examen plus approfondi de la variable étudiée.

La non-réponse partielle étant assez faible en général pour les variables de l'EQSJS 2010-2011, il n'est pas nécessaire, sauf exception, de faire un tel examen. Toutefois, pour les cas où cela s'avère nécessaire, la présente annexe explique comment procéder au diagnostic et à l'analyse de la non-réponse partielle sur des estimations de proportions lorsque la non-réponse est non négligeable.

Un exemple tiré de l'ESCC 2000-2001 (cycle 1.1) est présenté[70]. Il illustre, de manière qualitative, l'impact de la non-réponse partielle sur l'estimation de la répartition de l'indice de l'activité physique dans la région sociosanitaire de Chaudière-Appalaches. Bien qu'au niveau régional, les risques de biais importants associés à la non-réponse partielle dans l'EQSJS 2010-2011 soient, sauf exception, négligeables relativement à la précision des estimations, cet exemple est présenté pour illustrer la démarche associée au diagnostic de la non-réponse partielle et à son impact sur les estimations de proportions.

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

Il faut d'abord calculer le taux de non-réponse partielle pour l'indice de l'activité physique pour la région sociosanitaire de Chaudière-Appalaches au cycle 1.1 de l'ESCC. Ce taux est tiré directement du tableau de fréquences simples de la variable avant la répartition de la catégorie « NRP » (fourni dans les tableaux produits à l'Infocentre de santé publique). Ainsi qu'il apparaît au tableau A2.1, il est de 16,66 %. Une analyse plus approfondie s'avère donc nécessaire.

69. 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.

70. Cet exemple est tiré 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* (ISQ, 2009).

Tableau A2.1

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).

Pour étudier l'impact de la non-réponse partielle sur les estimations, on vérifie d'abord si le taux de cette non-réponse est uniforme selon les variables de croisement étudiées dans l'enquête (ex. : âge, sexe). Dans cet exemple-ci, il est pertinent d'examiner si le taux de non-réponse partielle diffère selon le sexe puisque ce dernier est lié au thème étudié, comme en témoigne le résultat significatif du test d'indépendance présenté au tableau A2.2 (p=0,0038). Les taux de non-réponse partielle pour chacun des sexes peuvent être obtenus par une analyse similaire à celle reproduite au tableau A2.1, mais pour les sous-populations des hommes et des femmes considérées séparément. Le tableau A2.3 montre les résultats obtenus pour chacune des catégories de la variable sexe (ces taux sont fournis dans les tableaux produits à l'Infocentre de santé publique). On constate dans ce tableau que le taux de non-réponse partielle à l'indice de l'activité physique est beaucoup plus important dans la sous-population des hommes (20,7 %) que dans celle des femmes (12,6 %). Ces taux pourraient également être déduits du tableau A2.2 par les calculs suivants :

Pour les hommes :

NR = 98,632
NTOT = 98,632 + 377,792 = 476,424
% = 98,632 / 476,424 = **20,7 %**

Pour les femmes :

NR = 59,865
NTOT = 59,865 + 415,044 = 474,909
% = 59,865 / 474,909 = **12,6 %**

Tableau A2.2

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).

Tableau A2.3

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).

2. Biais potentiels et mises en garde

En général, si les taux de non-réponse partielle diffèrent selon les catégories de certaines variables de croisement liées au thème étudié, on doit faire un examen des biais potentiels et les mises en garde qui s'imposent. Cet examen est fait à partir de l'hypothèse que les non-répondants ayant des caractéristiques semblables (définies par les variables de croisement considérées) aux répondants auraient fourni des réponses semblables s'ils avaient participé à l'enquête.

Dans un premier temps, on peut identifier les variables de croisement associées au thème étudié et selon lesquelles les taux de non-réponse partielle diffèrent le plus. On peut ensuite se faire une idée des biais induits par la non-réponse sur les estimations en s'intéressant à l'effet qu'aurait eu la présence des données des non-répondants sur les proportions estimées.

