



Plan national de surveillance

Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires

– Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017 –

Institut de la statistique du Québec

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

Février 2019

Pour tout renseignement concernant l'Institut de la statistique du Québec (ISQ) et les données statistiques dont il dispose, s'adresser à :

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

ou

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

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

Ce document est disponible seulement en version électronique.

Dépôt légal
Bibliothèque et Archives nationales du Québec
1^{er} trimestre 2019
ISBN : 978-2-550-83447-2 (PDF)

© Gouvernement du Québec, Institut de la statistique du Québec, 2019

Toute reproduction autre qu'à des fins de consultation personnelle
est interdite sans l'autorisation du gouvernement du Québec.

www.stat.gouv.qc.ca/droits_auteur.htm

Février 2019

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

Coordination et rédaction

France Lapointe, Direction de la méthodologie, ISQ

Collaboration à la rédaction

Micha Simard, Direction des enquêtes longitudinales et sociales, ISQ

Mathieu Langlois, Infocentre de santé publique, INSPQ

Sénaba Sambe, Infocentre de santé publique, INSPQ

Robert Courtemanche, Direction de la méthodologie, ISQ

Relecture

Outre les personnes déjà mentionnées ci-dessus :

Lucie Gingras, Direction des enquêtes longitudinales et sociales, ISQ

Christian Macé, MSSS

Révision linguistique

Sarah Bélanger, Direction de la diffusion et des communications, ISQ

Édition

Alexandra Lanthier, Direction des enquêtes longitudinales et sociales, ISQ

Notice bibliographique suggérée

INSTITUT DE LA STATISTIQUE DU QUÉBEC, en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC (2018). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017*, Québec, Gouvernement du Québec, 60 p.

Introduction.....	7
I. Description générale de l'Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017 (EQPPEM)	8
1.1 Contexte général.....	8
1.2 Objectifs de l'enquête	8
1.3 Accès aux données.....	9
II. Aspects méthodologiques et statistiques	10
2.1. Plan d'échantillonnage	12
2.1.1 Population visée	12
2.1.2 Base de sondage	13
2.1.3 Taille et répartition de l'échantillon	14
2.2 Collecte des données.....	16
2.2.1 Prétest	16
2.2.2 Stratégies et opérations de collecte.....	17
2.2.3 Résultats de collecte.....	18
2.3 Traitement des données	20
2.3.1 Validation et saisie.....	20
2.3.2 Pondération.....	21
2.3.3 Analyse de la non-réponse partielle	22
2.3.4 Imputation du revenu familial.....	24
2.3.5 Précision des estimations	25
2.3.6 Analyses et tests statistiques	26
2.4 Portée et limites de l'enquête	27
III. Aspects méthodologiques spécifiques de l'Infocentre	29
3.1 Mesures de précision supplémentaires.....	29
3.1.1 Intervalle de confiance	29
3.1.2 Erreur-type	30
3.2 Non-réponse partielle pour les indicateurs de l'Infocentre	30
3.2.1 Taux de non-réponse partielle pour les estimations à l'échelle régionale et locale	30
3.2.2 Traitement pour l'estimation des proportions et des effectifs.....	30
3.3 Ajustement des proportions selon l'âge ou le sexe.....	30
3.4 Tests statistiques et seuils de signification disponibles à l'Infocentre	31
3.4.1 Test global d'indépendance	31

3.4.2	Test d'égalité de deux proportions.....	31
3.4.3	Seuils de signification disponibles à l'Infocentre	32
IV.	Exploitation des données de l'EQPPEM 2017 sur le portail de l'Infocentre de santé publique.....	33
4.1	Production de tableaux et comparaison de proportions sur le portail de l'Infocentre de santé publique	33
4.1.1	Paramètres d'une requête sur le portail de l'Infocentre de santé publique	33
4.1.2	Exemples de résultats de requêtes sur le portail de l'Infocentre de santé publique et recommandations	37
4.2	Particularités des analyses produites à l'Infocentre de santé publique	42
4.2.1	Nombre trop important de tests demandés pour une requête	43
4.2.2	Non-réponse partielle	44
4.2.3	Diffusion à l'échelle locale pour deux régions sociosanitaires de résidence	46
4.2.4	Limites de certaines variables de croisement.....	47
4.2.5	Limites de certains indicateurs.....	48
V.	Protection des renseignements personnels	49
5.1	Confidentialité des tableaux destinés à la diffusion	49
5.2	Règles de masquage par suppression.....	51
VI.	Recommandations pour l'interprétation des résultats	53
	Références bibliographiques	54
	Annexe 1 - Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre de santé publique.....	58

L'*Enquête québécoise sur le parcours préscolaire des enfants de la maternelle 2017* (EQPPEM) est rattachée à l'*Enquête québécoise sur le développement des enfants à la maternelle* (EQDEM) de 2017. Celle-ci s'inscrit dans le Plan ministériel d'enquêtes sociales et de santé à titre d'enquête planifiée à des fins de surveillance. Elle vise notamment à répondre à l'un des besoins liés aux plans de surveillance prescrits par la Loi sur la santé publique.

