

www.stat.gouv.qc.ca

Institut de la statistique du Québec

SANTÉ ET BIEN-ÊTRE

Étude sur l'alimentation
des jeunes québécois :
présentation de l'enquête
et des méthodes

Enquête sur la santé dans les
collectivités canadiennes
(cycle 2.2)

Québec 

Pour tout renseignement concernant l'ISQ et les données statistiques qui y sont disponibles, s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec)
G1R 5T4
Téléphone : 418 691-2401
ou
Téléphone : 1 800 463-4090
(sans frais d'appel au Canada et aux États-Unis)

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

Cette publication a été réalisée et produite par
L'Institut de la statistique du Québec

Dépôt légal
Bibliothèque et Archives Canada
Bibliothèque et Archives nationales du Québec
ISBN 978-2-550-54092-2 (version PDF)

© Gouvernement du Québec, Institut de la statistique du Québec

Toute reproduction est interdite
sans l'autorisation expresse
de l'Institut de la statistique du Québec

Octobre 2008

Avant-propos

Il est aujourd'hui parfaitement reconnu que l'alimentation contribue de manière primordiale à l'état de santé des individus et des populations. En ce qui concerne les jeunes, une alimentation saine est essentielle à la croissance et au développement. Il est clair, également, que les comportements et choix alimentaires des jeunes constituent les bases des habitudes à l'âge adulte et jouent un rôle crucial dans l'état de santé futur de cette population.

Les résultats présentés dans la publication intitulée *L'alimentation des jeunes québécois : un premier tour de table* portent sur l'alimentation des enfants et des adolescents québécois âgés de un à 18 ans. Ils proviennent de l'analyse des données recueillies dans l'*Enquête sur la santé dans les collectivités canadiennes – nutrition* réalisée par Statistique Canada en 2004. L'objectif de l'Institut de la statistique du Québec était de fournir un portrait des enfants et des adolescents, sur le plan alimentaire et nutritionnel, en examinant les résultats obtenus en fonction des suggestions du *Guide alimentaire canadien* et des recommandations en matière de nutrition que sont les *Apports nutritionnels de référence* ou ANREF.

En 1999, l'Institut avait mené une enquête de nutrition auprès d'une population d'enfants et d'adolescents québécois, soit l'*ESSEA – volet nutrition*, enquête bien connue. Ensuite, l'*Enquête de nutrition auprès des enfants québécois de 4 ans* a été réalisée en 2002 dans le cadre de l'*Étude longitudinale du développement des enfants du Québec* (ÉLDEQ 1998-2010). Aujourd'hui, l'analyse et l'interprétation des résultats de l'enquête de Statistique Canada permettent de suivre l'évolution des habitudes alimentaires et de dégager certaines tendances à cet égard. À l'heure où de nombreux pays industrialisés élaborent des politiques visant la promotion de saines habitudes alimentaires, le Québec est particulièrement bien outillé pour surveiller l'état nutritionnel de sa jeune population. Sur la base de ces trois enquêtes, le réseau des décideurs et des intervenants en santé publique, travaillant à la mise en œuvre du *Plan d'action gouvernemental de promotion des saines habitudes de vie et de prévention des problèmes reliés au poids 2006-2012* et du *Programme national de santé publique 2003-2012*, dispose de données probantes. Tant le *plan d'action* que le *programme national* considèrent l'acquisition de saines habitudes alimentaires comme prioritaire.

Une enquête de nutrition peut être comparée à une denrée rare. D'une part, il s'agit d'enquêtes complexes, coûteuses, qui demandent un investissement important de la part des répondants et des intervieweurs. À l'échelle canadienne, la dernière enquête de nutrition remonte à plus de 25 ans, soit en 1972. D'autre part, ces enquêtes exigent énormément des statisticiens et analystes. Le champ des enquêtes populationnelles en nutrition est en évolution de même que les méthodes statistiques qui y sont associées. Ainsi, en plus de répondre au besoin d'une information statistique de qualité sur l'alimentation des jeunes, le mandat confié à l'ISQ par le MSSS témoigne de la volonté de nos décideurs de consolider l'expertise québécoise en matière d'enquêtes de nutrition. Il s'agit, sans nul doute, d'un investissement qui portera ses fruits.

Le directeur général,



Stéphane Mercier

Produire une information statistique pertinente, fiable et objective, comparable, actuelle, intelligible et accessible, c'est là l'engagement « qualité » de l'Institut de la statistique du Québec.

Rédaction :

Nathalie Plante
Maxime Boucher
Robert Courtemanche
Institut de la statistique du Québec

Avec la collaboration de :

Lucie Gingras
Rosanna Baraldi
Institut de la statistique du Québec

Avec l'assistance technique de :

France Vaillancourt

Direction des statistiques sociales, démographiques et de santé :

Daniel Tremblay

Étude subventionnée par :

Ministère de la santé et des services sociaux du Québec

Référence suggérée :

PLANTE, N., M. BOUCHER, R. COURTEMANCHE (2008). *L'alimentation des jeunes québécois : présentation de l'enquête et des méthodes. Enquête sur la santé dans les collectivités canadiennes (cycle 2.2)*, Québec, Institut de la statistique du Québec, 33 p.

Pour tout renseignement concernant le contenu de cette publication :

Direction des statistiques sociales, démographiques et de santé
Institut de la statistique du Québec
1200, avenue McGill College – 5^e étage
Montréal (Québec) H3B 4J8.
Téléphone : 514 873-4749
Télécopieur 514 864-9919
ou
Téléphone : 1 800 463-4090
(aucuns frais d'appel au Canada et aux États-Unis)

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

Table des matières

Avant-propos	3
Table des matières	5
Contexte et objectifs	7
Description de l'enquête ESCC 2.2.....	9
Population visée	9
Plan de sondage.....	9
Collecte des données	10
Résultats de la collecte.....	11
Données du volet alimentaire	13
Apports quotidiens c. apports habituels	13
Estimations produites et outils statistiques.....	15
Estimations produites à partir des premiers rappels seulement	15
Estimations produites à partir des premiers et seconds rappels.....	16
Analyse des données nutritionnelles avec SIDE.....	19
Étape 1 : Exclusion de certains répondants pour l'analyse.....	20
Étape 2 : Ajustements initiaux	20
Étape 3 : Normalisation des données	21
Étape 5 : Estimation des centiles et proportions relatives aux recommandations	21
Étape 6 : Contrôle de qualité et interventions	22
Écarts entre les statistiques produites avec SIDE par l'ISQ et par Statistique Canada.....	25
Portée et limites des données et des statistiques produites	27
Tableaux complémentaires	29
Bibliographie	33

Contexte et objectifs

L'Institut de la statistique du Québec (ISQ) s'est vu confier, par le ministère de la Santé et des Services sociaux du Québec, le mandat d'exploiter les données de *l'Enquête sur la santé dans les collectivités canadiennes (ESCC) – Nutrition* (cycle 2.2) réalisée en 2004 par Statistique Canada. Ce mandat consistait à tracer le portrait de l'état de santé nutritionnelle des enfants et des québécois âgés de 1 à 18 ans. Plus précisément, il s'agissait de produire différentes estimations liées aux habitudes alimentaires, tant au niveau des nutriments que des portions des divers groupes d'aliments consommés. Il était d'intérêt de produire ces estimations par groupe d'âge et de sexe de manière à permettre une comparaison, lorsque possible, aux recommandations d'usage. Par ailleurs, une comparaison aux estimations obtenues pour le regroupement des enfants des autres provinces ainsi qu'aux estimations issues de *l'Enquête sociale et de santé des enfants et des adolescents québécois (ESSEA 1999)* a été réalisée. Seront également étudiés d'autres aspects de l'alimentation tels que les pratiques alimentaires et les déterminants de l'alimentation.

Ce feuillet vise à présenter les aspects méthodologiques liés à l'ESCC 2.2, tant au niveau des procédures utilisées pour réaliser l'enquête qu'au niveau de l'exploitation des données recueillies auprès des enfants québécois. Il s'attarde principalement aux méthodes d'analyse des données nutritionnelles. En ce sens, ce document se veut un complément aux rapports d'analyse qui présentent les résultats de cette exploitation.