Dans cet exemple-ci, on observe, d'une part, que les hommes ont répondu en moins grande proportion et, d'autre part, qu'ils sont proportionnellement plus nombreux à se situer dans la catégorie « actif » de l'indice de l'activité physique (15,4 %) que les femmes (9,8 %). Pour avoir une idée de l'impact de cette non-réponse sur l'estimation de la répartition de l'indice de l'activité physique, on doit supposer que les non-répondants de sexe masculin présentent une répartition de l'indice similaire à celle des répondants de ce même sexe. La même hypothèse est faite chez les femmes. Cela sous-entend par ailleurs que la répartition des répondants ne diffère pas de celle des non-répondants selon d'autres caractéristiques non prises en compte dans l'analyse. Par conséquent, étant donné les écarts importants observés, il est permis de croire que la non-réponse entraîne une légère sous-estimation de la proportion de personnes dans la catégorie « actif ». Une mise en garde dans l'interprétation des résultats pour indiquer le sens du biais (sous-estimation dans ce cas-ci) serait donc justifiée.

Annexe 3

Exemples de calcul pour l'estimation d'un effectif

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

Tableau A3.1
Effectifs et proportions pondérés de la variable HV7.0 sur les élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

HV7.0 (A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement)	Effectif	Proportion	Effectif cumulatif	Proportion cumulative
NRP	1 478	0,5	1 478	0,5
Oui	112 136	36,9	113 614	37,4
Non	190 472	62,6	304 086	100,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau A3.2
Effectifs et proportions pondérés de la variable HV7.0 sur les élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011 (non-réponse répartie pour la proportion seulement)

HV7.0 (A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement)	Effectif*	Proportion	Effectif cumulatif	Proportion cumulative
NRP	.	.	.	.
Oui	112 136	37,1	112 136	37,1
Non	190 472	62,9	302 607	100,0

* Les valeurs présentées dans cette colonne ne sont pas des estimations sans biais de l'effectif des élèves, 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 2.3.5 (tableau 2.4).

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Détails des calculs pour la catégorie « A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement » de l'exemple présenté à la section 2.3.5

Effectif **biaisé** de la catégorie « A déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement » de la question d'analyse HV7.0	112 136
Divisé par	/
Taux de réponse de l'indicateur HV7.0 pour l'ensemble du Québec (1 – taux de non-réponse partielle)	99,514 % (1 – 0,486 %)
=	=
Estimation non biaisée du nombre d'élèves du secondaire de 14 ans et plus au Québec qui ont déjà eu des relations sexuelles (orales, vaginales ou anales) avec consentement	**112 683**

Conclusion

Le nombre d'élèves du secondaire de 14 ans et plus au Québec qui ont déjà eu des relations sexuelles avec consentement est estimé à 112 700 (voir tableau A3.3[71]).

Tableau A3.3
Proportion des élèves du secondaire de 14 ans et plus ayant eu au moins une relation sexuelle consensuelle au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

Statistiques					
Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
112 700	37,1	1,3	0,50	(36,2 - 38,1)	0,5

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

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

Question filtre (HV6.1)
Au cours des 12 derniers mois, as-tu déjà eu des sifflements dans la poitrine (bruits anormaux ou « râlements » lorsque tu respires), à un moment quelconque?

71. Il est à noter que les effectifs présentés sur le portail de l'Infocentre de santé publique sont arrondis à la centaine près.

Tableau A3.4

Effectifs et proportions pondérés de la question HV6.1 portant sur les sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, pour l'ensemble du Québec, EQSJS 2010-2011

HV6.1 (Au cours des 12 derniers mois, as-tu déjà eu des sifflements dans la poitrine (bruits anormaux ou « râlements » lorsque tu respires), à un moment quelconque?)	Effectif*	Proportion	Effectif* cumulatif	Proportion cumulative
NRP	1 654	.	1 654	.
Oui	78 735	18,3	80 388	18,3
Non	350 712	81,7	431 100	100,0

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Question d'analyse (HV6.4)

Au cours des 12 derniers mois, ces crises de sifflements ont-elles été si graves qu'elles ont limité ton discours à un ou deux mots maximum entre deux respirations?

Tableau A3.5

Effectifs et proportions pondérés de la question HV6.4 portant sur les crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, pour l'ensemble du Québec, EQSJS 2010-2011

HV6.4 (Au cours des 12 derniers mois, ces crises de sifflements ont-elles été si graves qu'elles ont limité ton discours à un ou deux mots maximum entre deux respirations?)	Effectif*	Proportion	Effectif* cumulatif	Proportion cumulative
NRP	218**	.	218	.
Oui	10 345	13,2	10 563	13,2
Non	68 172	86,8	78 735	100,0

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

** Effectifs pondérés des élèves qui ont répondu « NRP » à la question HV6.4 en enlevant les effectifs pondérés des élèves qui ont répondu « NRP » à la question HV6.1.