Plusieurs acteurs, tant sur le plan national qu'à l'échelle régionale, et divers experts de contenu ont été associés de près aux différentes étapes de réalisation de l'EQPPEM 2017. Dans la foulée des travaux visant à rehausser la capacité de surveillance au Québec, des efforts considérables de développement ont été consentis pour rendre disponible, sur le portail de l'Infocentre de santé publique, l'information nécessaire à l'analyse de chaque indicateur mesuré par l'EQPPEM 2017. Cette enquête est la huitième à être traitée à l'Infocentre de santé publique dans le cadre de la mise en œuvre du Plan national de surveillance (PNS). Contrairement à la plupart des enquêtes traitées à l'Infocentre, l'EQPPEM ne comporte qu'un cycle; la comparabilité dans le temps n'est donc pas un enjeu ici.

Ce guide vise à soutenir les utilisateurs dans l'analyse et l'interprétation des résultats de l'EQPPEM 2017 disponibles dans le rapport et sur le site Web de l'ISQ ainsi que sur le portail de l'Infocentre de santé publique, en leur fournissant l'information nécessaire pour leur permettre :

1. de comprendre les balises méthodologiques de l'enquête et de l'analyse des résultats;
2. d'interpréter les mesures de précision et les tests statistiques;
3. de choisir l'option méthodologique selon les besoins parmi celles offertes sur le portail.

Ce document a été rédigé en complément du *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le développement des enfants à la maternelle 2012*¹ (ISQ, 2013), du *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le développement des enfants à la maternelle 2017*¹ (ISQ, 2018), des documents *Méthodologie de l'Enquête québécoise sur le développement des enfants à la maternelle 2017* (Tremblay et Simard, 2018) et *Méthodologie de l'Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017* (Lapointe et autres, 2019), ainsi que des rapports de l'EQDEM 2017 (Simard et autres, 2018) et de l'EQPPEM 2017 (Lavoie et autres, 2019).

La première partie donne une description générale de l'EQPPEM 2017. La seconde porte sur les aspects méthodologiques et statistiques généraux de l'enquête, alors que la troisième traite des aspects méthodologiques spécifiques à l'analyse des données sur le portail de l'Infocentre de santé publique. La partie suivante aborde la question de l'exploitation proprement dite des données de l'EQPPEM 2017, alors que les deux dernières parties font référence à la protection des renseignements personnels et aux recommandations d'usage pour l'interprétation des résultats obtenus.

1. Le lecteur est invité à consulter les guides méthodologiques de l'EQDEM 2012 et 2017, notamment pour prendre connaissance de certains aspects de l'analyse statistique traités dans les annexes qui s'appliquent au traitement des données de l'EQPPEM 2017.

I. Description générale de l'Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017 (EQPPEM)

1.1 Contexte général

En 2017, au Québec, selon l'instrument de mesure du développement de la petite enfance (IMDPE)², un peu plus du quart (28 %) des enfants de maternelle appartient à la catégorie des enfants considérés comme vulnérables dans au moins un des aspects de leur développement, que ce soit sur le plan physique, affectif, social, langagier ou cognitif (Simard et autres, 2018). Si ces résultats attirent l'attention des différents acteurs œuvrant auprès des enfants, c'est qu'il est largement admis que la vulnérabilité à la maternelle diminue les chances de ces enfants de réussir tout au long de leur parcours, tant du point de vue social que sur le plan scolaire (Kershaw et autres, 2010; McCain et autres, 2007; Forget-Dubois et autres, 2007; Lemelin et autres, 2007).

On reconnaît également que les premières années de vie des enfants jouent un rôle fondamental dans la façon dont ils se développeront tout au long de leur vie. Plusieurs facteurs individuels, familiaux, sociaux et environnementaux peuvent influencer la façon dont se développent les enfants durant cette période charnière qu'est la petite enfance.

