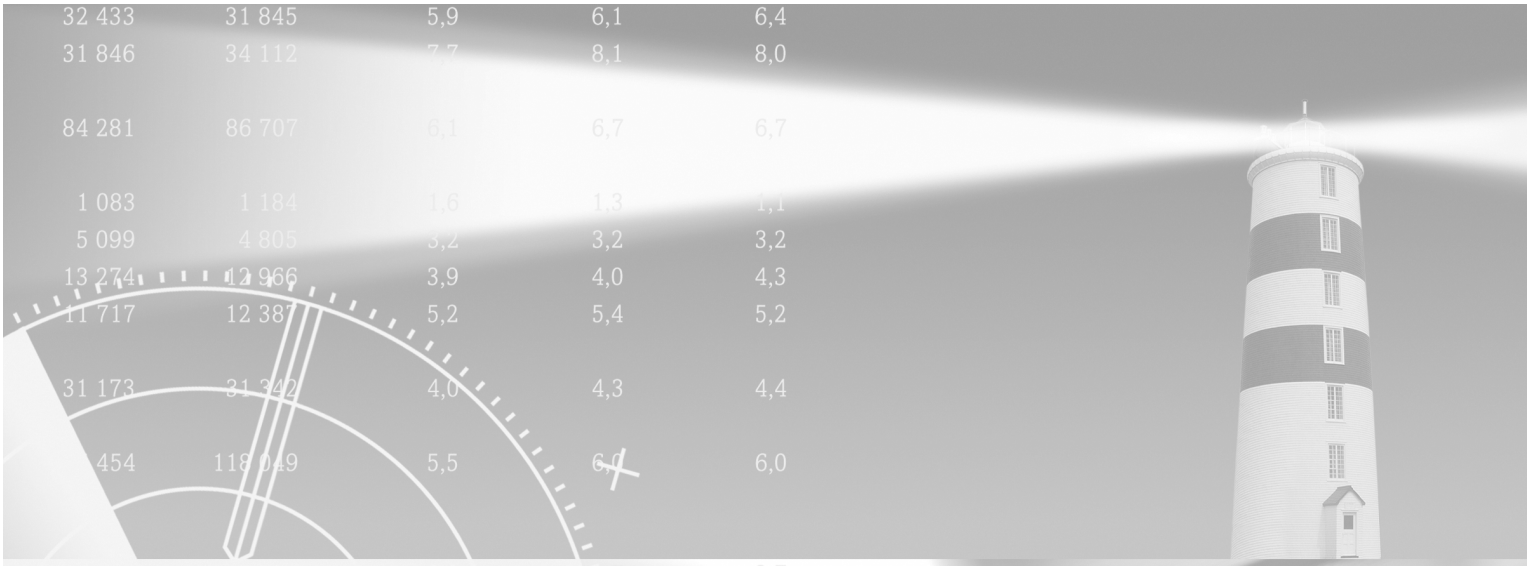
ALGORITHME DÉCISIONNEL relatif à l'utilisation des trois types de fichiers de données de l'Enquête sur la santé dans les collectivités canadiennes

Québec



ALGORITHME DÉCISIONNEL relatif à l'utilisation des trois types de fichiers de données de l'Enquête sur la santé dans les collectivités canadiennes

Document produit par le
Ministère de la Santé et des Services sociaux

En collaboration avec
l'Institut national de santé publique du Québec
et l'Institut de la statistique du Québec

Santé
et Services sociaux
Québec

Ce document a été produit dans le cadre des travaux du Comité tripartite MSSS-INSPQ-ISQ sur les aspects méthodologiques des données d'enquêtes sociosanitaires. Ce comité est composé de :

Ministère de la Santé et des Services sociaux
Josée Bourdages, Monique Godin et Johanne Labbé, Direction générale de la santé publique

Institut national de santé publique du Québec
Robert Choinière et Denis Hamel, Direction recherche, formation et développement
Suzanne Gingras, Direction des ressources informationelles

Institut de la statistique du Québec
Natalie Audet et Mikaël Berthelot, Direction générale adjointe aux statistiques et à l'analyse
Robert Courtemanche, Direction générale adjointe aux services à la productions statistique

Agence de santé et des services sociaux du Saguenay-Lac-Saint-Jean
Régis Couture, Direction de santé publique

Coordination et rédaction :
Monique Godin et Johanne Labbé

Collaboration à la rédaction :
Nathalie Audet, Robert Courtemanche, Denis Hamel

Secrétariat et mise en page :
Diane Brazeau, Direction générale de la santé publique, ministère de la Santé et des Services sociaux

Révision linguistique :
Virginie Rompré, Italiques

Édition :
La Direction des communications du ministère de la Santé et des Services sociaux du Québec

Le présent document s'adresse spécifiquement aux intervenants du réseau québécois de la santé et des services sociaux. Il a été édité en quantité limitée et n'est maintenant disponible qu'en version électronique sur les sites suivants : **http://intranetreseau.rtss.qc.ca** et **www.msss.gouv.qc.ca** section **Documentation**, rubrique **Publications**.

Le genre masculin utilisé dans ce document désigne aussi bien les femmes que les hommes.

Dépôt légal
Bibliothèque et Archives nationales du Québec, 2009
Bibliothèque et Archives Canada, 2009

ISBN : 978-2-550-55896-5 (version PDF)

Tous droits réservés pour tous pays. La reproduction, par quelque procédé que ce soit, la traduction ou la diffusion du présent document, même partielles, sont interdites sans l'autorisation préalable des Publications du Québec. Cependant, la reproduction partielle ou complète du document à des fins personnelles et non commerciales est permise, uniquement sur le territoire du Québec et à condition d'en mentionner la source.

© Gouvernement du Québec, 2009

Liste des acronymes

ASSS	Agence de la santé et des services sociaux
CADRISQ	Centre d'accès aux données de recherche de l'Institut de la statistique du Québec
CIQSS	Centre interuniversitaire québécois de statistiques sociales
DSQ	Direction Santé Québec
ESCC	Enquête sur la santé dans les collectivités canadiennes
FMGD	Fichier de microdonnées à grande diffusion
IMC	Indice de masse corporelle
ISQ	Institut de la statistique du Québec
INSPQ	Institut national de santé publique du Québec
MSSS	Ministère de la Santé et des Services sociaux

Table des matières

Introduction 7

I. Description de l'Enquête sur la santé dans les collectivités canadiennes 8

A. Caractéristiques générales de l'ESCC 8

B. Caractéristiques des différents fichiers de données de l'ESCC 9
1. Fichier de microdonnées à grande diffusion 9
2. Fichier de partage 9
3. Fichier maître 10

C. Comparabilité des résultats produits avec le fichier maître et le fichier de partage 12

II. Choix du fichier de l'ESCC 13

A. Critères de sélection du fichier à utiliser 13
1. Besoins des utilisateurs et fichiers correspondants 13
2. Accessibilité des fichiers 14
3. Précision statistique recherchée 14

B. Algorithme de décision 15

III. Exemples de produits de surveillance élaborés à partir de chacun des trois fichiers de l'Enquête sur la santé dans les collectivités canadienne 16

Conclusion 17

Références bibliographiques 18

ANNEXE I 19
ANNEXE II 25
ANNEXE III 33

Introduction

Les résultats de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) peuvent être accessibles par l'exploitation de trois types de fichiers de données, soit le fichier maître, le fichier de partage ainsi que le fichier de microdonnées à grande diffusion (FMGD). Il faut souligner qu'en février 2003, ce dernier a été désigné comme la source de données officielle québécoise, mais que, à l'usage, il s'est avéré qu'il ne pouvait répondre à tous les besoins.