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Tableau A3.6

Proportion des élèves du secondaire ayant eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, pour l'ensemble du Québec, EQSJS 2010-2011

Statistiques					
Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
10 400	13,2	4,9	0,65	(11,9 - 14,5)	0,7

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Détails des calculs de l'exemple présenté à la section 2.3.6 pour la catégorie « A eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois »

Effectif **biaisé** de la catégorie « A eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations au cours des 12 derniers mois » de la question d'analyse HV6.4	10 345
+	+
Effectif estimé des élèves n'ayant pas répondu à HV6.4 (« NRP » de la question d'analyse seulement) qui auraient eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations au cours des 12 derniers mois	13,2 % de 218
+	+
Effectif estimé des élèves n'ayant pas répondu à HV6.1, mais qui auraient eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations au cours des 12 derniers mois (« NRP » de la question filtre qui aurait dû répondre à la question HV6.4)	13,2 % de (18,3 % de 1 654)
=	=
Estimation non biaisée du nombre d'élèves du secondaire au Québec ayant eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois	**10 414**

Conclusion

Le nombre d'élèves du secondaire au Québec ayant eu des crises de sifflements limitant le discours à un ou deux mots maximum entre deux respirations, parmi ceux ayant eu des sifflements dans la poitrine à un moment quelconque au cours des 12 derniers mois, est estimé à 10 400 (voir tableau A3.6[72]).

72. Il est à noter que les effectifs présentés sur le portail de l'Infocentre de santé publique sont arrondis à la centaine près.

Annexe 4

Information supplémentaire concernant la méthode du « bootstrap »

Dans l'EQSJS 2010-2011, la méthode du « bootstrap » est utilisée à deux fins. Premièrement, elle l'est pour l'estimation de la variance associée à une estimation, pour laquelle 500 ensembles de poids « bootstrap » servent à produire 500 estimations ponctuelles du paramètre d'intérêt. La variance « bootstrap » est ensuite calculée en mesurant la variabilité de ces estimations. Deuxièmement, elle est utilisée pour produire des intervalles de confiance pour le paramètre d'intérêt.

Seule l'estimation de la variance 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'EQSJS 2010-2011.

Voici une brève description de la méthode par ré-échantillonnage du « bootstrap »[73]. Elle consiste à tirer plusieurs sous-échantillons à partir de l'échantillon initial incluant les unités non répondantes, inadmissibles et non valides, s'il y a lieu. Le nombre de sous-échantillons varie d'une enquête à l'autre selon les besoins et objectifs visés. Les sous-échantillons sont sélectionnés de façon indépendante les uns des autres, et sont le résultat d'un échantillonnage aléatoire simple (ÉAS) avec remise de *n-1* unités primaires d'échantillonnage (UPE), parmi les *n* UPE 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 décrits à la section 2.2 sur la pondération est appliqué à chaque sous-échantillon, dont l'ajustement aux effectifs de la population, pour en arriver à la création d'une pondération « bootstrap ». Dans le cas de l'EQSJS 2010-2011, 2 000 ensembles de poids « bootstrap » ont été créés. De ceux-ci, 500 ensembles de poids suffisent pour l'estimation de la précision associée aux estimations.

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

En pratique, voici les principales étapes de l'estimation de la variance associée à une estimation à l'aide de la méthode du « bootstrap » :

a) L'estimation ponctuelle 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. Cinq cents estimations « bootstrap » du paramètre d'intérêt sont ainsi obtenues;
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).

73. La littérature contient un bon nombre de documents décrivant ou évaluant la méthode du « bootstrap » dans un contexte d'utilisation de données provenant d'enquêtes à plan de sondage complexe. Parmi ceux-ci figurent les ouvrages de Lohr (1999), Rust et Rao (1996) ou Wolter (1985).

Annexe 5

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 efficace[74] 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 document[75]. Voici, néanmoins, les conditions d'application de la loi Normale pour une proportion et une taille d'échantillon efficace minimale :

Tableau des conditions d'application de la loi Normale[76]

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

74. La taille efficace d'un échantillon correspond à la taille de cet échantillon divisée 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.
75. Le lecteur qui y est intéressé peut consulter Korn et Graubard (1999).
76. Adaptation du tableau 3.3 de la page 58 de Cochran (1977).