Ce constat général a mené l'Institut de la statistique du Québec à réaliser en 2017 l'*Enquête québécoise sur le parcours préscolaire des enfants de maternelle* (EQPPEM) afin de recueillir des renseignements permettant de décrire certains aspects du parcours préscolaire des enfants et de vérifier, comme la littérature semble le soutenir, si ceux-ci sont associés à leur état de développement au moment où ils fréquentent la maternelle.

Cette enquête est d'ailleurs rattachée à l'*Enquête québécoise sur le développement des enfants à la maternelle* (EQDEM) 2017 : toutes deux sont issues d'un partenariat de cinq ministères et organismes, soit Avenir d'enfants, le ministère de la Santé et des Services sociaux (MSSS), le ministère de la Famille, le ministère de l'Éducation et de l'Enseignement supérieur (MEES) ainsi que l'Institut de la statistique du Québec (ISQ).

1.2 Objectifs de l'enquête

L'objectif principal de l'EQPPEM est de fournir plus d'information sur les divers facteurs qui pourraient être associés au niveau de développement de l'enfant à la maternelle 5 ans.

De manière plus détaillée, l'EQPPEM vise à :

- dresser un portrait du parcours préscolaire des enfants québécois fréquentant une classe maternelle (s'inscrivent dans ce parcours les services de garde, la maternelle 4 ans, le programme Passe-Partout et les haltes-garderies communautaires);
- relever d'autres facteurs qui pourraient être associés au développement de l'enfant : on retient ici des renseignements sur la santé de l'enfant, sur son environnement familial et sur certaines pratiques parentales, par exemple le fait pour les parents d'initier leur enfant à la lecture;
- fournir de l'information sur les caractéristiques socioéconomiques des parents et des familles des enfants à l'étude;

2. © McMaster University, Ontario. L'IMDPE est largement utilisé dans les autres provinces canadiennes et à l'étranger sous l'appellation *Early Development Instrument* (EDI).

- mettre en lien l'expérience préscolaire des enfants et leurs conditions de vie avec les différents domaines de développement mesurés dans l'EQDEM³.

C'est par la combinaison des données de l'EQPPEM et de l'EQDEM qu'il est possible d'atteindre ce dernier objectif.

1.3 Accès aux données

Le **fichier de données maître** contient l'ensemble des informations non nominatives recueillies pour les enfants de maternelle de 16 régions sociosanitaires; les régions du Nunavik (17) et des Terres-Cries-de-la-Baie-James (18) ne sont pas couvertes par l'enquête. Le fichier maître est accessible aux chercheurs au Centre d'accès aux données de recherche de l'ISQ (CADRISQ) dans un environnement contrôlé et sécuritaire pour la confidentialité des données.

Lors de la collecte de données, les parents participant à l'enquête devaient consentir à ce que les réponses fournies soient transmises au ministère de la Santé et des Services sociaux (MSSS), à l'Institut national de santé publique (Infocentre), au ministère de l'Éducation et de l'Enseignement supérieur (MEES), au ministère de la Famille et à l'organisme Avenir d'enfants (AE). Chacun de ces organismes a obtenu l'accès au **fichier de partage**, qui correspond au fichier maître duquel ont été retirées les informations nominatives concernant les « enfants » répondants à l'EQPPEM.

3. Cette enquête utilise l'Instrument de mesure du développement de la petite enfance (IMDPE). Copyright © McMaster University, Ontario. L'IMDPE est largement utilisé dans les autres provinces canadiennes et à l'étranger sous l'appellation Early Development Instrument (EDI). Pour plus d'information à propos des expériences canadiennes, se référer au site Web de l'Offord Centre for Child Studies (OCCS) : www.offordcentre.com/readiness/index.html.

II. Aspects méthodologiques et statistiques

Afin de faire une utilisation adéquate des données et des résultats issus de l'EQPPEM, il importe de connaître la méthodologie utilisée. En effet, le plan d'échantillonnage de l'enquête, les procédures de collecte et le traitement des données ont tous une incidence sur les estimations produites. La connaissance de ces éléments permettra à l'utilisateur des données d'en faire une analyse judicieuse, laquelle pourra être utile à l'élaboration de plans d'action appropriés et adaptés à la population visée. Cette partie du guide présente la méthodologie générale de l'enquête.

L'encadré 2.1 présente une synthèse des principaux aspects méthodologiques à prendre en considération, en particulier pour l'analyse des données de l'enquête. Chacun des éléments mentionnés est accompagné du numéro de section où l'on peut trouver l'information complète sur le sujet.