Le fait que trois types de fichiers soient offerts pour produire des résultats sur de mêmes variables peut susciter de la confusion relativement au choix d'un fichier. La comparabilité des résultats tirés des trois fichiers peut aussi préoccuper les utilisateurs des données. À ce sujet, les membres du Comité tripartite MSSS-INSPQ-ISQ sur les aspects méthodologiques des données d'enquêtes sociosanitaires ont convenu que, pour la majorité des variables, les différences entre les résultats produits à partir de chaque fichier sont minimes et qu'en conséquence, l'utilisation d'un fichier plutôt qu'un autre n'a pas d'incidence majeure sur le résultat. Cette conclusion s'appuie principalement sur les résultats d'une étude de comparabilité des fichiers maître et de partage (annexes I et II) conduite dans le cadre des travaux d'un groupe de travail composé de représentants des instances de santé du Québec.

Le choix du type de fichier à utiliser est donc plus tributaire des besoins des utilisateurs (ex. : couverture territoriale, choix de catégories de variables), de la précision statistique recherchée et de l'accessibilité des fichiers que de considérations liées à la comparabilité des résultats. C'est donc avec l'objectif de fournir un outil simple et convivial pour guider les utilisateurs des données de l'ESCC dans le choix du fichier à privilégier que le comité tripartite a élaboré le présent document.

Ce document est divisé en trois sections. La première décrit sommairement l'Enquête sur la santé dans les collectivités canadiennes. Ses caractéristiques générales ainsi que celles de chacun des trois fichiers y sont présentées. L'étude de comparabilité des estimations produites avec les fichiers maître et de partage y est également traitée. La section II porte sur la sélection d'un fichier de l'ESCC. Les critères devant être pris en considération au moment de choisir un fichier sont d'abord énoncés, puis est présenté un algorithme de décision relatif au choix du fichier à privilégier. La section III fournit quant à elle des exemples de produits conçus avec chacun des fichiers.

En terminant, il est à noter que cet outil ne comporte pas de renseignements méthodologiques ou opérationnels liés à l'exploitation des fichiers; ces renseignements sont offerts dans d'autres documents, dont le *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan commun de surveillance - Enquête sur la santé dans les collectivités canadiennes cycle 1.1, 2.1 et 3.1*.

I. Description de l'Enquête sur la santé dans les collectivités canadiennes

A. Caractéristiques générales de l'ESCC

L'Enquête sur la santé dans les collectivités canadiennes (ESCC) est une enquête transversale visant à recueillir des renseignements sur l'état de santé, l'utilisation des services de santé et les déterminants de la santé de la population canadienne (Statistique Canada, 2005).

Jusqu'en 2006, cette enquête était conduite sur un cycle de deux ans. La première année du cycle, indiquée par la notation « .1 », consistait en une enquête générale fournissant des données représentatives des populations régionales. Quant à elle, l'enquête menée durant la seconde année du cycle portait sur un sujet particulier (enquête thématique) et fournissait uniquement des données représentatives des provinces. La première enquête générale de l'ESCC a été faite en 2000-2001 (cycle 1.1), la deuxième en 2003 (cycle 2.1) et la troisième en 2005 (cycle 3.1) (Institut de la statistique du Québec, 2006).

Depuis le cycle 4.1, débuté en 2007, l'ESCC se déroule selon un nouveau protocole : plutôt que de recueillir des données tous les deux ans, on procède à une collecte continue. Des contenus thématiques documentés sur une ou deux années, planifiés sur une période de six ans, seront étudiés tour à tour (Institut de la statistique du Québec, 2006).

Les données de l'ESCC sont recueillies auprès des personnes de 12 ans et plus pour l'enquête générale et de 15 ans et plus pour l'enquête thématique, vivant dans des logements privés des dix provinces et trois territoires. Les habitants des réserves indiennes et des terres de la Couronne, les personnes vivant dans un logement collectif institutionnalisé[1], les membres à temps plein des Forces canadiennes et les habitants de certaines régions éloignées sont exclus de la base de sondage. L'enquête générale de l'ESCC couvre ainsi environ 98 % de la population canadienne âgée de 12 ans et plus (Statistique Canada, 2005).

1. Les établissements carcéraux, les centres hospitaliers, les foyers pour personnes âgées sont des exemples de logements collectifs.

B. Caractéristiques des différents fichiers de données de l'ESCC

Statistique Canada produit trois types de fichiers pour analyser l'Enquête sur la santé dans les collectivités canadiennes. Les principales caractéristiques de ces fichiers sont décrites ci-dessous.

1. Fichier de microdonnées à grande diffusion

Le fichier de microdonnées à grande diffusion de Statistique Canada (FMGD) est fourni aux instances régionales et nationales (agences de la santé et des services sociaux, ministère de la Santé et des Services sociaux, Institut national de santé publique du Québec, Institut de la statistique du Québec) qui peuvent utiliser cette banque de données. Ce fichier est masqué afin de préserver la confidentialité des répondants (Bordeleau, 2003). Il comporte de l'information sur presque l'ensemble des répondants à travers le Canada, ce qui permet de faire des analyses comparatives interprovinciales.

Accessibilité :	Institut de la statistique du Québec, ministère de la Santé et des Services sociaux, agences de la santé et des services sociaux, Institut national de santé publique du Québec
Délai d'accès :	Accessible quelques mois après le fichier maître (sauf par télé-accès) et le fichier de partage
Couverture territoriale :	Canada, provinces et régions à l'exception des régions de moins de 70 000 habitants. Certaines de ces régions sont exclues, d'autres sont regroupées
Répondants :	100 % des répondants
Contenu :	Certains modules et variables des sous-échantillons ne sont pas présentés. Les possibilités de regrouper ou de scinder des catégories sont restreintes, en particulier pour les groupes d'âge
Comparabilité :	Interprovinciale et interrégionale
Précision :	Effets de plan ou grilles fournis par Statistique Canada

2. Fichier de partage

Le fichier de partage, dont l'Institut de la statistique du Québec est l'unique dépositaire officiel, est un fichier non masqué contenant l'ensemble des données relatives à la population québécoise. Il est extrait du fichier maître et est constitué à partir des répondants québécois qui ont accepté que les données recueillies par Statistique Canada soient transmises à l'Institut de la statistique du Québec dans le cadre d'une entente de partage[2] (Bordeleau, 2003). Le taux global de répondants du fichier québécois par rapport au fichier maître est de 95 % (2000-2001) et de 93 % (2003 et 2005).

2. Les données recueillies auprès des répondants des autres provinces ne font pas partie du fichier de partage québécois, bien que les autres provinces aient accès aux données du Québec dans leur propre fichier de partage. Une démarche devra être menée auprès de Statistique Canada par le Québec pour rétablir la situation.

Accessibilité :	Institut de la statistique du Québec et Centre d'accès aux données de recherche de l'Institut de la statistique du Québec (pour les utilisateurs externes)
Délai d'accès :	Accessible six à douze mois avant la diffusion du fichier de microdonnées à grande diffusion par Statistique Canada
Couverture territoriale :	Province de Québec et régions (incluant celles de moins de 70 000 habitants)
Répondants :	Pas tous les répondants du Québec (le taux de partage pouvant varier d'une enquête à l'autre)
Contenu :	L'ensemble des variables, y compris les modules de contenu de sous-échantillon. Possibilité de regrouper ou de scinder des catégories, en particulier pour les groupes d'âge
Comparabilité :	Interrégionale
Précision :	Poids *bootstrap*

3. Fichier maître

Le fichier maître est un fichier non masqué qui comprend l'ensemble des répondants canadiens ainsi que la totalité des variables (Bordeleau, 2003).