La section 2 donne un bref aperçu du contenu de l'enquête. À la section 3 sont décrites les procédures d'enquête liées au plan de sondage et à la collecte des données. La section suivante introduit les concepts de rappels alimentaires de 24 heures, d'apports quotidiens et habituels, ainsi que la problématique de l'analyse de données nutritionnelles. Les types d'estimations produites à partir des données du volet alimentaire sont précisés à la section 5. À la section 6 sont ensuite expliquées les différentes étapes d'analyse des données nutritionnelles pour les statistiques nécessitant plus d'un rappel alimentaire par enfant. Les écarts observés entre les statistiques issues de la présente exploitation de données et celles publiées par Statistique Canada sont quant à eux justifiés à la section 7. Tenant compte des difficultés inhérentes à toute enquête de nutrition, la dernière section permet d'apprécier la portée et les limites des données et des statistiques produites dans cette enquête-ci.

Description de l'enquête ESCC 2.2

L'ESCC est une enquête transversale comportant un volet général sur la santé, de même qu'un volet alimentaire. Le volet général sur la santé porte sur l'état de santé du répondant ainsi que sur ses comportements liés à la santé. Il comporte également des mesures sociodémographiques et anthropométriques. Le volet alimentaire vise à recueillir des renseignements sur tous les aliments et boissons consommés par le répondant sur une période de 24 heures. Ces renseignements incluent une description détaillée des types et quantités d'aliments et boissons consommés, ainsi que de leur lieu de préparation et de leur moment de consommation.

Cette publication porte sur l'analyse des données du volet alimentaire du fichier de partage québécois de l'ESCC 2.2.

Population visée

Le cycle 2.2 de l'ESCC vise les personnes de tous âges, vivant en logement privé, dans les dix provinces canadiennes. Sont exclus du champ de l'enquête les membres à temps plein des Forces canadiennes, de même que les résidents des trois territoires, des réserves indiennes et des terres de la Couronne, des établissements carcéraux, des établissements de soins ou de certaines régions éloignées. Ainsi, la population cible représente près de 98 % de la population totale des dix provinces.

L'exploitation des données nutritionnelles réalisée par l'ISQ porte exclusivement sur la sous-population québécoise âgée de 1 à 18 ans. Les données portant sur les enfants de moins de un an n'ont pas été analysées en raison du trop petit nombre de répondants pour ce groupe d'âge. Pour ce dernier groupe, seules des estimations à l'échelle canadienne ont été produites par Statistique Canada .

Plan de sondage

Le nombre de répondants visé pour l'ESCC 2.2 était de 29 000 personnes pour l'ensemble du pays, dont 4 864 pour le Québec. Ces effectifs ont été établis en fonction du budget alloué et de manière à obtenir des estimations de la distribution des apports nutritionnels et alimentaires habituels pour 15 domaines d'âge et de sexe dans chaque province. Puisque des estimations étaient requises, tant à l'échelle canadienne que provinciale, les tailles d'échantillon provinciales ont été fixées selon la méthode de répartition à la puissance p (Bankier, 1988), après avoir au préalable alloué une taille d'échantillon de 80 personnes par province à chaque domaine. Ces tailles d'échantillon ont été gonflées afin de tenir compte de la non-réponse attendue ainsi que des logements hors du champ de l'enquête.

Plusieurs bases de sondage ont été employées afin d'obtenir un nombre suffisant de répondants de 18 ans ou moins. La base de sondage principale de l'enquête est tirée de la base aréolaire utilisée dans l'*Enquête sur la population active du Canada* (EPA). Cette base ne contenant pas de données sur l'âge des personnes des ménages, une base secondaire créée à partir des renseignements fournis par les répondants à la composante régionale de l'ESCC au cycle 2.1 a également été utilisée pour la sélection de l'échantillon. Cette base liste contenait l'adresse des logements des ménages répondants au cycle 2.1 qui comptaient au moins une personne de 18 ans ou moins.

L'EPA a recours à un plan de sondage stratifié à plusieurs degrés. Selon ce plan, le territoire est découpé en aires géographiques, les unités primaires d'échantillonnage (UPE), lesquelles sont formées d'un regroupement de 150 à 250 logements. Ces derniers sont rassemblés au préalable en strates homogènes. La formation des UPE tient compte d'un premier niveau de stratification selon que les logements sont situés dans un grand centre urbain, une ville ou une région rurale. À un niveau plus fin, les strates regroupent les logements des secteurs de dénombrement¹ semblables quant à leur situation géographique ainsi qu'au revenu moyen ou autres caractéristiques socioéconomiques des ménages les composant. Au premier degré d'échantillonnage, une sélection aléatoire de six UPE est effectuée au sein de chaque strate, de façon indépendante d'une strate à l'autre et avec probabilité proportionnelle au nombre de logements au sein d'une UPE. Les logements de chacune des UPE sélectionnées sont par la suite énumérés en vue d'en faire une sélection aléatoire. Cette base liste de logements a été utilisée pour la sélection d'une première partie de l'échantillon du cycle 2.2 de l'ESCC.

Tout comme la base tirée de l'EPA, la base liste tirée du cycle 2.1 a fait l'objet d'un échantillonnage stratifié à deux degrés, avec une stratification par province et par région urbaine ou rurale. Les municipalités constituent les UPE et ont été sélectionnées avec probabilité proportionnelle à la taille. Un échantillon systématique de logements de ces municipalités a par la suite été sélectionné.

En dernière étape, une seule personne par ménage a été sélectionnée avec une probabilité variant selon l'âge et selon la base de sondage.

Collecte des données

Les UPE ont été réparties aléatoirement en quatre trimestres de collecte, de janvier 2004 à janvier 2005, ceci afin d'équilibrer la charge des intervieweurs et de tenir compte d'éventuels effets saisonniers. Une première entrevue a été réalisée au domicile du répondant afin de recueillir des données pour le volet général sur la santé ainsi que le volet alimentaire. Pour quelque 30 % des répondants, une seconde entrevue a été réalisée par téléphone, de 3 à 10 jours après l'entrevue initiale, afin de répéter la cueillette des données du volet alimentaire pour une journée de référence différente. Cette seconde entrevue visait à estimer la variabilité de la consommation alimentaire d'une journée à l'autre.

Pour le volet alimentaire, les données ont été recueillies à l'aide d'un instrument appelé « rappel alimentaire de 24 heures ». Il s'agit d'un questionnaire d'entrevue assistée par ordinateur, administré en cinq étapes (« Automated Multiple-Pass Method ») pour favoriser le rappel de toutes les quantités d'aliments et boissons consommées par le répondant pendant 24 heures, de minuit à minuit la journée précédant l'entrevue. Le questionnaire passe en revue toute la période visée, tant pour les occasions des repas que des collations, et prévoit une vérification de la consommation d'aliments qui sont fréquemment oubliés.

Pour les enfants de moins de six ans, l'entrevue a été faite auprès d'un parent; pour les repas préparés à la garderie, lorsque nécessaire, on demandait à ce dernier de communiquer avec la personne responsable pour obtenir des détails sur les aliments consommés et leur mode de préparation. Les enfants de 6 à 11 ans ont quant à eux participé à l'entrevue avec l'aide d'un parent.

1. Un secteur de dénombrement (SD) correspond à la région géographique dénombrée par un recenseur. Le nombre de logements dans un SD varie généralement entre un maximum de 650 dans les grands centres urbains et un minimum de 125 dans les régions rurales.

Résultats de la collecte

À l'échelle canadienne, 35 107 personnes de tous âges ont réalisé la première entrevue, pour un taux de réponse non pondéré de 76,5 %. Au Québec, le taux de réponse est de 75,8 %; 4 561 des 4 780 répondants (95,4 %) ont accepté de partager avec l'ISQ les renseignements qu'ils ont fournis et 1 853 d'entre eux ont participé à une seconde entrevue pour le volet alimentaire. La répartition des répondants de 1 à 18 ans par domaine d'âge et de sexe est présentée au tableau I.