Annexe 6

Niveau de confiance

Les niveaux de confiance le plus couramment utilisés pour bâtir un intervalle de confiance sont les suivants : 99,9 %, 99 % et 95 %. Puisque les intervalles de confiance sont fréquemment retenus comme méthode approximative pour réaliser des tests de comparaison de proportions, le choix du niveau de confiance est étroitement associé au seuil de signification de ces tests. En fait, le niveau de confiance des intervalles produits correspond à 1-a, où a 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, soit sa capacité à détecter une différence entre les proportions, de manière à déterminer les caractéristiques présentant des disparités significatives. En fait, 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 peuvent être déclarées significatives, ce qui facilite d'autant l'interprétation des données pour les utilisateurs et permet de tirer des conclusions fortes.

Le contrôle de l'erreur de type I du test, c'est-à-dire la probabilité de rejeter l'hypothèse d'égalité de deux proportions alors qu'elles sont en fait égales, importe aussi dans la construction du test. Le seuil de signification du test (a) 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, lorsque l'on construit un test statistique, on choisit de contrôler l'erreur de type I en fixant un seuil de signification acceptable, tout en prenant en considération la puissance de ce test. Plus sa puissance est élevée, moins on acceptera de prendre le risque de commettre une erreur de type I. En effet, lorsqu'on abaisse le seuil du test, on réduit le nombre de différences qui pourront être 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, la perte attribuable à la diminution du seuil de signification sera compensée par la puissance élevée d'un test.

Annexe 7

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

Exemple pour le calcul de la proportion ajustée des élèves du secondaire de la région de Montréal ne se percevant pas en bonne santé (passable ou mauvaise)

Âge	Proportion brute	Erreur-type	Population de référence, somme des poids de sondage[1]	Poids d'ajustement	Proportion ajustée	Variance de la proportion ajustée
	(A)	(B)	(C)	(D)	(E)	(F)
12 ans et moins	0,03023	0,00802962	49 369	0,11451867	0,00346190	8,45556E-07
13 ans	0,03744	0,00851875	77 645	0,18010902	0,00674328	2,35409E-06
14 ans	0,04218	0,00748489	88 319	0,20486894	0,00864137	2,35138E-06
15 ans	0,04276	0,00808265	86 651	0,20099977	0,00859475	2,63936E-06
16 ans	0,06180	0,00833385	87 648	0,20331246	0,01256471	2,87091E-06
17 ans et plus	0,06016	0,00896475	41 468	0,09619114	0,00578666	7,43612E-07
Total	0,04591	0,00357662	431 100	1,00000000	**0,04579287**	1,18049E-05

1. La population de référence est celle visée par l'EQSJS 2010-2011.

Source : Institut de la statistique du Québec, *Enquête québécoise sur la santé des jeunes du secondaire 2010-2011*.

Compilation : Institut national de santé publique du Québec, 2012.

Étape 1 : Calcul des proportions brutes (A) et de leur erreur-type (B) pour chacun des âges à partir du fichier de données de l'EQSJS 2010-2011.

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

Exemple : 13 ans, (D) = 77 645 ÷ 431 100 = 0,18010902.

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 : 13 ans, (E) = 0,03744 X 0,18010902= 0,00674328.

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 : 13 ans, (F) = $(0{,}00851875)^2$ X $(0{,}18010902)^2$ =2,35409X10-6.

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 0,003435827.

Annexe 8

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 que l'EQSJS 2010-2011, voici un bref survol de certains concepts de base en inférence statistique.

On appelle *inférence statistique* l'ensemble des méthodes mises au point pour permettre d'obtenir de l'information concernant une population sans avoir à faire de recensement[77]. La procédure de l'inférence statistique se résume 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, on formulera automatiquement l'hypothèse que les dés sont truqués et l'on voudra tester cette hypothèse en prenant soi-même 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. Pour ce faire, une statistique est calculée à partir des données de l'échantillon. La confirmation ou l'infirmation d'une hypothèse est basée sur la probabilité d'obtenir une statistique au moins aussi « extrême » que celle observée, à partir de l'hypothèse posée et selon un seuil fixé, de manière à contrôler la probabilité d'infirmer l'hypothèse par erreur (ex. : seuil de signification de 5 %).