Accessibilité :	Télé-accès de Statistique Canada (service offert gratuitement avec le fichier de microdonnées à grande diffusion), Centre interuniversitaire québécois de statistiques sociales au Québec, base de données CANSIM, totalisations spéciales (payantes)
Délai d'accès :	Accessible six à douze mois avant le fichier de microdonnées à grande diffusion au Centre interuniversitaire québécois de statistiques sociales et sur la base de données CANSIM et au même moment que le fichier de microdonnées à grande diffusion par le service de télé-accès
Couverture territoriale :	Canada, provinces et régions
Répondants :	100 % des répondants
Contenu :	L'ensemble des variables, y compris les modules de contenu de sous-échantillon pour l'ESCC 2003. Certains résultats sont absents en raison de règles de diffusion (ex. : coefficient de variation supérieur à 33,3 %, cellules avec un nombre d'observations inférieur à dix)
Comparabilité :	Interprovinciale et interrégionale
Précision :	Poids *bootstrap*

TABLEAU 1 : Résumé des caractéristiques des trois fichiers de l'Enquête sur la santé dans les collectivités canadiennes

	FMGD	Fichier de partage	Fichier maître
Considérations liées à la couverture			
■ Couverture territoriale	- Canada - Provinces - Régions	- Québec - Régions 🔔	- Canada - Provinces - Régions
Régions de moins de 70 000 habitants	- Regroupées ou exclues 🔔	- Incluses dans le fichier	- Incluses dans le fichier
■ Comparabilité	- Interprovinciale - Interrégionale	- Interrégionale 🔔	- Interprovinciale - Interrégionale
■ Contenu	- Certains modules et variables 🔔	- Ensemble des variables - Contenu du sous-échantillon inclus	- Ensemble des variables - Contenu du sous-échantillon inclus - Certains résultats absents en raison des règles de diffusion
Regroupements de catégories en fonction des besoins (ex. : les groupes d'âge)	- Regroupements restreints 🔔 - Groupes d'âge fixes	- Regroupements possibles	- Regroupements possibles
Considérations méthodologiques			
■ Précision	- Seulement effets de plan ou grilles fournis par Statistique Canada 🔔	- Poids *bootstrap* accessibles	- Poids *bootstrap* accessibles
■ Répondants	- Ensemble des répondants à l'exception des régions exclues (< 70 000 habitants) 🔔	- Pas tous les répondants (les répondants qui ont accepté le partage de leurs données) 🔔	- Ensemble des répondants
Considérations liées à l'accès			
■ Accessibilité	- ISQ - MSSS - ASSS - INSPQ	- Traitement par l'ISQ 🔔 - Accès par le CADRISQ	- Télé-accès 🔔 - CIQSS au Québec - CANSIM
■ Délai d'accès	- Quelques mois après les fichiers maître (sauf par télé-accès) et de partage 🔔	- De six à douze mois avant le FMGD	- De six à douze mois avant le FMGD, sauf par télé-accès

🔔 : *En fonction des besoins, indique que le fichier présente des limites.*

C. Comparabilité des résultats produits avec le fichier maître et avec le fichier de partage

En 2002, l'Institut de la statistique du Québec a mené une étude avec les données de l'ESCC 1.1 dans le but d'évaluer l'incidence statistique que pouvait avoir le fait d'utiliser le fichier de partage (comportant environ 95 % des répondants québécois) au lieu du fichier maître (100 % des répondants québécois) (Côté et Courtemanche, 2002). L'étude faisait suite à une analyse de la comparaison directe des estimations obtenues à partir du fichier maître de Statistique Canada, par l'intermédiaire de CANSIM, avec celles du fichier de partage que détient l'Institut de la statistique du Québec. Cette analyse a porté sur les 25 caractéristiques sanitaires alors présentes dans CANSIM. L'incidence de la différence entre les estimations obtenues à partir des deux fichiers sur la précision statistique a été évaluée pour les caractéristiques affichant les plus grandes différences, et ce, pour l'ensemble des répondants du Québec et pour les répondants des deux régions sociosanitaires ayant le taux de partage le plus faible, soit l'Estrie et Laval.

Les résultats suggèrent que, pour les statistiques obtenues sur l'ensemble des répondants du Québec, l'incidence des « non-partageants » consiste en des changements minimes dans les statistiques provinciales, la différence entre la statistique obtenue du fichier maître et celle obtenue du fichier de partage étant inférieure ou égale à 0,3 % lorsqu'on exclut les non-répondants partiels. Ces différences sont négligeables en matière de précision statistique.

Pour les estimations régionales, on remarque une différence importante entre la statistique obtenue du fichier maître et celle obtenue du fichier de partage pour très peu de variables et très peu de régions. Pour ces variables, la différence au regard de la précision statistique est non négligeable.

Les résultats de l'analyse de la comparaison directe des estimations obtenues des deux fichiers ainsi que le rapport sur l'incidence statistique sont respectivement fournis aux annexes I et II.

II. Choix du fichier de l'ESCC

A. Critères de sélection du fichier à utiliser

1. Besoins des utilisateurs et fichiers correspondants

Le choix du fichier de données de l'ESCC à utiliser dépend, entre autres, des besoins des utilisateurs. Ces besoins se répartissent en trois catégories :

1. Couverture territoriale
   - Possibilité de faire des comparaisons interprovinciales (données extérieures au Québec).
   - Présence des résultats pour les territoires de moins de 70 000 habitants.
   - Possibilité de faire des comparaisons interrégionales (Québec).

2. Variables ou catégories de variables
   - Variables et catégories de variables présentes dans le fichier de microdonnées à grande diffusion.
   - Autres variables ou catégories de variables absentes du fichier de microdonnées à grande diffusion :
     - catégories (ex. : groupes d'âge);
     - variables (ex. : enfants exposés à la fumée secondaire, idéation suicidaire, variable à créer, comme le niveau d'activité physique).

3. Données récentes, non versées encore au fichier de microdonnées à grande diffusion

Le tableau ci-après présente les trois types de besoins des utilisateurs et indique quels fichiers de l'ESCC possèdent les caractéristiques nécessaires pour répondre à ces besoins.

TABLEAU 2 : Besoins des utilisateurs et types de fichiers de l'ESCC correspondants

		Variables ou catégories de variables		Données récentes non versées au FMGD
		Présentes dans le FMGD	Autres (variables exclues du FMGD)	
Couverture territoriale	Comparabilité interprovinciale	FMGD Fichier maître	Fichier maître	Fichier maître
	Régions de moins de 70 000 habitants	Fichier de partage Fichier maître	Fichier de partage Fichier maître	Fichier de partage Fichier maître
	Comparabilité interrégionale	FMGD Fichier de partage Fichier maître	Fichier de partage Fichier maître	Fichier de partage Fichier maître

2. Accessibilité des fichiers

D'autres critères doivent être considérés au moment de choisir le fichier de données à utiliser pour produire des résultats de l'ESCC. Un de ces critères est l'accessibilité des fichiers, laquelle varie selon les organisations impliquées.