Tableau I
Nombre de répondants québécois ayant accepté le partage des données par domaine âge-sexe, 1^{ère} et 2^{ème} entrevues

Domaine	1 ^{ère} entrevue ²	2 ^{ème} entrevue ³
1 à 3 ans	317	133
4 à 8 ans	493	135
Garçons 9 à 13 ans	277	138
Filles 9 à 13 ans	282	139
Garçons 14 à 18 ans	339	146
Filles 14 à 18 ans	324	133
1 à 18 ans	2 032	824

Le lecteur est invité à consulter le guide méthodologique de Statistique Canada (2005) pour plus de détails sur les procédures d'enquête⁴.

2. Répondants pour lesquels des données d'au moins un rappel alimentaire de 24 heures sont disponibles.

3. Répondants pour lesquels des données de deux rappels alimentaires de 24 heures sont disponibles.

4. http://www.statcan.ca/francais/sdds/document/5049_D8_T9_V1_F.pdf

Données du volet alimentaire

Les rappels alimentaires de 24 heures permettent de connaître les aliments et boissons consommés par le répondant au cours d'une journée. Pour l'analyse, les aliments ont été regroupés en grands groupes⁵ et le nombre de portions consommées par le répondant lors de cette journée a été calculé. Sur la base du Fichier canadien sur les éléments nutritifs (FCÉN), qui donne la composition des aliments consommés au Canada, les quantités d'aliments consommés ont également été décomposées en apports nutritionnels pour cette journée de référence⁶. Une liste d'aliments spécifiques à l'ESCC 2.2 et qui n'apparaissent pas dans le FCÉN, de même qu'une base de données de recettes provenant de la base de données du ministère de l'Agriculture des États-Unis (USDA)⁷ ont été utilisées en complément pour faire la décomposition des aliments en nutriments.

Apports quotidiens c. apports habituels

Tant au niveau des portions des grands groupes d'aliments consommés que des apports nutritionnels, l'intérêt principal d'une enquête de nutrition est d'en arriver à connaître les habitudes de consommation d'une population, habitudes auxquelles se rapportent les valeurs de référence établies par les experts pour évaluer la qualité de l'alimentation.

Théoriquement, pour pouvoir décrire adéquatement les habitudes alimentaires d'une population, il serait nécessaire de recueillir un grand nombre de rappels par personne sur une période de temps assez longue. On comprend bien qu'un seul rappel alimentaire, pour une journée donnée, peut ne pas refléter les habitudes de consommation alimentaire pour une durée de 24 heures. Un rappel fournit en effet l'apport alimentaire quotidien pour une journée seulement. Par exemple, si un répondant est interviewé un lundi, il est possible que le dimanche précédant l'interview, la journée de référence, ait été une journée spéciale avec un copieux repas festif. L'apport alimentaire habituel d'une personne fait plutôt référence à la moyenne de ses apports quotidiens considérée sur un grand nombre de jours⁸. Il est par ailleurs reconnu que les apports alimentaires varient de façon importante d'une journée à l'autre, pour une même personne, et que cette variabilité est en général supérieure à celle observée d'une personne à l'autre d'un même groupe d'âge et de sexe.

Il en découle que certaines statistiques d'intérêt seraient biaisées si elles étaient basées sur la distribution des données qui s'appliquent à une seule journée, soit celle des apports quotidiens. Par exemple, la proportion de personnes qui consomment plus que l'apport recommandé de cholestérol pour une journée est supérieure à la proportion de celles qui le font de façon habituelle, c'est-à-dire dont la quantité moyenne consommée est supérieure à la recommandation. Les statistiques telles que les centiles des habitudes alimentaires seraient également biaisées si elles étaient basées sur la distribution des apports quotidiens.

-
5. Seules les données portant sur les grands groupes alimentaires du *Guide alimentaire canadien* de 1992 ont été rendues disponibles pour l'analyse.
 6. Pour certains nutriments, le FCÉN n'indique l'apport nutritionnel que pour une partie des aliments en contenant. C'est le cas, par exemple, de la vitamine E et des acides gras trans présents dans les aliments. Ces nutriments n'ont pas été analysés dans cette étude-ci puisque les quantités consommées auraient été sous-estimées de façon importante.
 7. Des recettes de mets typiquement canadiens ont été ajoutées à la base de données de l'USDA.
 8. Cet apport habituel ne représente pas l'apport alimentaire d'une journée typique pour une personne. Il s'agit plutôt de l'apport alimentaire cumulé sur un grand nombre de jours et ramené par une moyenne à une période de 24 heures.

Pour limiter ces biais et en arriver à tracer le portrait des habitudes alimentaires d'une population, il n'est évidemment pas possible d'avoir recours à un grand nombre de rappels alimentaires par personne, compte tenu des coûts qui seraient associés à un tel projet et au fardeau ainsi imposé aux répondants.

C'est pour cette raison qu'un second rappel alimentaire de 24 heures a été complété pour certains répondants. Les données des premiers et seconds rappels permettent d'estimer la distribution des apports habituels de la population visée et d'en tirer des statistiques sur les habitudes de consommation telles que les centiles de la distribution des apports habituels (nutriments ou groupes alimentaires) et la proportion de la population consommant habituellement trop ou trop peu d'un nutriment ou d'un groupe d'aliment. Ces dernières permettent de situer une population par rapport aux recommandations d'usage, en l'occurrence, le *Guide alimentaire canadien* (GAC) pour les groupes d'aliments et les *Apports nutritionnels de référence* (ANREF) (Santé Canada, 2006) dans le cas des nutriments.

Précisons qu'en l'absence d'apports habituels observés, les données des premiers et seconds rappels sont utilisées pour estimer la distribution des apports habituels à l'aide d'un modèle statistique. Pour ce faire, les données des seconds rappels ne sont requises que pour une partie des répondants. Les statistiques issues de la distribution estimée des apports habituels sont sujettes à des biais moins importants que si elles étaient basées directement sur la distribution des apports quotidiens observés.

Estimations produites et outils statistiques

Différentes statistiques ont été produites à partir des données du volet alimentaire. Une première partie de ces statistiques a été obtenue à partir des données des premiers rappels seulement : apports moyens en macronutriments et en micronutriments, nombre moyen de portions pour chacun des grands groupes du GAC, contribution des macronutriments aux apports en énergie et contribution des apports des groupes alimentaires aux apports en énergie et en nutriments. Les mêmes statistiques ont également été produites pour la population des enfants de l'ensemble des autres provinces, par groupes d'âge et de sexe, aux fins de comparaison. Les apports moyens estimés pour un autre découpage selon l'âge et le sexe (garçons et filles de 6 à 8 ans, garçons et filles de 9 à 13 ans, garçons et filles de 14 à 16 ans) ont finalement été utilisés dans la comparaison avec les résultats de l'ESSEA de 1999.

Une seconde partie des statistiques produites est basée sur les données des premiers et seconds rappels : centiles de la distribution des apports habituels en macronutriments et en micronutriments, centiles de la distribution des portions habituelles des grands groupes alimentaires, proportions d'apports habituels en macronutriments et en micronutriments relatives aux ANREF et proportions des quantités de portions habituelles relatives aux recommandations du GAC. Les statistiques portant sur les apports en macronutriments et en micronutriments ont été comparées à celles portant sur la population des enfants de l'ensemble des autres provinces, lesquelles ont été produites suivant la même démarche.

Estimations produites à partir des premiers rappels seulement

Bien que la distribution des apports quotidiens présente en général une variabilité beaucoup plus grande que celle des apports habituels, ces distributions partagent une même moyenne. Il en est ainsi puisqu'un apport habituel représente, par définition, une moyenne d'apports quotidiens considérée sur un grand nombre de jours. Ainsi, les estimations de moyennes ont été faites sur la base des données des premiers rappels seulement plutôt que sur la distribution des données ajustées à partir des premiers et seconds rappels puisque ces dernières comptent, en plus d'une erreur due à l'échantillonnage, une erreur de modélisation. C'est dire que les estimations ont été tirées des données déclarées par le répondant (apports quotidiens observés), plutôt que sur celles obtenues par modélisation (apports habituels estimés).