Encadré 2.1

Aperçu des principaux aspects méthodologiques

- La population visée par l'EQPPEM est la même que celle ciblée par l'EQDEM 2017. Elle est constituée de l'ensemble des enfants fréquentant la maternelle 5 ans à temps plein dans les écoles francophones et anglophones, publiques et privées (subventionnées ou non), du Québec, à l'exception des enfants fréquentant une école des commissions scolaires crie et Kativik et ceux fréquentant une école relevant du gouvernement fédéral. Les écoles spécialisées et les enfants handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDA)⁴ dans les classes spéciales sont aussi exclus de l'EQPPEM. Les enfants avec un code d'EHDA dans une classe ordinaire où 50 % et moins des enfants sont considérés comme des EHDA sont inclus dans l'enquête, mais sont exclus du fichier d'analyse disponible à l'Infocentre à des fins de cohérence avec les données de l'EQDEM 2017 accessibles (section 2.1.1);
- L'EQPPEM est une enquête-échantillon à portée régionale. Elle repose ainsi sur un échantillon aléatoire stratifié tiré d'une base de sondage contenant les renseignements pertinents sur les enfants visés par l'enquête. Un échantillon d'un peu plus de 18 000 enfants a été préparé pour répondre aux objectifs régionaux et infrarégionaux de l'EQPPEM (section 2.1.2);
- La collecte multimodale (mode Web ou téléphonique) auprès des parents des enfants de l'échantillon a eu lieu entre avril et juillet 2017. Seuls les parents des enfants de l'échantillon pour lesquels un questionnaire rempli par l'enseignante ou l'enseignant (EQDEM 2017) était disponible ont été sollicités pour la collecte de l'EQPPEM (section 2.2);
- Au total, 11 827 questionnaires ont été remplis lors de l'EQPPEM. Le taux de réponse *global* pondéré, mesure de la qualité des données, est d'un peu plus de 66 %; ce taux de réponse prend en considération la non-réponse à l'EQDEM 2017 (à l'échelle du Québec, environ 5 %) et la non-réponse particulière à l'EQPPEM (à l'échelle du Québec, environ 30 %). Le fichier de données accessible à l'Infocentre, qui tient compte de l'exclusion des EHDA se trouvant dans des classes ordinaires de maternelle admissibles à l'enquête, contient les renseignements recueillis auprès des parents de 11 545 enfants (section 2.2.3);

4. Pour obtenir la liste des codes de difficulté accompagnés des définitions, voir MELS, 2007. Pour les détails sur la définition retenue en 2017, consulter Tremblay et Simard (2018).

- Afin qu'il soit possible d'inférer les résultats à la population visée, toutes les estimations publiées par l'ISQ et obtenues sur le portail de l'Infocentre de santé publique sont pondérées (section 2.3.2). Cette façon de faire permet de tenir compte des probabilités d'inclusion variables dans l'échantillon et de la non-réponse plus importante observée chez certains groupes d'enfants. Mentionnons au passage que l'analyse du lien entre les données sur le développement de l'enfant mesuré dans le cadre de l'EQDEM 2017 et les renseignements recueillis par l'EQPPEM, au sujet, par exemple, du parcours préscolaire, repose sur les répondants à l'EQPPEM. Ce sont donc les pondérations de l'EQPPEM qui servent à produire les estimations.
- Afin que soient pris en considération le plan d'échantillonnage et la complexité du mécanisme de réponse, des poids d'autoamorçage sont utilisés aux fins d'estimation de la précision statistique et de production de tests statistiques (section 2.3.4);
- Recommandations pour l'analyse des données de l'EQPPEM et l'interprétation des résultats :
 - Sauf pour les indicateurs associés au revenu des familles des enfants visés par l'EQPPEM, les risques de biais dus à la non-réponse partielle, relativement à la précision des estimations, sont négligeables pour presque tous les indicateurs (section 2.3.3). L'ISQ a traité la non-réponse partielle plus importante pour le revenu des familles en procédant par l'imputation (section 2.3.4); cette façon de faire réduit les risques de biais pour l'indicateur de revenu (en quatre catégories, basé sur la mesure de faible revenu [MFR]), l'indicateur de faible revenu basé sur la MFR et celui basé sur le seuil de faible revenu [SFR]. Par contre, toutes les estimations d'effectifs (populations estimées) doivent prendre en compte l'effet de la non-réponse partielle s'il y a lieu, et ce, peu importe le taux observé (section 2.3.3);
 - Les estimations de l'EQPPEM sont disponibles selon le découpage géographique défini par la région sociosanitaire (RSS) et, pour les RSS de la Capitale-Nationale et de la Chaudière-Appalaches, selon les réseaux locaux de services (RLS) ou regroupements de RLS qui ont été définis par les autorités de ces deux RSS;
 - L'analyse du lien entre l'expérience préscolaire des enfants ou leurs conditions de vie et les différents domaines de développement mesurés dans l'EQDEM devrait se faire à l'échelle provinciale. Les échantillons régionaux n'offrent pas une puissance statistique permettant une analyse à l'échelle des régions;
 - L'interprétation des résultats devrait suivre les recommandations présentées à la partie VI.