Pour les agences de santé et des services sociaux, le ministère de la Santé et des Services sociaux et l'Institut national de santé publique du Québec, le fichier le plus accessible est le fichier de microdonnées à grande diffusion, chacune de ces organisations en possédant une copie. Lorsque les besoins des membres de ces organisations ne peuvent être comblés avec ce fichier, ils ont trois options : 1) traiter le fichier maître à distance avec le service de télé-accès de Statistique Canada, 2) se rendre dans les locaux du Centre d'accès aux données de recherche de l'Institut de la statistique du Québec pour traiter le fichier de partage ou 3) acheminer à l'Institut de la statistique du Québec une demande de totalisations spéciales à partir du fichier de partage. Dans ce dernier cas, précisons que des demandes d'envergure pourraient engendrer des frais pour l'utilisateur.

Pour l'Institut de la statistique du Québec, le fichier de microdonnées à grande diffusion et le fichier de partage sont aussi accessibles l'un que l'autre. Cette organisation pourrait utiliser le fichier maître seulement pour faire des comparaisons interprovinciales avec des variables ou catégories de variables absentes du fichier de microdonnées à grande diffusion.

3. Précision statistique recherchée

La précision statistique (variance, intervalle de confiance, coefficient de variation, etc.) peut être obtenue en utilisant soit l'effet de plan et les poids à moyenne un, soit les poids *bootstrap*. Ces derniers permettent d'obtenir une meilleure estimation de la précision, la combinaison « effet de plan – poids à moyenne un » restant une méthode approximative pour l'estimer. Malheureusement, les poids *bootstrap* ne sont pas accessibles avec le fichier de microdonnées à grande diffusion.

Si la précision statistique fournie par le fichier de microdonnées à grande diffusion est suffisante dans bien des situations d'analyse, les utilisateurs doivent néanmoins opter pour les poids *bootstrap* et le fichier maître ou le fichier de partage si :

- dans les différents tests statistiques, le seuil observé est près du seuil de signification fixé au départ (par exemple, une valeur *p* de 0,04 pour un test au seuil de 5 %);
- des analyses multivariées, comme la régression logistique, sont envisagées afin de tenir compte adéquatement du plan de sondage;
- le fichier maître ou le fichier de partage est retenu pour les analyses.

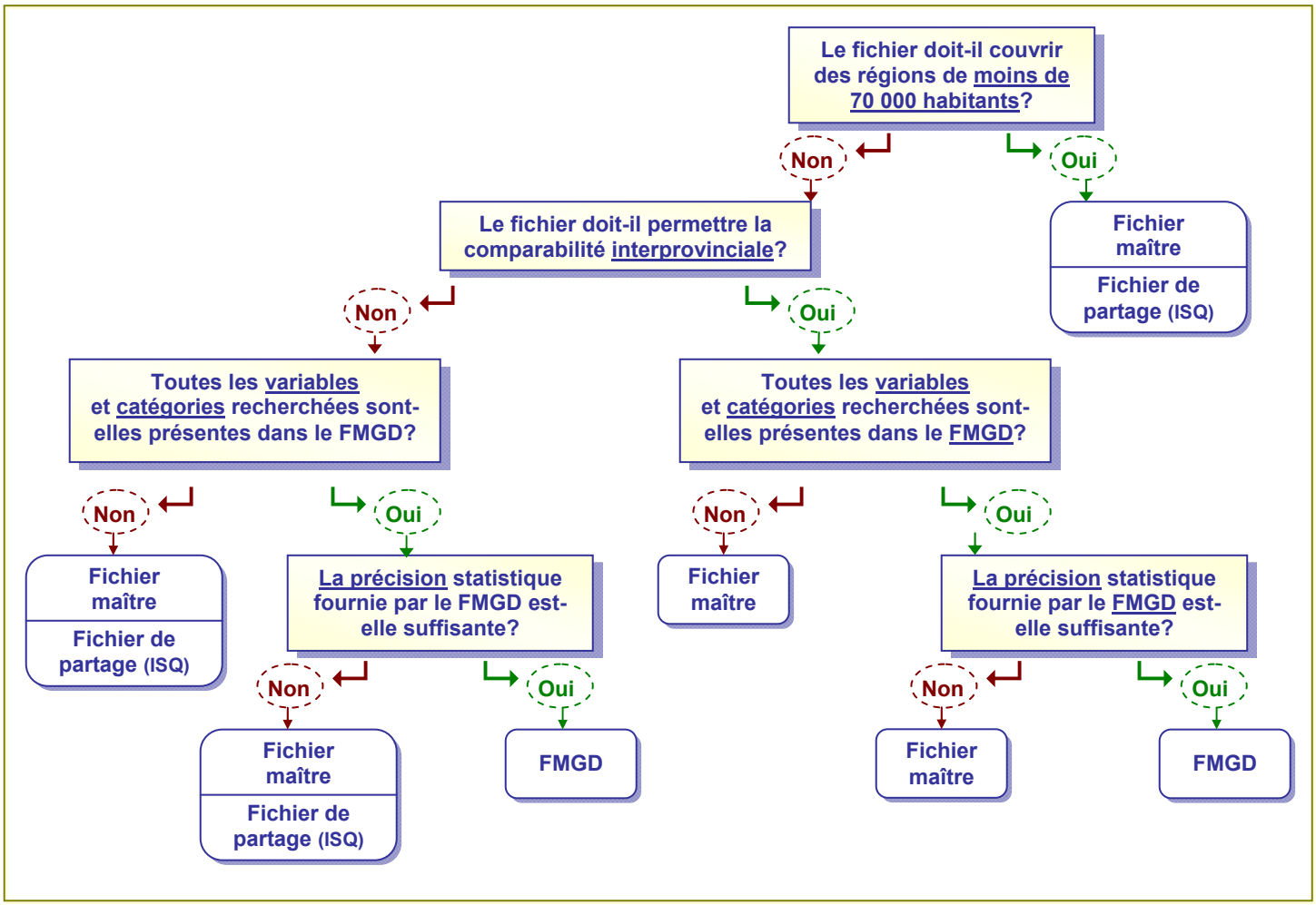
B. Algorithme de décision

Comme indiqué au tableau 2, plus d'un fichier de l'ESCC peut répondre aux mêmes besoins des utilisateurs. Pour guider le choix du fichier à privilégier, un algorithme de décision est présenté à la figure 1.

Celui-ci a été élaboré en considérant que le fichier de microdonnées à grande diffusion devrait être privilégié, car ce fichier est celui dont le ministère de la Santé et des Services sociaux, les agences de la santé et des services sociaux et l'Institut national de santé publique du Québec disposent.

Le critère lié à la non-accessibilité du fichier de microdonnées à grande diffusion n'est pas considéré, l'algorithme de décision s'appliquant lorsque les trois fichiers sont accessibles, soit à partir du moment où le fichier de microdonnées à grande diffusion est constitué.

FIGURE 1 : Algorithme décisionnel relatif à l'utilisation des trois types de fichiers de données de l'Enquête sur la santé dans les collectivités canadiennes



III. Exemples de produits de surveillance élaborés à partir de chacun des trois fichiers de l'Enquête sur la santé dans les collectivités canadiennes

Le recours à chacun des trois fichiers de l'ESCC pour la création de produits de surveillance est illustré ci-dessous.