L'estimation de rapports de moyennes a également été faite à partir des premiers rappels uniquement⁹. C'est le cas, par exemple, de l'estimation de la contribution du groupe des produits céréaliers à l'apport en énergie qui s'obtient en divisant la moyenne estimée des apports en énergie par enfant provenant de ce groupe alimentaire par la moyenne estimée des apports en énergie totale par enfant. Tant les moyennes que les rapports de moyennes peuvent être interprétés en termes de consommation habituelle.

Notons qu'avec les outils d'analyse utilisés, il y a dans certains cas impossibilité d'estimer adéquatement la distribution des apports habituels. Cela se produit, par exemple, pour les nutriments ou les aliments qui ne sont pas consommés à peu près tous les jours par la plupart des enfants. On ne peut alors produire des statistiques telles que les centiles de la distribution des apports habituels ou les proportions d'apports habituels relatives aux recommandations d'usage. Dans cette situation, il demeure toutefois pertinent de présenter des statistiques telles

9. Un rapport de moyennes s'obtient en divisant l'apport moyen par personne pour une première variable par l'apport moyen pour une seconde variable.

que les moyennes ou les rapports de moyennes décrivant la consommation habituelle. Bien que ne fournissant pas d'information permettant de situer une population par rapport aux recommandations, ces statistiques permettent de connaître son niveau de consommation habituelle et de le comparer à celui d'autres populations.

Les estimations de moyennes et de rapports de moyennes ont été produites à l'aide du logiciel SUDAAN, sur la base des données issues du fichier de partage. Toutes les estimations ont été pondérées pour tenir compte du plan de sondage. Les 500 pondérations « bootstrap » fournies par Statistique Canada avec ce fichier ont par ailleurs été utilisées pour tenir compte du plan de sondage dans les estimations de précision. Un coefficient de variation (CV) ainsi qu'un intervalle de confiance à 95 % accompagnent les estimations produites. L'intervalle de confiance est construit à partir d'une estimation de l'erreur-type, sous l'hypothèse d'une distribution normale.

Le CV est une mesure de précision relative qui s'obtient par le rapport de l'erreur-type de l'estimation sur l'estimation elle-même. Plus le CV est élevé, moins l'estimation est précise et inversement. L'intervalle de confiance à 95 %, quant à lui, complète l'information fournie par le CV, en ce sens qu'il permet d'apprécier la portée des estimations dont le CV est élevé. Pour un niveau de confiance de 95 %, l'intervalle de confiance signifie que si l'on tirait un échantillon de la population visée un grand nombre de fois selon le même plan de sondage, alors 95 % des échantillons sélectionnés fourniraient un intervalle de confiance contenant la vraie valeur du paramètre concernant la population visée.

À chaque CV est par ailleurs accolée une cote permettant d'identifier rapidement les estimations dont la précision est faible. En ce sens, un CV ne dépassant pas 15 % ne requiert aucune annotation. Une estimation dont le CV est supérieur à 15 % est accompagnée d'une note indiquant si le CV se situe entre 15 % et 25 % inclusivement, ou encore au-delà de 25 %; cette note appelle également une interprétation tenant compte de l'intervalle de confiance. Dans les deux cas, les estimations sont publiées seulement lorsque cet intervalle fournit une information valable. En effet, en dépit d'une faible précision relative, certaines statistiques demeurent informatives lorsqu'elles sont interprétées avec l'intervalle de confiance correspondant.

Estimations produites à partir des premiers et seconds rappels

Les données des premiers et seconds rappels issues du fichier de partage ont été utilisées pour produire des estimations de centiles et de proportions relatives aux ANREF ou au GAC, selon le cas. L'estimation de la distribution des apports habituels ainsi que de la distribution des portions habituelles des grands groupes d'aliments du GAC a été faite à l'aide du logiciel *Software for Intake Distribution Estimation* (SIDE - version SIDE-IML) développé par l'Université de l'Iowa, tel que recommandé par Statistique Canada. Ce logiciel est également utilisé pour l'analyse des données des grandes enquêtes nutritionnelles américaines. Toutes les estimations ont été produites séparément par domaine d'âge et de sexe.

Les analyses réalisées avec SIDE sur les habitudes de consommation alimentaires portent uniquement sur les nutriments consommés presque quotidiennement et les aliments regroupés selon les cinq grands groupes du GAC. Avec la version SIDE-IML et sur la base de deux rappels de 24 heures uniquement, il n'est en effet pas possible d'estimer la distribution habituelle des nutriments ou des aliments qui ne sont pas consommés tous les jours ou presque par la plupart des répondants. C'est le cas, en particulier, pour l'alcool et la caféine consommés par les personnes de 18 ans et moins, de même que pour les aliments considérés un à un. Les aliments regroupés selon les sous-groupes d'aliments (les 54 sous-groupes) du GAC n'ont pas non plus été analysés parce que les renseignements permettant ces regroupements n'ont pas été diffusés par Statistique Canada.

Notons que les statistiques sur les apports alimentaires ont été comparées à la version du GAC en vigueur au moment de l'enquête en 2004. Bien que le GAC ait fait l'objet d'une révision en 2007, il n'était pas possible, au moment de l'analyse statistique, d'obtenir les outils permettant l'assignation des aliments aux divers groupes alimentaires de la version 2007, ni ceux requis pour l'assignation de la taille des portions aux aliments ayant fait l'objet de modifications dans la nouvelle version. Des travaux en ce sens sont en cours à Santé Canada et permettront éventuellement de procéder à de telles comparaisons.

Les 500 pondérations « *bootstrap* » fournies par Statistique Canada avec le fichier de partage ont été utilisées pour tenir compte du plan de sondage dans les estimations de précision avec SIDE. Pour ce faire, des programmes ont été développés afin d'utiliser le programme SIDE de façon répétitive pour chacune des répliques « *bootstrap* », pour ensuite rassembler les résultats obtenus et en tirer une estimation de la précision. Tout comme les estimations produites à partir des premiers rappels seulement, les estimations sont dans ce cas-ci accompagnées d'un CV ainsi que d'un intervalle de confiance à 95 % lorsque pertinent. Ce dernier est déduit de la distribution observée pour les 500 répliques « *bootstrap* ».

Analyse des données nutritionnelles avec SIDE

L'analyse des habitudes nutritionnelles avec SIDE comporte plusieurs étapes. Le but visé est d'estimer la distribution d'apports habituels à partir des données observées d'apports quotidiens, puis de tirer de cette distribution différentes statistiques décrivant l'état de santé nutritionnelle des enfants québécois de six groupes d'âge et de sexe. Le processus de modélisation est décrit en détail dans les articles de Nusser et al. (1996;1997). Le guide de l'utilisateur de SIDE (1996) ainsi que la documentation afférente produite par Statistique Canada (2006) peuvent être consultés pour obtenir plus de renseignements concernant l'utilisation de ce logiciel.

L'estimation de la distribution d'apports habituels passe par l'ajustement d'un modèle d'erreur de mesure visant à estimer les composantes de variance intra-individuelle et interindividuelle, sous l'hypothèse d'une distribution normale des apports quotidiens. Selon ce modèle, l'apport quotidien observé pour une personne est égal à la somme de l'apport habituel de cette personne et d'une erreur de mesure. Les estimations produites ainsi que leur précision estimée tiennent compte du plan de sondage.

Antérieurement, l'analyse des données de l'Enquête québécoise sur la nutrition (1990), du volet nutrition de l'Enquête sociale et de santé auprès des enfants et des adolescents québécois (1999) et du volet nutrition de l'Étude longitudinale sur le développement des enfants du Québec (2002) avait été faite à partir de programmes produits par Santé Canada. Tout comme SIDE, ces programmes visaient à estimer une distribution d'apports habituels à partir de données d'apports quotidiens, sur la base de données de rappels de 24 heures. Les deux méthodes ont recours à un modèle d'erreur de mesure pour estimer les composantes de variance intra-individuelle et interindividuelle des apports quotidiens déclarés.