2.1. Plan d'échantillonnage

Cette section présente la population visée, la base de sondage à partir de laquelle l'échantillon d'enfants nécessaire à l'enquête a été sélectionné ainsi que la taille et la répartition de l'échantillon préparé pour l'EQPPEM. Les cibles de répondants associées aux objectifs de précision des estimations à produire sont également mentionnées.

2.1.1 Population visée

La population visée par l'EQPPEM est la même que celle visée par l'EQDEM. Celle-ci correspond à l'ensemble des enfants fréquentant la maternelle 5 ans⁵ à temps plein⁶ dans les écoles francophones et anglophones, publiques et privées (subventionnées ou non), du Québec.

Quelques enfants de maternelle sont cependant exclus de la population visée. Les enfants fréquentant une école faisant partie des commissions scolaires crie et Kativik ou étant située dans les régions sociosanitaires du Nunavik (17) et des Terres-Cries-de-la-Baie-James (18) ne font pas partie de la population à l'étude⁷. Il en est de même pour les enfants fréquentant des établissements relevant du gouvernement fédéral dans les réserves autochtones. Ces élèves ont été exclus pour les raisons suivantes : les difficultés à les joindre, les modifications qu'il aurait fallu apporter au questionnaire et les modalités de fréquentation de la maternelle qui peuvent être différentes.

Tous les élèves fréquentant une école spéciale⁸ sont également exclus de l'enquête. À ces exclusions s'ajoutent les enfants qui, selon les fichiers administratifs du MEES ou selon les renseignements obtenus de l'école, ont un code d'élèves handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDAA)⁹ et font partie d'une classe composée de plus de 50 % d'EHDAA. Par contre, les EHDAA fréquentant une classe ordinaire où 50 % et moins des élèves sont considérés comme des EHDAA ont été conservés dans la population visée et sont retenus pour la collecte des données. Ils sont donc pris en compte dans le calcul des taux de réponse présentés. Cependant, ils sont exclus du domaine retenu pour l'analyse présentée dans le rapport de l'enquête produit par l'ISQ, à la suite des recommandations de l'Offord Centre for Child Studies (OCCS).

Somme toute, la population visée par l'EQDEM et l'EQPPEM représente 98 % de l'ensemble des enfants de maternelle en 2016-2017¹⁰. Le tableau 2.1 en offre la distribution selon le secteur d'éducation, en termes d'écoles visées et d'élèves de maternelle 5 ans. La colonne « Domaine d'analyse » correspond au nombre d'enfants de la population visée, à l'exception des EHDAA fréquentant une classe admissible à l'enquête.

-
5. Bien que toutes les commissions scolaires du Québec soient tenues d'offrir la maternelle 5 ans, sa fréquentation demeure facultative. Soulignons toutefois que la presque totalité des enfants de 5 ans y sont inscrits.
 6. Il n'est pas possible d'inscrire un enfant à la maternelle 5 ans à demi-temps, sauf dans de rares cas (MEESR, 2015).
 7. Pour obtenir la liste des codes de difficulté accompagnés des définitions, voir MELS, 2007. Pour les détails sur la définition retenue en 2017, consulter Tremblay et Simard (2018). Un peu plus de 600 enfants étaient inscrits à la maternelle dans ces commissions scolaires en 2016-2017 (renseignements transmis par la Direction de la méthodologie et des études du MEES, fichier d'avril 2017).
 8. Sont considérées comme « écoles spéciales » les écoles des hôpitaux et des centres de réadaptation.
 9. Pour obtenir la liste des codes de difficulté accompagnés des définitions, consulter le *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le développement des enfants à la maternelle 2017* (ISQ, 2018).
 10. Compilation de l'ISQ à partir de renseignements transmis par la Direction de la méthodologie et des études du MEES (fichier d'avril 2017).