TABLEAU 3 : Quelques exemples de produits de surveillance en fonction du type de fichier de l'ESCC utilisé

Fichiers	Référence des produits
Fichier de microdonnées à grande diffusion	■ BERNIER, Sylvie, et Denis HAMEL (2006). *Évolution de l'usage de la cigarette chez les Québécois de 15 ans et plus, de 1994-1995 à 2003*, Québec, Institut national de santé publique du Québec, 12 p. ■ BORDELEAU, Monique, et Issouf TRAORÉ (2007). « Santé générale, santé mentale et stress au Québec », *Zoom Santé*, juin, 4 p. ■ CAZALE, Linda, et Valeriu DIMITRU (2008). « Les maladies chroniques au Québec : quelques faits marquants », *Zoom Santé*, mars, 4 p. ■ FORTIN, Éric, et Issouf TRAORÉ (2007). « Les Québécois sont-ils satisfaits de leur vie et du cadre de vie dans lequel ils évoluent? », *Zoom Santé*, 4 p. ■ MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX DU QUÉBEC (2007). *Troisième rapport national sur l'état de santé de la population du Québec. Riches de tous nos enfants*, Québec, Gouvernement du Québec, 162 p.
Fichier de partage	■ AUBIN, Jacinthe, et Issouf TRAORÉ (2007). « L'excès de poids en France et au Québec en 2003 : un regard sur les caractéristiques socioéconomiques associées », *Zoom Santé*, juin, 4 p. ■ AUDET, Nathalie (2007). « L'évolution de l'excès de poids chez les adultes québécois de 1990 à 2004 : mesures directes », *Zoom Santé*, juin, 4 p. ■ BARALDI, Rosanna, et autres (2007). « L'embonpoint et l'obésité chez les jeunes québécois de 12 à 17 ans », *Zoom Santé*, juin, 4 p. ■ INSTITUT DE LA STATISTIQUE DU QUÉBEC (2007). *Compendium pour la région Nord-du-Québec, cycles 1.1, 2.1 et 3.1*, Québec, L'Institut. ■ LAPRISE, Patrick, et Brigitte BEAUVAIS (2007). « Les Québécois de plus en plus grands? », *Portraits et trajectoires*, novembre, 2 p.
Fichier maître	■ ANCTIL, Mélanie, et Serge CHEVALIER (2008). *Enquête de santé auprès des Cris 2003. Enquête sur la santé dans les collectivités canadiennes, cycle 2.1. Iiyiyiu Aschii, 2003 : Habitudes de vie en matière de consommation d'alcool, de drogues et les pratiques de jeux de hasard et d'argent*, Québec, Institut national de santé publique du Québec, 27 p. ■ LAMONTAGNE, Patricia, et Denis HAMEL (2008). *Poids corporel de la population adulte québécoise : Mise à jour 2005. Exploitation des données de l'ESCC 3.1*, Québec, Institut national de santé publique du Québec, 30 p. ■ NOLIN, Bertrand, Carole BLANCHET et Elena KUZMINA (2008). Enquête de santé auprès des Cris 2003. Enquête sur la santé dans les collectivités canadiennes, cycle 2.1. Iiyiyiu Aschii : Habitudes alimentaires, activité physique et poids corporel, Québec, Institut national de santé publique du Québec, 24 p.
Fichier maître et fichier de microdonnées à grande diffusion	■ INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC en collaboration avec le MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX DU QUÉBEC et l'INSTITUT DE LA STATISTIQUE DU QUÉBEC (2006). *Portrait de santé du Québec et de ses régions 2006 : Les statistiques – Deuxième rapport national sur l'état de santé de la population*, Québec, Gouvernement du Québec, 659 pages.

Conclusion

En résumé, le choix du fichier de l'ESCC à privilégier se fait sur la base de trois critères :

1) les besoins des utilisateurs (ex. : couverture territoriale, choix de catégories de variables)
2) l'accessibilité des fichiers
3) la précision statistique recherchée

Par ailleurs, il faut reconnaître que, bien que cela ait peu d'incidence sur les résultats, les estimations provenant des trois fichiers de données peuvent présenter de très légères différences. Pour ces raisons et pour éviter toute confusion éventuelle, il est fortement recommandé de mentionner la source de données utilisée dans la section des rapports portant sur la méthodologie et d'inscrire une note à ce sujet au bas des tableaux de résultats.

Références bibliographiques

BORDELEAU, Monique (2003). *ESCC : cycle 1.1. Synthèse de l'information diffusée lors de la formation de juin 2003*, Québec, Institut de la statistique du Québec, 4 pages.

CÔTÉ, Luc, et Robert COURTEMANCHE (2002). *Comparaison entre l'estimation du fichier maître et celle du fichier de partage pour quelques caractéristiques sanitaires, ESCC 2000-2001*, Québec, Institut de la statistique du Québec, 4 pages.

INSTITUT DE LA STATISTIQUE DU QUÉBEC – DIRECTION SANTÉ QUÉBEC (2006). *Document d'information pour la consultation sur le contenu optionnel québécois de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) cycle 4.1*, Québec, Institut de la statistique du Québec, 13 pages.

STATISTIQUE CANADA (2005). *Enquête sur la santé dans les collectivités canadiennes (ESCC)*, [En ligne]. [www.statcan.ca/cgi-in/imdb/p2SV_f.pl?Function=getSurvey&SDDS=3226&lang=fr&db=IMDB&dbg=f&adm=8&dis=2]. (Consulté le 10 janvier 2007).

ANNEXE I

Comparabilité des estimations du fichier maître et du fichier de partage Résumé de la comparaison directe des estimations

Extrait du document de travail *ESCC 2000-2001. Comparaisons (fichier maître de Statistique Canada et fichier de partage de l'ISQ) des estimations obtenues* de Monique Bordeleau, Direction Santé Québec, Institut de la statistique du Québec, 27 novembre 2002.

Résumé de la comparaison directe des estimations

En 2002, la Direction Santé Québec de l'Institut de la Statistique du Québec a effectué une comparaison des estimations produites avec le fichier de partage avec celles produites avec le fichier maître de l'ESCC 1.1. Le fichier de partage contient les données des répondants québécois qui ont accepté que l'information recueillie au cours de l'enquête soient transmises à l'Institut de la statistique du Québec, soit, en l'occurrence, 95 % des répondants initiaux. Les estimations du fichier maître ont été obtenues par l'intermédiaire de CANSIM, et les estimations du fichier de partage ont été produites par l'Institut de la statistique du Québec, dépositaire de ce fichier.

Le choix des indicateurs a été fait selon que les variables étaient présentes ou non dans CANSIM. Au total, 25 indicateurs ont été examinés pour la province et 24 pour les régions. Les résultats de cette étude sont basés sur des données pondérées; la non-réponse partielle est incluse dans le pourcentage total, car elle est considérée comme une catégorie de réponse. La Direction Santé Québec a sélectionné sept territoires d'analyse : l'ensemble du Québec, les deux régions ayant le taux de partage le plus faible (Laval, 92 %, et l'Estrie, 90 %), les deux régions ayant le taux de partage le plus élevé (Lanaudière, 97 %, et la Mauricie, 98 %) et les deux régions ayant les plus grandes tailles échantillonnales (Montréal et la Montérégie).

Résumé des résultats

La plupart des différences examinées sont négligeables. Les différences[1] entre les estimations produites avec le fichier de partage et celles produites avec le fichier maître sont inférieures à 1 %, et ce, pour tous les indicateurs de tous les territoires examinés, à l'exception des deux régions ayant le taux de partage le plus faible (Laval et Estrie). Pour ces dernières, seulement 2 indicateurs parmi les 24 examinés montrent des différences de plus de 1 % : l'activité physique et l'IMC.