Par rapport à la méthode utilisée pour les enquêtes antérieures, SIDE permet de faire certains ajustements initiaux aux données d'apports quotidiens pour réduire la variabilité associée au déroulement de la collecte de données. De plus, bien que les anciennes méthodes d'analyse intégraient aussi une normalisation des données, SIDE a recours à une méthode semi-paramétrique plutôt qu'à une simple transformation par la puissance telle que la transformation par la racine carrée ou logarithmique, par exemple. En effet, cette dernière ne permettait pas toujours d'obtenir une distribution normale, hypothèse sous-jacente à l'estimation des composantes de variance basée sur le modèle d'erreur de mesure. Finalement, en répétant l'analyse avec SIDE pour chacune des pondérations « *bootstrap* », il est possible de tenir compte à la fois de la variance due à l'échantillonnage et de la variance liée à la modélisation dans l'estimation de la variance des statistiques produites. Dans la procédure d'estimation des enquêtes antérieures basée sur les programmes de Santé Canada, cette deuxième composante de variance était considérée négligeable, ce qui pouvait mener à une sous-estimation. Dans la présente exploitation de données, le processus complet de modélisation a été répété pour chaque réplique « *bootstrap* », ce qui fait en sorte que la variabilité due à la modélisation a pu être prise en compte.

Étape 1 : Exclusion de certains répondants pour l'analyse

À partir du fichier de partage, ont été exclues de l'analyse quelques adolescentes enceintes ou allaitantes. Ces dernières sont par conséquent exclues de la population visée pour l'analyse des apports nutritionnels et alimentaires. Les rappels alimentaires étiquetés comme étant non valides dans le fichier¹⁰ ont également été exclus, de même qu'un rappel pour lequel le fichier ne contenait aucun aliment déclaré¹¹.

Étape 2 : Ajustements initiaux

La cueillette de données du second rappel alimentaire de 24 heures est semblable à celle du premier rappel, si ce n'est que le mode de collecte diffère et que la participation du répondant au second rappel peut être influencée par sa participation au premier rappel. En effet, le répondant enquêté pour une seconde fois a une idée de ce qui lui sera demandé. Or, il est raisonnable de croire que les quantités réellement consommées devraient être égales, en moyenne, pour les premiers et seconds rappels; il appert toutefois que les quantités moyennes déclarées sont en général inférieures lors du second rappel. Par ailleurs, une part de la variabilité observée dans les quantités déclarées provient du fait que les quantités consommées par une personne varient d'un jour de la semaine à l'autre. Relativement à l'étude des habitudes alimentaires, il est souhaitable de contrôler cette source de variabilité.

Dans cette enquête-ci, des ajustements initiaux ont par conséquent été faits afin de ramener, pour chaque domaine étudié, la moyenne des seconds rappels et celle de chacun des jours de la semaine (premiers et seconds rappels regroupés) à la moyenne globale des premiers rappels. Ces ajustements ont été effectués sur les données préalablement normalisées par une transformation par la puissance. Une fois tous les ajustements initiaux effectués, les données ont été retransformées dans l'échelle originale par une transformation inverse.

Dans le programme SIDE, l'ordre du rappel (premier ou second) ainsi que le jour de la semaine ont été spécifiés explicitement à l'énoncé CLASSVAR, de sorte que les ajustements initiaux n'ont pas modifié la moyenne globale des premiers rappels normalisés pour chaque domaine. Ces considérations sont importantes puisque la modélisation est faite sous l'hypothèse que les quantités moyennes déclarées au premier rappel sont fiables. Une modification de ces quantités, à tort, donnerait lieu à une sous-estimation ou une surestimation des quantités déclarées et, par le fait même, des statistiques telles que les centiles ou les proportions relatives aux recommandations¹².

Notons que les rappels alimentaires de 24 heures ne sont pas répartis de façon homogène selon le jour de la semaine. Plus précisément, pour l'ensemble des répondants québécois de 1 à 18 ans, on ne compte que 26 rappels se rapportant aux aliments consommés un samedi (soit de 1 à 7 rappels par domaine âge-sexe, premiers et seconds rappels confondus). Par conséquent, les rappels se rapportant au vendredi et samedi ont été regroupés avant l'étape des ajustements initiaux. Il semble toutefois que l'impact de ce regroupement sur les estimations demeure négligeable. Il y a également peu de seconds rappels se rapportant au vendredi, mais puisqu'il n'y a pas d'interaction entre l'ordre des rappels et le jour de la semaine dans le modèle sous-jacent aux ajustements initiaux, cela ne semble pas problématique et aucune action supplémentaire n'a été faite pour pallier ce problème.

10. Dix enfants pour lequel R24DVAL=2.

11. Un enfant pour lequel R24DFLOW=1. Selon Statistique Canada, les données pour cet enfant sont manquantes.

12. Il y a divergence entre les façons de faire de l'ISQ et de Statistique Canada quant aux ajustements initiaux. Pour l'analyse des apports nutritionnels, Statistique Canada n'a pas spécifié explicitement l'ordre des rappels à l'énoncé CLASSVAR.

*Étape 3 : Normalisation des données*¹³

L'ajustement du modèle d'erreur de mesure utilisé pour l'estimation des composantes de variance présuppose une distribution normale des données de consommation quotidienne ainsi que des erreurs de mesure. Or, une distribution normale ne peut être obtenue pour certains nutriments à l'aide d'une simple transformation par la puissance. SIDE propose alors d'avoir recours à une transformation semi-paramétrique normalisatrice. Un des paramètres définissant la méthode de normalisation dans les queues de distribution peut être modifié dans SIDE. La valeur de départ choisie pour ce paramètre (option LINFRAC = 0,01) a été augmentée lorsque l'ajustement à la loi normale n'était pas satisfaisant (valeurs retenues de 0,01 à 0,04)¹⁴.

Dans l'analyse des portions des grands groupes alimentaires, étant donné la présence d'un certain nombre de portions nulles (entre 1 % et 3 % des rappels), la modification de l'option LINFRAC n'a pas permis dans la majorité des cas d'obtenir une distribution normale. Ainsi, la difficulté à normaliser les portions d'aliments est généralisée, à l'exception des produits céréaliers qui présentent moins de valeurs nulles. Statistique Canada recommande de ne pas utiliser SIDE pour estimer des quantités pour les apports comportant plus de 20 % de valeurs nulles. Cette proportion varie de 1 % à 3 % pour les groupes d'âge et de sexe étudiés pour les portions alimentaires, ce qui est assez faible. Il est toutefois difficile d'évaluer l'impact de la difficulté à normaliser sur les résultats.

Étape 4 : Estimation de la distribution des apports habituels

Les composantes de variance intra-individuelle et interindividuelle sont estimées à l'aide d'un modèle d'erreur de mesure ajusté aux apports quotidiens normalisés (après ajustements initiaux). Une fois la distribution des apports habituels normalisés estimée, une transformation est appliquée afin d'obtenir une distribution estimée des apports habituels dans l'échelle originale. C'est de cette dernière distribution que sont tirées les statistiques portant sur les habitudes de consommation.

Étape 5 : Estimation des centiles et proportions relatives aux recommandations

Les statistiques telles que les centiles et les proportions relatives aux ANREF et aux portions recommandées du GAC sont obtenues à partir de la distribution estimée des apports habituels. Les estimations de précision sont faites à partir des 500 poids « *bootstrap* » fournis avec le fichier de partage par Statistique Canada. Ces poids sont utilisés un à un pour reproduire le processus d'ajustement complet avec SIDE et en tirer une estimation de variance.

13. Contrairement à l'étape 2, où la normalisation ne constitue qu'une étape temporaire préalable aux ajustements initiaux, la méthode de normalisation des données exposée à l'étape 3 est fondamentale et sous-jacente au modèle proposé par Nusser et al. (1996).