Tableau 2.1

Population visée par l'EQDEM et l'EQPPEM 2017 et domaine d'analyse des rapports d'enquête

Unités visées ou analysées	Secteur public	Secteur privé	Population visée	Domaine d'analyse – (EHDAA des classes admissibles exclus)
Écoles	1 748	147	1 895	1 895
Élèves	82 567	5 372	87 939	85 851

2.1.2 Base de sondage

La base de sondage utilisée par l'EQPPEM a été construite à partir de l'information tirée des systèmes administratifs du MEES. L'ISQ a entre autres reçu divers renseignements nominaux sur les commissions scolaires, les écoles¹¹ et les élèves¹². Il faut savoir que les fichiers des systèmes administratifs du MEES sont mis à jour de façon continue pendant l'année scolaire. C'est pourquoi deux extractions ont été faites des banques administratives, soit une en décembre 2016 et une autre en avril 2017. Ce sont les renseignements de l'extraction de décembre 2016 qui ont servi à élaborer la base de sondage de l'enquête. L'information du fichier du MEES extraite en avril 2017, mise à jour avec les renseignements recueillis lors de la collecte de l'EQDEM 2017, a servi à établir de manière définitive la taille de la population visée et a permis d'évaluer la couverture de la base de sondage.

Il faut dire que cette base de sondage comporte de la sous-couverture et de la surcouverture. La sous-couverture de la base est associée au fait que des élèves à la maternelle 5 ans ne se trouvaient pas dans les systèmes administratifs du MEES en décembre 2016, mais y figuraient en avril 2017, ou ont été ajoutés par les enseignants de classes de maternelle 5 ans en 2016-2017 sur la liste des élèves de leurs classes respectives. Quant à la surcouverture de la base, elle est due au fait que certains élèves, présents dans le fichier du MEES de décembre 2016, ne figuraient plus dans le fichier du MEES après la mise à jour d'avril 2017 et n'avaient pas non plus été retrouvés lors de la collecte de l'EQDEM. De plus, certains élèves de la base de sondage peuvent avoir quitté le Québec, s'être retrouvés dans une école spéciale, s'être déplacés vers des régions ou des commissions scolaires exclues de l'enquête ou encore s'être trouvés, en 2016-2017, dans une classe composée de plus de 50 % d'EHDAA (classe non admissible à l'enquête).

Malgré cela, l'utilisation de l'information provenant du MEES dans la construction de la base de sondage garantit une bonne couverture de la population visée et des renseignements de qualité sur les écoles et les enfants. En effet, le fichier du MEES de décembre 2016 a permis d'obtenir une base de sondage comportant une surcouverture¹³ estimée à moins de 2 % et une sous-couverture¹⁴ à peu près équivalente des enfants de la population visée. De plus, la comparaison de l'information du fichier de décembre 2016 du MEES concernant chaque enfant avec celle du fichier mis à jour en avril 2017 a montré des différences marginales. Par ailleurs, les renseignements nécessaires pour joindre les parents des enfants de l'échantillon étaient rarement manquants dans la base de sondage.

11. Comme les coordonnées des commissions scolaires et des écoles, la langue d'enseignement de l'école, le réseau d'enseignement, etc.

12. Comme le nom, le prénom, l'âge, le sexe, la langue maternelle, le code postal, le code d'admissibilité exceptionnelle, etc.

13. La surcouverture correspond au rapport entre le nombre d'enfants de la base de sondage inadmissibles à l'EQPPEM et le nombre d'enfants de la population visée tel qu'établi à la fin de la collecte de l'EQDEM et après la mise à jour du fichier du MEES en avril 2017.

14. On évalue la sous-couverture en comparant le nombre d'enfants présents dans la base de sondage avec le nombre d'enfants de la population visée tel qu'établi à la fin de la collecte de l'EQDEM et après la mise à jour du fichier du MEES en avril 2017.

2.1.3 Taille et répartition de l'échantillon

Avant de détailler l'information concernant la taille et la répartition de l'échantillon de l'EQPPEM, il faut rappeler que cette enquête fait suite à l'EQDEM 2017 et qu'elle tire profit de l'échantillon de l'*Étude pour l'évaluation du programme Passe-Partout* (ÉÉPPP) menée simultanément aux deux autres enquêtes. La combinaison des trois enquêtes a une influence sur les paramètres du plan de sondage de l'EQPPEM, qui est le dernier à être élaboré. Celui-ci doit d'abord tenir compte de l'effet de la combinaison de l'EQPPEM avec l'EQDEM 2017. Puis, la contribution de l'échantillon de l'ÉÉPPP à l'EQPPEM est examinée afin que soient déterminées la taille et la répartition de l'échantillon d'enfants qui s'ajoute à celui de l'ÉÉPPP.