COMPARAISON DES ESTIMATIONS ENTRE LE FICHIER DE PARTAGE ET LE FICHIER MAÎTRE DE L'ESCC, CYCLE 1.1

Territoire	Étendue des différences[1]	Nombre d'indicateurs présentant des différences[1] entre 0,5 % et 1,0 %	Indicateurs présentant la plus grande différence
Province	0 % à 0,3 %	0	• Consultation de professionnels pour la santé mentale • Risque de dépression • Vaccination contre la grippe
Laval et Estrie	0 % à 1,6 %	11	• Activité physique de loisir • IMC
Lanaudière et Mauricie	0 % à 0,7 %	5	• Consultation de professionnels pour la santé mentale
Montréal et Montérégie	0 % à 0,5 %	6	• Activité physique de loisir • Autoévaluation de la santé • État de santé fonctionnel • IMC • Douleur ou malaise qui a un effet sur les activités • Intensité des douleurs

[1] Différences de points de pourcentage entre les estimations produites avec le fichier de partage et celles produites avec le fichier maître.

Résultats détaillés

Pour l'ensemble du Québec

On observe des différences qui varient de 0,1 % à 0,3 % entre les estimations produites avec les données du fichier de partage de l'Institut de la statistique du Québec et celles produites à partir des données du fichier maître de Statistique Canada. Les indicateurs pour lesquels les plus grandes différences sont observées (0,3 %) sont les suivants :

- Consultation de professionnels pour la santé mentale :
  - Pas de consultation 12 derniers mois 0,3 %
  - Non déclaré 0,3 %
- Risque de dépression :
  - Aucun risque 0,3 %
- Vaccination contre la grippe :
  - Moins d'un an 0,3 %

Pour les deux régions où le taux de partage est le plus faible (Laval et Estrie)

Indicateurs pour lesquels une différence supérieure à 1 % est observée :

- Activité physique durant les loisirs :
  - Estrie : physiquement inactif 1,3 %
- IMC – norme internationale :
  - Estrie : embonpoint 1,1 %
- IMC – norme canadienne :
  - Laval : poids un peu excédentaire 1,6 %
  - Estrie : poids un peu excédentaire 1,2 %

Indicateurs pour lesquels on observe une différence allant de 0,5 % à 1 % :

- Auto-évaluation de la santé :
  - Estrie : excellente 0,7 %
  - Estrie : bonne 0,5 %
  - Estrie : passable ou mauvaise 0,6 %
- Consultation de médecins :
  - Estrie : consultation 12 derniers mois 0,6 %
  - Estrie : pas de consultation 12 derniers mois 0,5 %
- Consultation en médecine non traditionnelle :
  - Estrie : oui 0,5 %
  - Estrie : non 0,5 %

- Consommation de fruits et légumes :
  - Estrie : moins de cinq fois par jour 0,8 %
  - Estrie : plus de dix fois par jour 0,6 %

- Hypertension :
  - Laval : oui 0,6 %
  - Laval : non 0,6 %

- IMC – norme internationale :
  - Laval : poids acceptable 0,7 %
  - Estrie : poids acceptable 1,0 %
  - Laval : embonpoint 1,0 %

- IMC – norme canadienne :
  - Laval : poids acceptable 0,6 %
  - Estrie : poids acceptable 1,0 %
  - Laval : embonpoint 1,0 %

- Limitation des activités :
  - Estrie : oui 1,0 %
  - Estrie : non 1,0 %

- Stress dans la vie quotidienne :
  - Laval : stress assez intense 0,6 %

- Test de Pap :
  - Estrie : trois ans et plus 0,9 %

- Usage du tabac :
  - Laval : fumeurs réguliers 0,5 %
  - Estrie : fumeurs réguliers 1,0 %
  - Estrie : n'a jamais fumé 1,0 %

Pour les deux régions où le taux de partage est le plus élevé (Lanaudière et Mauricie)

Indicateurs pour lesquels une différence supérieure à 1 % est observée : aucun

Indicateurs pour lesquels on observe une différence allant de 0,5 % à 1 % :

- Consultation de professionnels pour la santé mentale :
  - Mauricie : pas de consultation 12 derniers mois 0,7 %

- Consommation de fruits et légumes :
  - Mauricie : entre cinq et dix fois par jour 0,5 %

- IMC – norme internationale :
  - Lanaudière : poids acceptable 0,5 %

- Risque de dépression :
  Mauricie : aucun risque 0,6 %

- Test de Pap :
  Mauricie : moins d'un an 0,6 %

Pour les deux régions ayant les plus grandes tailles échantillonnales (Montréal et Montérégie)

Indicateurs pour lesquels une différence supérieure à 1 % est observée : aucun

Indicateurs pour lesquels on observe une différence allant de 0,5 % à 1 % :

- Activité physique durant les loisirs :
  Montréal : physiquement inactif 0,5 %

- Autoévaluation de la santé :
  Montréal : passable ou mauvaise 0,5 %

- Douleur ou malaise qui a un effet sur les activités
  Montréal : aucune douleur ou malaise 0,5 %

- État de santé fonctionnel :
  Montréal : santé fonctionnelle très bonne ou parfaite 0,5 %

- IMC – norme canadienne :
  Montérégie : poids acceptable 0,5 %

- Intensité des douleurs :
  Montréal : aucune douleur ou malaise 0,5 %

ANNEXE II

Comparaison entre l'estimation du fichier maître et celle du fichier de partage pour quelques caractéristiques sanitaires, ESCC 2000-2001

Luc Côté et Robert Courtemanche
Direction de la méthodologie, de la démographie et des enquêtes spéciales
Institut de la statistique du Québec
Version du 17 janvier 2002 (révision linguistique de février 2009)

Comparaison entre l'estimation du fichier maître et celle du fichier de partage pour quelques caractéristiques sanitaires, ESCC 2000-2001

Nous avons analysé l'incidence statistique que pouvait avoir le fait d'utiliser le fichier de partage que détient l'Institut de la statistique du Québec au lieu du fichier maître, duquel est tiré le fichier de microdonnées à grande diffusion (FMGD), aussi connu sous l'acronyme anglais PUMF. L'étude a porté sur 25 caractéristiques sanitaires analysées au préalable par Monique Bordeleau de la Direction Santé Québec. Ces caractéristiques constituaient les seules variables accessibles sur le site de Statistique Canada, seules variables à partir desquelles on pouvait donc établir une base de comparaison.

Pour que les statistiques obtenues du fichier de partage soient comparables à celles tirées du fichier maître, nous avons d'abord procédé à la redistribution du taux de non-réponse partielle[1] lorsque nous avons traité des données provenant de l'un ou l'autre des deux fichiers. Cette étape s'est avérée essentielle parce que le taux de non-réponse partielle n'est pas nécessairement le même dans le fichier de partage[2] et le fichier maître. Il importe de souligner que, sur le site Internet de Statistique Canada (fichier maître) et dans le document préparé par Monique Bordeleau (fichier de partage), la non-réponse partielle, ce que Statistique Canada appelle les « non-déclarés », paraît sous la forme d'une modalité de réponse. Autrement dit, la non-réponse partielle constitue en soi une réponse possible à une question et elle est traitée comme telle, ce qui peut poser des problèmes de comparaison si les taux de non-réponse partielle du fichier maître et du fichier de partage sont assez différents. L'exemple de la page suivante illustre ce cas.

Par la suite, pour les caractéristiques sanitaires pour lesquelles les statistiques obtenues des fichiers maître et de partage affichaient les plus grandes différences, nous avons examiné si ces différences étaient ou non négligeables. Pour ce faire, nous avons supposé que la différence entre la statistique obtenue avec le fichier de partage et celle obtenue avec le fichier maître constituait une estimation du biais. En se rapportant à l'exemple présenté ci-après, on constate que l'estimation du biais peut être assez différente avant et après redistribution de la non-réponse partielle, celle-ci passant de 0,3 % à 0,04 % dans le cas illustré.