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, au seuil fixé. Cette décision étant basée sur une statistique calculée à partir d'un seul échantillon, il est entendu que différents échantillons tirés d'une même population auraient fourni des statistiques également différentes entre elles. Ainsi, on rejettera l'hypothèse lorsque la probabilité d'obtenir une statistique au moins aussi « extrême » que celle observée à partir de l'échantillon, à partir de l'hypothèse posée, est inférieure au seuil de signification fixé (l'hypothèse posée apparaissant donc peu vraisemblable).

77. Un recensement consiste à recueillir l'information auprès de tous les individus de la population.

Comment interpréter un test statistique?

Lorsqu'un test d'égalité des proportions donne un résultat significatif, on peut affirmer que ces proportions sont statistiquement différentes puisque l'on contrôle la probabilité d'infirmer l'hypothèse posée alors que celle-ci est vraie (seuil de signification). Par contre, lorsqu'un test n'est pas significatif, on ne peut pas nécessairement conclure que les proportions sont égales si on ne s'assure pas d'une forte puissance statistique. En effet, si la puissance statistique est limitée, la probabilité de détecter un écart, si cet écart existe, pourrait être trop faible pour obtenir un résultat significatif. On ne peut pour autant conclure que les proportions comparées sont égales. C'est pourquoi on utilise en général l'une ou l'autre des formulations suivantes pour interpréter un résultat non significatif : « Pour ce qui est de la proportion de la population qui possède telle caractéristique, l'enquête ne permet pas de conclure à une différence significative entre les deux groupes » ou bien « Aucune différence significative n'a été observée entre les deux groupes quant à la proportion de la population qui possède telle caractéristique ».

Exemple

Voici un exemple lié aux données de l'EQSJS 2010-2011. Supposons que l'on désire comparer la proportion des élèves du secondaire de la région Chaudière-Appalaches ayant consommé de la drogue au cours de leur vie avec celle du reste du Québec.

Les données de l'enquête permettent d'estimer que la proportion des élèves du secondaire ayant consommé de la drogue au cours de leur vie en Chaudière-Appalaches s'élève à 29,2 %, alors que celle du reste du Québec est de 27,0 %. À première vue, les résultats ne semblent pas très différents. Si tous les élèves de la région Chaudière-Appalaches étaient répondants à l'enquête, la valeur 29,2 % représenterait non pas une estimation mais la proportion exacte d'élèves du secondaire déclarant avoir consommé de la drogue au cours de leur vie dans cette région. Et s'il en était de même pour le reste du Québec, la proportion exacte d'élèves du secondaire déclarant avoir consommé de la drogue au cours de leur vie pour la même période, pour le reste du Québec, serait de 27,0 %. Dès lors, on pourrait conclure que les proportions sont différentes, car les valeurs ne sont pas identiques.

On ne peut tirer d'emblée cette conclusion puisque l'EQSJS 2010-2011 est une enquête par échantillon; les résultats qui en sont tirés correspondent par conséquent à des estimations des valeurs exactes. Ainsi, l'estimation de la proportion d'élèves du secondaire déclarant avoir consommé de la drogue au cours de leur vie dans la région de Chaudière-Appalaches est associée à une mesure de précision qui reflète l'incertitude liée à l'échantillonnage (voir le tableau A8.1). La précision est souvent représentée par l'intervalle de confiance de l'estimation. Dans le cas de la proportion de la région Chaudière-Appalaches, cet intervalle à 95 % est [27,2 % ; 31,6 %], ce qui veut dire qu'on estime que la valeur exacte de la proportion se situe entre ces deux valeurs[78]. De la même manière, on estime que la valeur exacte de la proportion d'élèves du secondaire ayant consommé de la drogue au cours de leur vie mesurée pour le reste du Québec se situe entre 26,2 % et 28,1 %.

Ainsi qu'il a déjà été mentionné, 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 d'égalité des proportions (voir section 5.1.2) ne permettait pas de conclure que la proportion des élèves du secondaire ayant consommé de la drogue au cours de leur vie en Chaudière-Appalaches est différente de celle du reste du Québec.

78. Plus précisément, un intervalle de confiance à 95 % pour une proportion signifie que si l'on répétait l'échantillonnage un grand nombre de fois selon les mêmes paramètres et que l'on produisait une estimation de la proportion à partir de chacun de ces échantillons, 95 % de ceux-ci généreraient un intervalle de confiance contenant la valeur exacte de la proportion observée dans la population.