Bien que l'EQDEM soit une enquête par recensement à caractère obligatoire menée auprès des commissions scolaires et des écoles privées, elle comporte tout de même un certain niveau de non-réponse globale. Donc, pour les enfants de l'échantillon de l'EQPPEM, on a établi la taille de l'échantillon de départ pour l'enquête en prévoyant des pertes associées à la non-réponse à l'EQDEM.

Quant à la combinaison de l'EQPPEM avec l'ÉÉPPP, on doit considérer que cette dernière est une enquête comportant un volet d'étude cas-témoin ayant pour objectif principal de mettre en lien le niveau de développement des enfants à la maternelle avec la participation au programme Passe-Partout. Il faut dire que la population visée par l'ÉÉPPP correspond à un peu moins du tiers de celle visée par l'EQPPEM, puisque le programme n'est pas offert dans toutes les écoles. Par ailleurs, le plan de sondage de l'étude comporte une stratification très fine de la base de sondage de manière à assurer le meilleur « pairage » des enfants ayant participé au programme Passe-Partout avec les témoins. De fait, on a construit quelque 342 strates secondaires en considérant l'école ou le regroupement d'écoles fréquentées par les enfants en maternelle 5 ans en 2016-2017 (stratification secondaire). Le tirage des enfants Passe-Partout est fait indépendamment du tirage des enfants témoins; une stratification (principale) des enfants est ainsi d'abord faite selon la participation (ou non) au programme. Le plan d'échantillonnage de l'ÉÉPPP comporte donc 684 strates. La répartition de l'échantillon des témoins entre les strates est choisie de manière à ce que l'atteinte de l'objectif de la comparaison avec les enfants Passe-Partout soit optimisée et qu'une description fiable de la sous-population des enfants qui n'ont pas participé au programme en 2015-2016 soit obtenue. Ce choix optimal pour l'ÉÉPPP rend l'échantillon des enfants témoins moins efficace pour la description de cette sous-population des enfants de maternelle. Pour la combinaison de l'ÉÉPPP à l'EQPPEM, il faut aussi se rappeler que le plan de sondage de l'ÉÉPPP n'est pas conçu pour répondre à des objectifs régionaux. Mentionnons enfin que le nombre de répondants visés pour l'ÉÉPPP est de 4 300 répondants (2 200 enfants ayant participé au programme et 2 100 enfants témoins).

Pour sa part, l'EQPPEM doit fournir des estimations à l'échelle des régions administratives (RA). En plus, pour les régions sociosanitaires (RSS) de la Capitale-Nationale et de la Chaudière-Appalaches, des estimations sont désirées pour des regroupements de réseaux locaux de services (RLS)¹⁵. Le plan de sondage a donc été construit de manière à ce qu'il soit possible :

- d'estimer un phénomène présent chez environ 10 % des enfants avec une bonne précision relative, c'est-à-dire qu'un coefficient de variation (CV)¹⁶ d'au plus 15 % dans chacune des dix-sept RA visées par l'enquête est obtenu;

15. Les besoins statistiques de l'enquête se situent au niveau du lieu de résidence des enfants. Par contre, le plan d'échantillonnage comporte une stratification géographique en fonction de l'adresse de l'école. Cette décision a été prise en raison du fait que la qualité des renseignements concernant les adresses des enfants dans la base de sondage n'était pas uniforme pour tous les enfants (ex. adresse parfois manquante ou case postale seulement). En outre, une vérification sommaire a montré une excellente correspondance entre la région administrative où se situe l'école et celle où se situe la résidence des enfants : dans moins de 2 % des cas, on a observé une différence entre les deux renseignements.

16. Voir la section 2.3.5 pour la définition du coefficient de variation.

- d'estimer un phénomène présent chez environ 10 % des enfants avec une bonne précision relative, c'est-à-dire qu'un coefficient de variation (CV) d'au plus 15 % dans chacun des regroupements de RLS définis pour les RSS de la Capitale-Nationale et de la Chaudière-Appalaches est obtenu.

Le nombre de répondants nécessaires à l'atteinte de ces objectifs de précision a été fixé à 440 pour chaque RA et pour chacun des cinq regroupements de RLS des RSS de la Capitale-Nationale et de la Chaudière-Appalaches pour lesquels des estimations sont requises. Pour qu'il soit possible d'offrir des estimations provinciales ayant une meilleure précision statistique, la cible minimale de répondants est fixée à 990 pour la RA de Montréal et à 770 pour la Montérégie. En tout, le nombre total de répondants nécessaires à l'atteinte des objectifs de l'EQPPEM s'élève donc à près de 9 700.