1. Le taux de non-réponse partielle est le rapport entre le nombre pondéré de personnes ayant répondu à une question donnée et le nombre pondéré de personnes devant y répondre. Nous utilisons le « nombre pondéré », car les personnes sélectionnées n'ont pas toutes la même probabilité d'être sélectionnées. À ne pas confondre avec le taux de réponse à l'enquête, qui est défini comme le rapport entre le nombre pondéré de personnes ayant répondu au questionnaire de l'enquête et le nombre pondéré de personnes admissibles au questionnaire de l'enquête.

2. Le fichier de partage n'inclut que les « partageants », c'est-à-dire les répondants qui ont accepté que les données qu'ils ont fournies soient transmises à l'Institut de la statistique du Québec. Ceux qui ont refusé (les « non-partageants ») ne s'y retrouvent pas. Le fichier maître inclut donc les partageants et les non-partageants.

Exemple

Redistribution de la non-réponse partielle pour la question concernant la consultation d'un professionnel pour un problème de santé mentale

Réponses possibles	**Fichier maître (%)**	**Fichier de partage (%)**	**Différence (biais) (%)**
Consultation 12 derniers mois	8,8	8,8	0
Pas de consultation 12 derniers mois	90,4	90,7	0,3
Non déclaré (non-réponse partielle)	0,8	0,5	0,3
Après redistribution de la non-réponse partielle[1]			
Consultation 12 derniers mois	8,87	8,83	0,04
Pas de consultation 12 derniers mois	91,13	91,17	0,04

1. Redistribution de la non-réponse partielle : on suppose que les non-répondants partiels se seraient comportés de la même façon que ceux qui ont répondu à cette question. Autrement dit, 90,4 % des personnes qui n'ont pas répondu à cette question (0,8 %) n'auraient pas consulté de professionnel au cours des douze derniers mois, tandis que 8,8 % de ce même groupe en aurait consulté un.

Rappelons que nous voulons savoir si le fait de disposer d'un sous-échantillon de répondants (les « partageants ») affecte la qualité des données. Autrement dit, nous voulons savoir jusqu'à quel point le biais associé au fait que les « non-partageants » puissent se comporter différemment des « partageants » affecte la précision de la statistique obtenue à partir du fichier de partage.

La précision est une quantité qu'on estime à l'aide de l'erreur totale. Celle-ci est constituée de deux composantes : l'erreur-type (ou la variance échantillonnale) et le biais. L'erreur-type (ou la variance échantillonnale) est la statistique la plus couramment utilisée pour quantifier la précision. Comme il a été fait mention plus haut, on pose comme hypothèse que la différence entre les deux fichiers constitue une estimation du biais, ce qui implique que nous sommes en mesure d'estimer l'erreur totale. Nous voulons donc voir quelle est l'incidence de ce biais sur l'estimation de la précision en utilisant l'erreur-type plutôt que l'erreur totale. Autrement dit, nous voulons déterminer l'importance de ce biais par rapport à l'erreur totale. On peut dire que la précision de la statistique obtenue n'est pas compromise lorsque le biais représente moins de 10 % de l'erreur totale.

L'analyse porte en premier lieu sur l'ensemble des répondants du Québec et, en second lieu, sur les répondants des deux régions sociosanitaires ayant le taux de partage le plus faible, soit l'Estrie et Laval. Elle se termine sur la discussion d'autres aspects touchant les fichiers maître et de partage.

Analyse pour l'ensemble du Québec

Pour les statistiques obtenues sur l'ensemble des répondants du Québec pour les 25 variables analysées par Monique Bordeleau, on remarque que la différence entre la statistique obtenue du fichier maître et celle obtenue du fichier de partage est inférieure ou égale à 0,3 % lorsque l'on exclut les non-répondants partiels.

Nous avons observé que l'erreur-type associée à quelques statistiques obtenues à partir du fichier de partage sous-estime de 10 à 20 % l'erreur totale (voir tableau I).

Cependant, pour la majorité des statistiques obtenues du fichier de partage, cette sous-estimation est négligeable (moins de 10 %).

Analyse régionale (Estrie et Laval)

L'Estrie et Laval sont les deux régions sociosanitaires qui présentent le taux de partage le plus faible. L'analyse a été effectuée séparément pour chacune de ces deux régions.

Ainsi, pour les statistiques obtenues sur l'ensemble des répondants pour les 25 variables analysées par Monique Bordeleau, on remarque que la différence entre la statistique obtenue du fichier maître et celle obtenue du fichier de partage est plus importante que pour la province, toujours en excluant les non-répondants partiels (voir tableau I). Pour les régions ayant un taux de partage plus élevé, cette différence s'amenuise.

Par ailleurs, nous avons observé que l'utilisation de l'erreur-type pour estimer l'erreur totale fait en sorte que la sous-estimation peut atteindre 50 %. Cependant, la sous-estimation est généralement négligeable (moins de 10 %).

Tableau I
Sous-estimation de l'erreur totale en utilisant le fichier de partage pour certaines caractéristiques sanitaires ciblées

		Sous-estimation de l'erreur totale (%)
Pour l'ensemble du Québec		
Consultation d'un professionnel pour santé mentale		
	Pas de consultation 12 derniers mois	2,35
Risque de dépression		
	Aucun risque	1,34
	Risque probable	17,53
Vaccination contre la grippe		
	Moins d'un an	12,20
	Jamais	4,88

Pour l'Estrie et Laval		
Activité physique durant les loisirs		
	Estrie : physiquement actif	16,34
	Estrie : physiquement inactif	10,89
IMC : norme internationale		
	Estrie : embonpoint	18,09
IMC : norme canadienne		
	Estrie : poids un peu excédentaire	23,52
	Laval : poids un peu excédentaire	51,97

Autres aspects

L'analyse effectuée ci-dessus a porté uniquement sur la répartition des répondants entre les différentes modalités de réponse relatives à une caractéristique sanitaire (par exemple, ne fume pas, ancien fumeur, fumeur). Nous n'avons pas examiné le cas où des statistiques sont désirées pour des sous-populations comme les groupes d'âge ou le sexe. Cependant, si on calculait les différences pour de telles sous-populations, elles devraient être plus grandes que celles observées pour l'ensemble des répondants.

Également, à partir du fichier de microdonnées à grande diffusion de Statistique Canada, lequel exclut la région du Nord-du-Québec, les statistiques seront légèrement différentes de celles obtenues avec le fichier maître de Statistique Canada. Ces différences devraient être inférieures à celles observées entre le fichier maître et le fichier de partage.