Tableau A8.1

Proportion des élèves du secondaire ayant consommé de la drogue au cours de leur vie, EQSJS 2010-2011

Territoire	Statistiques						
Région de l'école	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test
12 Chaudière-Appalaches	6 200	29,2	4,0	1,16	(27,2 - 31,6)	0,0	ns
Reste du Québec	110 800	27,0	1,7	0,47	(26,2 - 28,1)	0,0	

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Annexe 9

Validité du test du khi-deux

Le test du khi-deux n'est valide qu'à certaines conditions. La condition qui s'applique 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 telle fréquence, 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 possible, 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 lignes ou colonnes du tableau avant de faire l'interprétation du test. Les lignes ou colonnes posant problème correspondent généralement à celles présentant de faibles effectifs. Autant que possible, ces dernières devront être regroupées avec d'autres qui sont assez homogènes par rapport à la variable d'étude.

Annexe 10

Tests d'égalité de deux proportions – deux méthodes

Étant donné la nature catégorique[79] de plusieurs variables et indicateurs étudiés, le test d'égalité de deux proportions est l'un des outils importants de l'analyse des données de l'EQSJS 2010-2011. Bien entendu, pour être adéquates, les conclusions doivent être appuyées par un test statistique approprié qui tient compte du plan de sondage de l'enquête. Les analyses doivent notamment intégrer la pondération.

Aux fins de l'exploitation des données de l'EQSJS 2010-2011 sur le portail de l'Infocentre de santé publique, deux méthodes ont été implantées pour effectuer des comparaisons de proportions. Elles reposent sur le concept des poids « bootstrap » et font intervenir la construction d'intervalles de confiance « bootstrap » (voir la section 3.3.3). Cette annexe présente ces deux méthodes ainsi que les avantages et les inconvénients de chacune d'elles.

Méthode 1 : test empirique basé sur l'intervalle de confiance de la différence de deux proportions

Comme il a été 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 son utilisation dans tous les cas, contrairement au test de Student (test *t*) d'égalité de deux proportions.

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 2 000[80] valeurs obtenues, construire l'intervalle de la différence des proportions en prenant les 2,5e et 97,5e centiles de cette distribution comme bornes inférieure et supérieure, pour un intervalle de confiance à 95 %.

Le résultat de ce test s'obtient très simplement : on peut conclure que deux proportions sont différentes au seuil de signification 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, 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.

La méthode 1 présente également l'avantage de tenir compte de la corrélation entre les estimations (ou de la covariance entre les deux estimations de proportions $\hat{p}_1$ et $\hat{p}_2$) dans l'estimation de la variance de la différence. En effet, l'utilisation d'une méthode par ré-échantillonnage, comme celle du « bootstrap », permet de tenir compte adéquatement du plan de sondage.

79. Une variable catégorique (aussi appelée variable qualitative) est une variable pour laquelle chaque réponse peut être classée dans une catégorie particulière.
80. Le nombre d'ensembles de poids « bootstrap » a été fixé à 2 000 dans le cas de l'estimation des centiles 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.

Méthode 2 : construction de deux intervalles de confiance

Il existe une autre avenue pour la comparaison de proportions lorsque les résultats des comparaisons ne sont pas affichés par la requête effectuée sur le portail de l'Infocentre de santé publique (voir la section 6.3 pour les différentes situations menant à l'absence des résultats des comparaisons deux à deux) ou que la méthode 1 ne peut être utilisée. Celle-ci consiste à comparer les intervalles de confiance de chacune des proportions (calculés séparément pour chaque proportion) plutôt que d'utiliser l'intervalle de confiance de la différence des proportions, comme c'est le cas avec la méthode 1. Les intervalles calculés sont encore une fois de type « bootstrap » (section 3.3.3).

La façon d'interpréter un test selon la méthode 2 est la suivante : lorsque les intervalles « bootstrap » des deux proportions comparées ne présentent aucun chevauchement, on peut affirmer que ces 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 tirer de conclusion. Un écart significatif selon la méthode 2 implique nécessairement un écart significatif selon la méthode 1, mais non l'inverse. En effet, le test basé sur la construction de deux intervalles de confiance (méthode 2) est conservateur, ce qui veut dire que la probabilité de rejeter l'hypothèse d'égalité des proportions, lorsque celles-ci sont différentes, est inférieure au seuil de signification fixé à 5 % (Schenker et Gentleman, 2001). Il importe donc de mentionner que la méthode 2 est imparfaite et constitue une option à envisager uniquement lorsque la méthode 1 ne peut être appliquée.