14. Un second paramètre définissant la méthode de normalisation peut être augmenté dans SIDE dans le but d'obtenir un meilleur ajustement à la loi normale lorsque nécessaire (option MAXJP). Pour certaines combinaisons nutriment-domaine, ce paramètre a été modifié. Dans tous les cas, cela n'a eu aucun impact sur l'estimation ponctuelle. Seule l'estimation de variance a diminué, de telle sorte que l'estimation initiale de variance, plus conservatrice, a été retenue.

Étape 6 : Contrôle de qualité et interventions

Les estimations obtenues à l'aide de SIDE doivent faire l'objet d'une validation. En effet, pour certaines combinaisons nutriment-domaine, il peut arriver que le nombre de répondants (en particulier le nombre de seconds rappels) soit insuffisant pour permettre un ajustement adéquat des données d'apports quotidiens en données d'apports habituels¹⁵. Dans cette situation, il arrive que SIDE fournisse tout de même des estimations de centiles et de proportions relatives aux recommandations. Il serait par conséquent hasardeux d'interpréter ces dernières sans procéder au préalable à certaines vérifications quant à leur validité.

Une première situation problématique se présente lorsqu'un ajustement adéquat ne peut être réalisé sur la base de l'ensemble de l'échantillon des répondants. Une variance intra-individuelle estimée supérieure à 90 % est en ce sens suspecte. Une comparaison avec cette même composante de variance au niveau canadien (ESCC 2.2) ainsi qu'avec les données de l'enquête américaine NHANES peut dans ce cas permettre de déterminer si les estimations préliminaires doivent être invalidées.

Le cas échéant, une solution possible, suggérée par Statistique Canada, consiste à imputer les estimations de variance intra-individuelle canadiennes, pour le même sous-groupe d'âge et de sexe, en remplacement des variances québécoises, sous l'hypothèse d'homogénéité de la variance intra-individuelle au Canada pour ce sous-groupe. Ces composantes de variance sont alors utilisées tant au niveau des estimations ponctuelles (centiles ou proportions relatives aux ANREF) que des estimations de précision. Cette solution a été appliquée au besoin pour l'estimation des statistiques portant sur les apports nutritionnels¹⁶.

Il appert toutefois que les estimations obtenues selon que l'on utilise ou non les paramètres canadiens diffèrent de façon majeure, dans plusieurs cas, surtout pour les centiles des queues de distribution. Non seulement les estimations ponctuelles diffèrent-elles, mais l'erreur-type ainsi obtenue est beaucoup plus petite pour ces centiles. C'est également le cas pour les estimations de proportions relatives aux ANREF. Le fait de décider ou non d'imputer a donc un impact majeur sur la statistique produite et lorsqu'il y a imputation, la précision apparaît bien meilleure. De plus, les critères selon lesquels on prend la décision d'intervenir ou non sont quelque peu arbitraires.

En effet, les estimations produites en imputant les paramètres canadiens le sont sous l'hypothèse d'homogénéité de la variance intra-individuelle d'une province à l'autre, pour un groupe d'âge et de sexe donné. Une erreur-type plus petite est obtenue puisque la précision estimée est liée à la variabilité de la variance intra-individuelle d'une réplique à l'autre et que la taille d'échantillon du Canada est beaucoup plus grande que celle du Québec. Advenant que l'hypothèse d'homogénéité ne tienne pas tout à fait la route, on aura donc une fausse impression de précision. Il y a en effet une possibilité non négligeable de biais, mais l'estimation de précision n'en tient pas compte. En fait, les cas où on a le plus de difficulté à obtenir des estimations valables pour le Québec seront ceux où la précision estimée paraîtra finalement la meilleure lorsque les paramètres canadiens sont utilisés.

Pour toutes ces raisons, il est recommandé d'annoter ces estimations lorsqu'elles sont publiées pour indiquer sous quelle hypothèse elles ont été produites et les risques de biais encourus. Pour l'analyse des apports nutritionnels, les variances estimées au niveau canadien n'ont été utilisées que pour 14 combinaisons nutriment-domaine. Pour ces dernières, les estimations initiales (basées sur la variance intra-individuelle québécoise) ont également été produites et ont été retenues pour publication lorsque les conclusions tirées étaient appuyées par les estimations basées sur la variance intra-individuelle canadienne.

15. Rappelons que dans l'ESCC 2.2, pour les enfants âgés de 1 à 18 ans, on compte environ 300 répondants par domaine, dont quelque 135 répondants au second rappel (tableau I).

16. Simultanément à l'imputation des paramètres de variance intra-individuelle, le quatrième moment centré de l'erreur de mesure est imputé.

Une autre solution envisageable pour pallier le problème d'ajustement consiste à imputer les paramètres québécois pour un groupe d'âge et de sexe élargi. L'ordre de grandeur de la précision estimée est ainsi plus comparable à celui obtenu pour le domaine initial ainsi que pour les autres domaines et, par conséquent, le risque de se tromper est vraisemblablement plus faible. Selon un examen préliminaire des données, les estimations produites avec les variances intra-individuelles du domaine élargi se situent entre les deux autres estimations (Québec avec domaine non élargi et ensemble du Canada) et ont une précision plus comparable à celle du Québec, ce qui reflète mieux la difficulté d'estimation pour ce domaine. Cette solution, qui semble plus avantageuse que la première, a été appliquée lorsque requis aux données d'apports alimentaires (viandes et substituts chez les enfants de 4 à 8 ans seulement) mais n'a pu être appliquée aux données d'apports nutritionnels dans les délais et coûts impartis pour la présente exploitation des données.

Dans d'autres situations d'analyse, c'est l'estimation de précision à l'aide des 500 poids « *bootstrap* » qui cause problème. En effet, chacun des poids « *bootstrap* » prend une valeur nulle pour une proportion importante de l'échantillon des répondants, de sorte que pour chaque réplique « *bootstrap* », SIDE répète l'ajustement des données sur la base d'un échantillon de taille réduite, d'où la difficulté accrue d'obtenir des estimations adéquates pour certaines répliques¹⁷.

Dans cette situation, il arrive que des répliques ne donnent lieu à aucune estimation, indiquant que l'ajustement n'a pu être réalisé. Par exemple, SIDE émet un avertissement en présence d'une variance interindividuelle estimée négative pour une réplique (code 65). Si le nombre de répliques rejetées n'est pas négligeable, il est possible que les estimations de précision soient biaisées. Cela peut également suggérer une instabilité dans le processus d'ajustement, même pour les estimations ponctuelles initiales correspondantes. Ces dernières peuvent être remises en cause surtout s'il y a un problème de convergence dans l'estimation de la variance. Autrement dit, lorsque l'on estime la variance à l'aide d'un grand nombre de répliques « *bootstrap* », on s'attend à ce que l'estimation de variance se stabilise à mesure que l'on ajoute des répliques, c'est-à-dire qu'elle converge vers une valeur. En l'absence d'une convergence acceptable, la solution adoptée est encore d'avoir recours aux estimations de variance intra-individuelles canadiennes. Les estimations ainsi obtenues ne sont conservées que si l'impact de cette imputation est important, question de minimiser les interventions au processus d'ajustement.

Il arrive également que des estimations soient obtenues pour une réplique, sans pour autant que le processus d'ajustement ait bien fonctionné. Le but de la démarche de validation consiste alors à déterminer si certaines répliques ont fourni des estimations de centiles aberrantes, c'est-à-dire qui sont très éloignées de celles obtenues avec les autres répliques valides. Dans ce dernier cas, un examen graphique de la convergence des estimations de variance permet d'identifier les répliques qui posent un problème important, en ce sens qu'elles modifient à elles seules les estimations de variance de façon majeure (saut important dans la courbe de convergence). Certaines répliques problématiques peuvent également être identifiées par une somme des variances intra-individuelle et interindividuelle sensiblement supérieure à 1. En effet, le processus d'ajustement inclut une standardisation telle que la somme des composantes de variance soit égale à un, mais cela ne fonctionne pas pour certaines répliques. Toutes les répliques ainsi identifiées ont été éliminées pour l'estimation de la variance. Ce n'est que lorsque l'écart entre l'erreur-type ainsi obtenue et l'erreur-type obtenue initialement était supérieur à 5 % pour au moins un des centiles estimés que les nouveaux résultats ont été conservés.