Conclusion

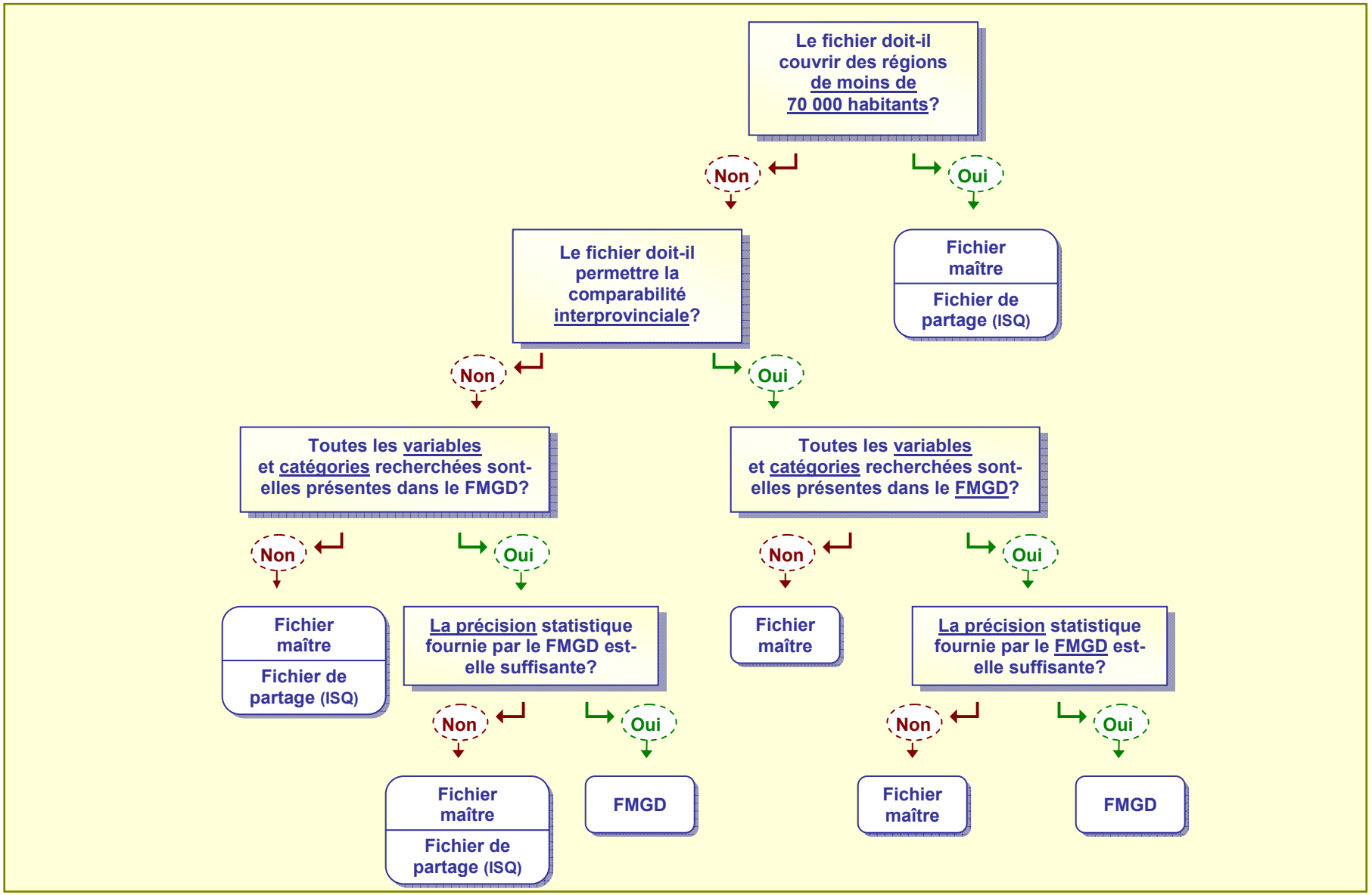
À la suite de ces analyses, nous pouvons conclure que l'incidence des « non-partageants » consiste en des changements négligeables dans les statistiques provinciales et que l'erreur totale est peu modifiée. Pour ce qui est des estimations régionales, on observe un changement non négligeable pour très peu de variables et très peu de régions. Quant à l'erreur totale, il y a augmentation seulement pour ces variables.

ANNEXE III

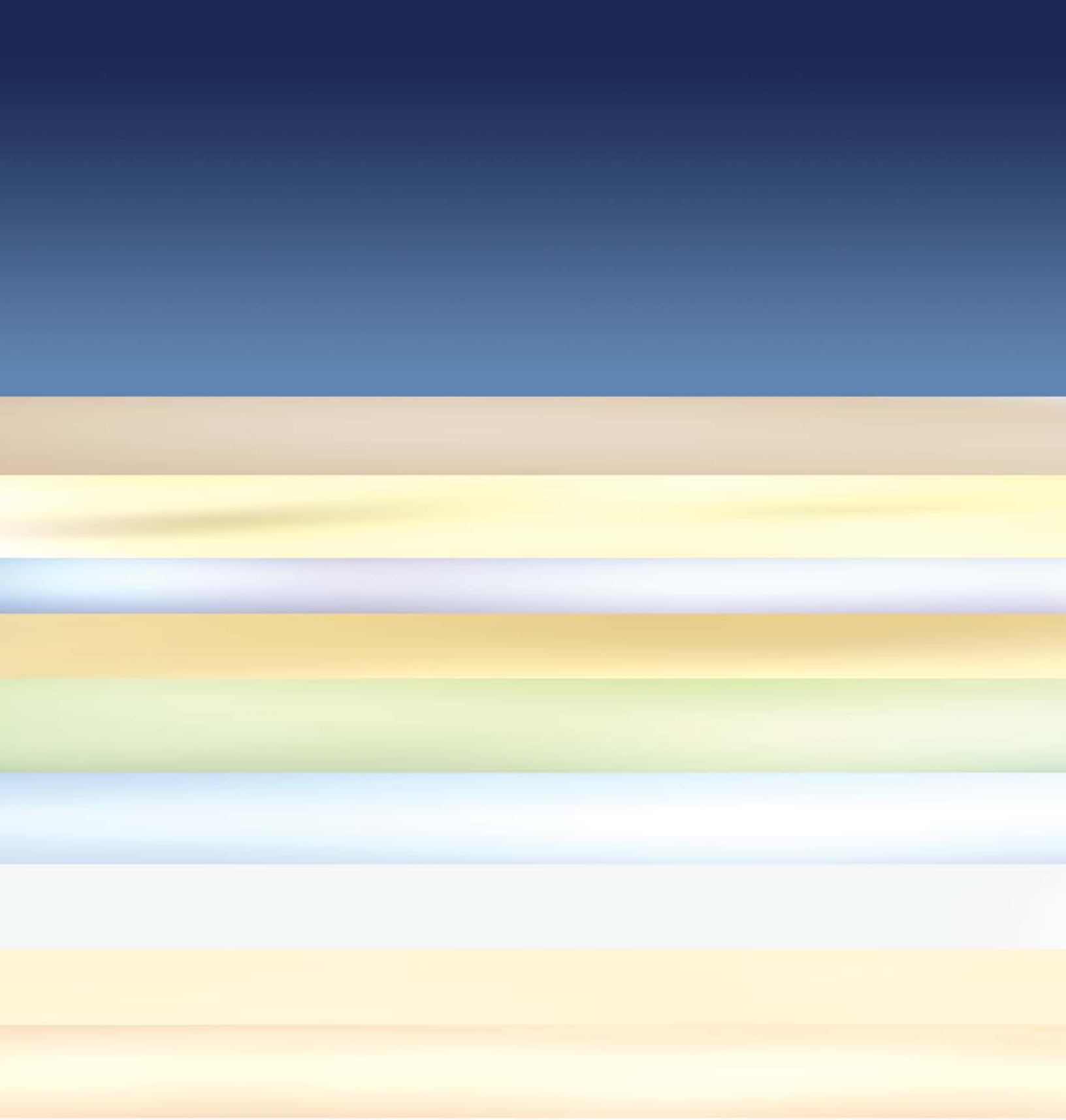
Aide-mémoire

Algorithme décisionnel relatif à l'utilisation des trois types de fichiers de données de l'Enquête sur la santé dans les collectivités canadiennes

Algorithme décisionnel relatif à l'utilisation des trois types de fichiers de données de l'Enquête sur la santé dans les collectivités canadiennes



MSSS en collaboration avec l'INSPQ et l'ISQ (2009)



www.msss.gouv.qc.ca

Santé
et Services sociaux
Québec

09-202-01F