Le tableau A10.1 est un exemple de tableau produit sur le portail de l'Infocentre de santé publique pour la méthode 1 décrite précédemment. On remarque que le résultat est significatif au seuil de 5 % selon cette dernière (résultat du test illustré dans le tableau A10.1). Avec la méthode 2, on ne peut pas conclure qu'il y a une différence significative puisque les deux intervalles de confiance à 95 % se chevauchent. Cet exemple illustre le fait que la méthode 2 est plus conservatrice que la méthode 1.

Tableau A10.1

Proportion des élèves du secondaire dont les crises d'asthme ont été provoquées par la pollution atmosphérique ou le smog, parmi ceux ayant eu des crises d'asthme au cours de leur vie, pour l'ensemble du Québec, EQSJS 2010-2011

Variable de croisement	Statistiques					
Sexe	Nombre[1]	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	6 500	16,9	7,4	1,25	(14,7 - 19,2)	4,8
Féminin	8 500	21,2	5,6	1,19	(18,8 - 23,6)	2,8
Total	15 000	19,1	4,6	0,88	(17,4 - 20,8)	3,8

Test de comparaison de la proportion des élèves du secondaire dont les crises d'asthme ont été provoquées par la pollution atmosphérique ou le smog, parmi ceux ayant eu des crises d'asthme au cours de leur vie, selon la variable « Sexe »
-Proportion brute-

Sexe		Statistiques		
Catégorie 1	Catégorie 2	Proportion 1 (%)	Proportion 2 (%)	Résultat du test
Masculin	Féminin	16,9	21,2	sig

Notes

sig : Différence significative au seuil de 5 %.

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, mai 2012.

Annexe 11

Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre de santé publique

Cette annexe expose une façon différente de présenter l'algorithme d'une requête effectuée sur le portail de l'Infocentre de santé publique de celle apparaissant à la section 6.2.3.

Niveau géographique de l'école : Ensemble du Québec

Sous-population	Sexe	Niveau de scolarité[1]	Première variable de croisement	Deuxième variable de croisement	Test statistique	Type de proportion[4]
Totale			- Aucune			
			- Région de l'école[2]		- Aucun	
					- Indicateur selon var. de croisement	- Brute - Ajustée
			- Sexe - Autres variables[3]	- Aucune	- Aucun	
					- Indicateur selon var. de croisement	- Brute - Ajustée
				- Région de l'école - Groupe d'âge - Autres variables		
			- Groupe d'âge - Niveau de scolarité	- Aucune	- Aucun - Indicateur selon var. de croisement	
				- Région de l'école - Autres variables		
Sexe	- Par sexe - Masculin - Féminin		- Aucune			
			- Région de l'école		- Aucun	
					- Indicateur selon var. de croisement	- Brute - Ajustée
			- Groupe d'âge - Niveau de scolarité	- Aucune	- Aucun - Indicateur selon var. de croisement	
				- Région de l'école - Autres variables		
			- Autres variables	- Aucune	- Aucun	
					- Indicateur selon var. de croisement	- Brute - Ajustée
				- Région de l'école - Groupe d'âge - Autres variables		
Niveau de scolarité		Niveau de scolarité	- Aucune			
			- Région de l'école		- Aucun - Indicateur selon var. de croisement	
			- Sexe - Autres variables	- Aucune	- Aucun - Indicateur selon var. de croisement	
				- Région de l'école - Autres variables		
Sexe et niveau de scolarité	- Par sexe - Masculin - Féminin	Niveau de scolarité	- Aucune			
			- Région de l'école - Autres variables		- Aucun - Indicateur selon var. de croisement	

1. Il s'agit du niveau de scolarité de l'élève qui est disponible seulement chez les 12 ans et plus, les 13 ans et plus et les 14 ans et plus.
2. Il s'agit de la région sociosanitaire de l'école.
3. Il s'agit des autres variables de la fiche, à l'exception du sexe, du niveau de scolarité, du groupe d'âge et de la région sociosanitaire de l'école
4. Dans certaines situations particulières, la standardisation peut être différente pour les indicateurs dont l'âge est ciblé.