Finalement, la validation des estimations passe par une vérification du fonctionnement de l'étape de normalisation des données réalisée préalablement à l'estimation des composantes de variance. Lorsque la transformation normalisatrice n'a pu être effectuée correctement pour une réplique, SIDE émet un avertissement (code 61). Lorsque plusieurs répliques ont donné lieu à un tel avertissement, le paramètre de SIDE impliqué dans la transformation semi-paramétrique (option LINFRAC) a été modifié. Plusieurs valeurs différentes ont en général été évaluées afin de minimiser le nombre de répliques problématiques.

17. Puisque chaque réplique « *bootstrap* » est le fruit d'un échantillonnage avec remise d'UPE au sein de l'échantillon total, certaines UPE ne sont pas sélectionnées pour une réplique donnée; ces dernières se voient par conséquent allouer un poids nul.

Les options modifiées dans l'analyse des apports nutritionnels et alimentaires avec SIDE ainsi que les interventions effectuées quant au rejet de répliques ou à l'imputation des variances estimées pour un domaine élargi sont documentées aux annexes 1 (apports nutritionnels) et 2 (apports alimentaires) pour chacune des combinaisons de variables et domaines d'âge et de sexe où cela s'applique.

En résumé, il ressort que les effectifs limités par domaine d'analyse appellent une grande précaution dans l'analyse des apports nutritionnels et alimentaires avec SIDE, tant au niveau des estimations ponctuelles (p. ex. : proportion, moyenne) que des estimations de précision. Dans la présente exploitation de données, différentes stratégies ont été appliquées afin de solutionner les problèmes liés à l'estimation, optimisant ainsi la qualité des statistiques produites. Les écarts entre les quantités moyennes déclarées aux premiers et seconds rappels ont été pris en compte dans l'analyse par un ajustement initial des données des seconds rappels. Les difficultés d'ajustement du modèle d'erreur de mesure ont été résolues par l'utilisation de paramètres provenant d'un domaine élargi (Canada ou groupe d'âge élargi pour le Québec). Cette même solution ainsi que le rejet de répliques « *bootstrap* » invalides ont été appliqués pour pallier les problèmes d'estimation de la précision. Finalement, les paramètres de SIDE ont été modifiés pour faire face aux difficultés de normalisation des données.

Écarts entre les statistiques produites avec SIDE par l'ISQ et par Statistique Canada

Les statistiques produites à l'ISQ peuvent différer de celles publiées par Statistique Canada pour les enfants québécois de 1 à 18 ans pour plusieurs raisons. D'abord, il y a divergence entre la méthode utilisée par les deux organismes pour les ajustements initiaux. Pour l'analyse des apports nutritionnels et alimentaires avec SIDE, Statistique Canada n'a pas spécifié explicitement l'ordre des rappels à l'énoncé CLASSVAR, comme l'a fait l'ISQ. Cela fait en sorte que les estimations ponctuelles produites par Statistique Canada sont en général supérieures à celles produites par l'ISQ, pour un même sous-groupe de répondants. Par exemple, pour la variable d'énergie totale, les estimations de centiles¹⁸ sont de l'ordre de 2 % à 8 % plus élevées selon le domaine étudié (écart stable d'un centile à l'autre pour un même domaine). L'écart observé dans les résultats est d'autant plus grand que les moyennes initiales des premiers et seconds rappels diffèrent. Les variances sont également plus élevées en général lorsque l'on omet l'ordre des rappels.

Dans une moindre mesure, les interventions effectuées dans l'analyse avec SIDE telles que le choix des paramètres de SIDE pour la transformation normalisatrice, le rejet des répliques « *bootstrap* » non valides, l'identification des cas où l'on impute la variance intra-individuelle d'un domaine élargi et la méthode utilisée pour ce faire, ont toutes pu avoir un impact plus ou moins important tant sur les estimations ponctuelles que sur les estimations de précision. Les combinaisons de nutriments (ou groupes d'aliment) et de domaines d'âge et de sexe où ces interventions ont été requises et retenues pour publication sont toutefois limitées en nombre.

Finalement, dans la présente exploitation de données, l'estimation de la variance des estimations de centiles et de proportions à l'aide des poids « *bootstrap* » a été faite en répétant le processus complet de modélisation pour chacune des 500 répliques. Statistique Canada a plutôt eu recours à une même estimation de variance intra-individuelle pour toutes les répliques, ce qui entraîne une légère sous-estimation de la variance des estimations.

18. Les 5ème, 10ème, 25ème, 50ème, 75ème, 90ème et 95ème centiles ont été étudiés.

Portée et limites des données et des statistiques produites

Le questionnaire d'entrevue automatisé administré en cinq étapes aide le répondant à se remémorer tous les aliments consommés durant les dernières 24 heures, favorisant ainsi une bonne qualité des quantités déclarées. Puisque les quantités sont recueillies par déclaration du répondant, il est toutefois probable qu'elles soient affectées par certains effets de sous-déclaration ou de surdéclaration (Santé Canada, 2006). Ces effets peuvent varier selon les types d'aliments consommés ainsi que selon les caractéristiques des enfants telles que l'âge, par exemple. Il importe de mentionner que les quantités d'aliments déclarées au premier rappel alimentaire sont en moyenne supérieures à celles déclarées au second rappel, laissant croire à une sous-déclaration plus importante des répondants lors du second rappel. Ce dernier problème semble inhérent à la plupart des enquêtes de nutrition et a été pris en compte dans l'analyse, sous l'hypothèse que ce sont les quantités déclarées au premier rappel qui reflètent le mieux la réalité de consommation.

La qualité des données dépend également de l'exhaustivité et de l'exactitude des bases de données utilisées pour la conversion des aliments en nutriments et la classification des aliments au sein des grands groupes alimentaires. Plusieurs bases de données ont été utilisées afin de maximiser la qualité des données en ce sens.

L'estimation de la distribution des apports habituels n'est possible que lorsque la taille de l'échantillon pour chaque domaine d'âge et de sexe d'intérêt est suffisante. Qui plus est, il faut non seulement un nombre de répondants suffisant, mais également un nombre suffisant de données de deuxième rappel pour produire des statistiques décrivant les habitudes alimentaires d'une population et non uniquement la consommation alimentaire pour une journée. Une limite importante inhérente à cette enquête-ci tient au fait que les effectifs par domaine sont limités, entraînant des difficultés au niveau de l'ajustement du modèle d'erreur de mesure pour certaines combinaisons nutriment-domaine. Il n'en demeure pas moins que différentes stratégies d'analyse ont été mises en place pour résoudre la plupart des problèmes rencontrés et obtenir des estimations valides.

Par ailleurs, même valides, les estimations obtenues sont dans plusieurs cas associées à un CV élevé (supérieur à 15 %). Cela se produit fréquemment pour l'estimation des proportions relatives aux recommandations car ces quantités sont souvent petites. Comme il a été mentionné précédemment, en dépit d'une faible précision relative, une bonne part de ces statistiques demeurent informatives lorsqu'elles sont interprétées avec l'intervalle de confiance correspondant. Elles ont par conséquent été présentées lorsque pertinentes. Par exemple, la proportion de garçons de 14 à 18 ans pour lesquels les habitudes de consommation de viandes ou de substituts se situe en-deçà de la recommandation du GAC est estimée à 2,1 %, avec un CV de plus de 25 % et un intervalle de confiance de (0,0 %; 9,4 %). Un tel CV signifie que la proportion ne peut être interprétée à elle seule car elle est trop imprécise. L'intervalle de confiance fournit quant à lui une borne supérieure pour cette proportion, indiquant ici qu'un maximum de 9,4 % des garçons de 14 à 18 ans consomment, de façon habituelle, moins que la quantité minimale recommandée de 100 g par jour de viandes ou de substituts. C'est donc dire que la grande majorité des garçons de cet âge consomment suffisamment de viandes ou de substituts.

Enfin, puisque les estimations produites à l'aide de SIDE sont basées sur la distribution estimée des apports habituels pour la population visée et non sur une donnée ajustée pour chaque répondant, il n'est pas possible de réaliser des analyses multivariées à partir des données des rappels de 24 heures qui soient cohérentes avec les analyses univariées et bivariées. Pour ce faire, une donnée prédite de consommation habituelle serait requise pour chaque répondant. De plus, ces valeurs prédites devraient être accompagnées d'une mesure d'erreur liée à l'ajustement des données¹⁹. À ce jour, contrairement aux statistiques produites à l'aide de SIDE, les données individuelles après ajustement produites par ce même logiciel (données de consommation habituelles estimées) ne permettent pas d'obtenir des estimations non biaisées des paramètres d'intérêt de la population. C'est la raison pour laquelle aucune analyse multivariée n'a été réalisée à partir des données des rappels de 24 heures.

Tenant compte des difficultés inhérentes à toute enquête de nutrition ainsi que des effectifs limités pour l'estimation par domaine d'âge et de sexe, toutes les précautions prises dans l'analyse des apports nutritionnels et alimentaires permettent de croire que les statistiques présentées et interprétées avec leur mesure de précision apportent une information valable et pertinente sur les habitudes alimentaires des enfants québécois âgés de 1 à 18 ans.

19 .Les estimations de centiles et de proportions produites à l'aide de SIDE sont accompagnées d'une mesure de précision tenant compte à la fois de l'erreur due à l'échantillonnage et de l'erreur due à la modélisation.

Tableaux complémentaires

Tableau A1

Interventions effectuées dans l'analyse avec SIDE pour les apports nutritionnels – enfants québécois de 1 à 18 ans

Nutriments		Domaine Âge – sexe	Option LINFRAC > 0,01	Imputation Paramètres canadiens	Nb répliques « bootstrap » éliminées
Glucides totaux	H	9-13			2
		1-3			6
Sucres totaux	F	9-13			6
	H	14-18			8
Fibres alimentaires totales	H	14-18			2
	F	14-18			2
Lipides totaux	F	14-18			3
		4-8			5
Acides gras saturés	H	14-18	0,03		
Acides gras polyinsaturés totaux	F	14-18		√	
Acide linoléique	H	9-13			2
		1-3		√	
Cholestérol	H	9-13		√	
	F	14-18		√	
		4-8			3
Protéines/poids corporel mesuré	F	9-13	0,02		
	H	14-18	0,02		
	F	14-18	0,02		
% Apport énergétique en glucides	F	9-13			6
	F	14-18		√	
% Apport énergétique en protéines		1-3			3
	F	14-18			1
% Apport énergétique en lipides ²⁰	H	14-18	0,03		
	F	14-18		√	
% Apport énergétique en AGS ²¹		1-3			1
	H	14-18	0,02		
	F	14-18		√	
% Apport énergétique en AGMI	F	14-18		√	
% Apport énergétique en AGPI	H	14-18	0,02		
	F	14-18		√	
% Apport énergétique en acide linoléique	F	14-18		√	
		1-3	0,02		
Vitamine A		4-8	0,02		
	F	9-13			1
	H	14-18			2
Vitamine D	F	14-18	0,02		
	F	9-13		√	
Vitamine C	H	14-18	0,02		
Thiamine		4-8			1
Riboflavine	H	9-13			4
	F	9-13			3
Niacine	H	14-18			1
		4-8		√	
Vitamine B ₆	H	14-18	0,02		
Folate – Équivalents de folate alimentaire		4-8		√	
		1-3	0,03		1
Acide folique ²²		4-8		√	
	F	14-18			3
Vitamine B ₁₂	F	9-13		√	
Calcium	H	14-18	0,03		
		4-8			3
Magnésium	H	14-18			3
	H	14-18	0,03		
Zinc	F	14-18	0,03		
		4-8			2
Sodium	F	9-13			4

20. Pour l'ensemble des provinces autres que le Québec, l'option LINFRAC diffère de 0,01 pour le groupe des filles de 14 à 18 ans (0,04).

21. Pour l'ensemble des provinces autres que le Québec, une réplique « bootstrap » a été éliminée pour le groupe des enfants de 1 à 3 ans.

22. Pour l'ensemble des provinces autres que le Québec, huit répliques « bootstrap » ont été éliminées pour le groupe des filles de 14 à 18 ans. De plus, l'option LINFRAC est supérieure à 0,01 pour les enfants de 1 à 3 ans (0,02), les garçons de 9 à 13 ans (0,03) et de 14 à 18 ans (0,03).

Tableau A2

Interventions effectuées dans l'analyse avec SIDE pour les portions des grands groupes alimentaires - enfants québécois de 1 à 18 ans

Groupe alimentaire		Domaine Âge – sexe	Option LINFRACTION > 0,01	Imputation paramètres domaine élargi	Nb répliques « bootstrap » éliminées
Fruits et légumes		1-3	0,03		
	H	9-13	0,02		
	F	9-13	0,03		
	H	14-18			1
Produits céréaliers		1-3	0,03		
		4-8	0,02		
Produits laitiers	H	9-13	0,02		
	F	9-13	0,02		
	H	14-18	0,04		
	F	14-18	0,04		
Viandes et substituts		1-3	0,03		
		4-8		√	
	F	14-18	0,03		1

Bibliographie

BANKIER, M.D. (1988). « Power Allocations: determining sample sizes for subnational areas ». *The American Statistician* 42, 174-177.

DEPARTMENT OF STATISTICS AND CENTER FOR AGRICULTURAL AND RURAL DEVELOPMENT (2006). *A User's Guide to SIDE*, version 1.0, Technical Report 96-TR 30, Iowa State University.

NUSSER, S. M., CARRIQUIRY, A. L., DODD, K.W., AND FULLER, W. A. (1996). « A SEMIPARAMETRIC TRANSFORMATION Approach to Estimating Usual Daily Intake Distributions ». *Journal of the American Statistical Association*, 91, 1440-1449.

NUSSER, S.M., FULLER, W.A. AND GUENTHER, P.M. (1997). « ESTIMATING USUAL DIETARY INTAKE DISTRIBUTIONS : ADJUSTING for Measurement Error and Nonnormality in 24-Hour Food Intake Data ». *Survey Measurement and Process Quality*, New York : John Wiley & Sons.

SANTÉ CANADA (2006). *Enquête sur la santé dans les collectivités canadiennes, cycle 2.2, Nutrition* (2004), Guide d'accès et d'interprétation des données, 123p.

STATISTIQUE CANADA (2005). *Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 2.2, Nutrition (2004) - volet général sur la santé*, Guide du fichier de microdonnées à grande diffusion.

STATISTIQUE CANADA (2007). *Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 2.2 – Nutrition (2004)*, Documentation sur Software for Intake Distribution Estimation (SIDE).

L'analyse des données recueillies dans le cadre de l'*Enquête sur la santé dans les collectivités canadiennes (cycle 2.2)* de Statistique Canada a permis de tracer le portrait alimentaire et nutritionnel des jeunes québécois ainsi que de mieux comprendre les déterminants sociaux de leur alimentation. Cette publication présente une brève description de l'enquête et fait état de façon détaillée de la méthodologie statistique utilisée pour analyser et interpréter les données des rappels alimentaires de 24 heures, entre autres.

« L'Institut a pour mission de fournir des informations statistiques qui soient fiables et objectives sur la situation du Québec quant à tous les aspects de la société québécoise pour lesquels de telles informations sont pertinentes. L'Institut constitue le lieu privilégié de production et de diffusion de l'information statistique pour les ministères et organismes du gouvernement, sauf à l'égard d'une telle information que ceux-ci produisent à des fins administratives. Il est le responsable de la réalisation de toutes les enquêtes statistiques d'intérêt général. »

*Loi sur l'Institut de la statistique du Québec (L.R.Q., c. I-13.011)
adoptée par l'Assemblée nationale du Québec le 19 juin 1998.*

**Institut
de la statistique**

Québec 