Niveau géographique de l'école : Régional

Région sociosanitaire de l'école	Sous-population	Sexe	Niveau de scolarité[1]	Comparaison régionale[2]	Région de comparaison	Variable de croisement	Test statistique	Type de proportion[4]
Choix d'une région de l'école	Totale			Aucune		Aucune		
						Sexe Autres variables[3]	Aucun	
							Indicateur selon var. de croisement	Brute Ajustée
						Groupe d'âge Niveau de scolarité	Aucun Indicateur selon var. de croisement	
				Région de l'école vs reste du Québec				Brute Ajustée
				Région de l'école vs une autre région	Choix région de comparaison			Brute Ajustée
	Sexe	Par sexe Masculin Féminin		Aucune		Aucune		
						Autres variables	Aucun	
							Indicateur selon var. de croisement	Brute Ajustée
						Groupe d'âge Niveau de scolarité	Aucun Indicateur selon var. de croisement	
				Région de l'école vs reste du Québec				Brute Ajustée
				Région de l'école vs une autre région	Choix région de comparaison			Brute Ajustée
	Niveau de scolarité		Niveau de scolarité	Aucune		Aucune		
						Sexe Autres variables	Aucun Indicateur selon var. de croisement	
				Région de l'école vs reste du Québec				
				Région de l'école vs une autre région	Choix région de comparaison			
	Sexe et niveau de scolarité	Par sexe Masculin Féminin	Niveau de scolarité	Aucune				
				Région de l'école vs reste du Québec				
				Région de l'école vs une autre région	Choix région de comparaison			

1. Il s'agit du niveau de scolarité de l'élève qui est disponible seulement chez les 12 ans et plus, les 13 ans et plus et les 14 ans et plus.
2. Il s'agit de la région sociosanitaire de l'école.
3. Il s'agit des autres variables de la fiche, à l'exception du sexe, du niveau de scolarité, du groupe d'âge et de la région sociosanitaire de l'école
4. Dans certaines situations particulières, la standardisation peut être différente pour les indicateurs dont l'âge est ciblé.

Niveau géographique de l'école : Local (RLS) ou Local (Commission scolaire)

Région sociosanitaire de l'école avec suréchantillon[1]	Sous-population	Sexe	Niveau de scolarité[2]	Comparaison locale	Variable de croisement	Test statistique	Type de proportion[4]
Choix d'une région de l'école	Totale			Aucune	Aucune		
					Sexe Autres variables[3]	Aucun	
						Indicateur selon var. de croisement	Brute Ajustée
					Groupe d'âge Niveau de scolarité	Aucun Indicateur selon var. de croisement	
				RLS vs reste de la région (ou commissions scolaires vs reste des commissions scolaires de la région)			Brute Ajustée
	Sexe	Par sexe Masculin Féminin		Aucune	Aucune		
					Autres variables	Aucun	
						Indicateur selon var. de croisement	Brute Ajustée
					Groupe d'âge Niveau de scolarité	Aucun Indicateur selon var. de croisement	
				RLS vs reste de la région (ou commissions scolaires vs reste des commissions scolaires de la région)			Brute Ajustée
	Niveau de scolarité		Niveau de scolarité	Aucune	Aucune		
					Sexe Autres variables	Aucun Indicateur selon var. de croisement	
				RLS vs reste de la région (ou commissions scolaires vs reste des commissions scolaires de la région)			

1. Pour les commissions scolaires, les régions sociosanitaires de l'école avec suréchantillon sont l'Estrie (05) et Chaudière-Appalaches (12). Pour les RLS, les régions sociosanitaires de l'école avec suréchantillon sont la Capitale-Nationale (03), la Mauricie et le Centre-du-Québec (04), l'Estrie (05), l'Abitibi-Témiscamingue (08), la Gaspésie–Îles-de-la-Madeleine (11), Lanaudière (14) et la Montérégie (16).
2. Au niveau local, le niveau de scolarité de l'élève est regroupé par cycle : 1er cycle (1re et 2e secondaire) et 2e cycle (3e, 4e et 5e secondaire). Le niveau de scolarité de l'élève est disponible seulement chez les 12 ans et plus, les 13 ans et plus et les 14 ans et plus.
3. Il s'agit des autres variables de la fiche, à l'exception du sexe, du niveau de scolarité, du groupe d'âge et de la région sociosanitaire de l'école.
4. Dans certaines situations particulières, la standardisation peut être différente pour les indicateurs dont l'âge est ciblé.