Comme indiqué précédemment, une partie des besoins statistiques de l'EQPPEM sont comblés par l'utilisation des répondants de l'ÉÉPPP¹⁷. En tenant compte de l'efficacité de l'échantillon de cette étude pour les objectifs de l'EQPPEM, l'ISQ a établi qu'environ 6 400 répondants seraient nécessaires en plus de ceux de l'ÉÉPPP (4 300 répondants). Finalement, en tout, le nombre de répondants visés pour l'EQPPEM avoisine donc 10 700. Grâce à la combinaison de l'EQPPEM avec l'ÉÉPPP, une économie d'environ 3 300 répondants a pu être réalisée.

Afin que la cible de répondants déterminée soit atteinte, la taille d'échantillon préparé pour l'EQPPEM est ajustée de manière à ce qu'elle tienne compte du taux d'admissibilité prévu et des taux de réponse attendus pour les questionnaires destinés à l'enseignant (EQDEM) et aux parents (EQPPEM). Un échantillon total d'un peu plus de 18 000 enfants a été sélectionné pour l'EQPPEM. Le tableau 2.2 présente les hypothèses retenues pour les différents facteurs d'ajustement de la taille d'échantillon et le tableau 2.3 présente la composition de l'échantillon préparé ainsi que le nombre de répondants attendu par RA et regroupements de RLS pour les RSS de la Capitale-Nationale et de la Chaudière-Appalaches.

Tableau 2.2

Hypothèses sur le taux d'admissibilité et les taux de réponse pour le calcul de la taille d'échantillon initiale de l'EQPPEM

Taux d'admissibilité attendu	Taux de réponse attendu EQDEM (enseignants)	Taux de réponse attendu EQPPEM (parents)	Taille de l'échantillon
99 %	90 %	66 %	18 062

17. Rappelons que l'ÉÉPPP ne couvre pas l'ensemble du territoire parce que le programme Passe-Partout n'est pas offert partout.

Tableau 2.3

Taille d'échantillon et nombre de répondants attendu selon le découpage géographique, Québec, 2017

Région administrative/réseau local de services (adresse de l'école)	Taille d'échantillon	Nombre de répondants attendu ¹
01 Bas-Saint-Laurent	749	450
02 Saguenay–Lac-Saint-Jean	936	555
03 Capitale-Nationale		
RLS Québec-Sud	789	450
RLS Québec-Nord	860	515
Autres RLS ²	165	100
04 Mauricie	791	460
05 Estrie	893	530
06 Montréal	1 764	995
07 Outaouais	803	460
08 Abitibi-Témiscamingue	723	450
09 Côte-Nord	775	450
10 Nord-du-Québec	171	100
11 Gaspésie–Îles-de-la-Madeleine	686	405
12 Chaudière-Appalaches		
RLS Alphonse-Desjardins	971	580
RLS de Beauce	794	460
Autres RLS	756	450
13 Laval	787	445
14 Lanaudière	873	510
15 Laurentides	848	490
16 Montérégie	2 147	1 360
17 Centre-du-Québec	781	485
Ensemble du Québec	18 062	10 700

1. Le nombre de répondants attendu est parfois plus élevé que les cibles de départ en raison de la combinaison avec l'ÉÉPPP qui n'est pas efficace. C'est le cas par exemple pour l'échantillon de la Montérégie. Pour la région 10, un recensement des enfants de la population visée mène à 100 répondants attendus, si l'on considère les hypothèses retenues pour les taux d'admissibilité et de réponse.

2. Pour ce regroupement de RLS, il n'y a pas d'objectif particulier de précision.

Source : Institut de la statistique du Québec, *Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017*.

2.2 Collecte des données

2.2.1 Prétest

Aucun prétest n'était prévu pour l'enquête. Par contre, le questionnaire entier a été vérifié auprès d'un échantillon de parents de 100 enfants entre le 3 et le 5 avril 2017. Cette vérification a permis de valider la compréhension des questions après la réalisation de 58 entrevues par téléphone. À la suite de cette opération, quelques modifications très mineures ont été apportées au questionnaire de l'enquête. Il faut dire que les changements apportés au questionnaire sont marginaux, de sorte que les 58 entrevues réalisées sont conservées dans le fichier de données de l'enquête et servent aux analyses produites dans le rapport.

2.2.2 Stratégies et opérations de collecte

La collecte de données de l'EQPPEM a été menée auprès des parents une fois la collecte de l'EQDEM pour leur enfant réalisée auprès des enseignants. En effet, seuls les enfants pour lesquels un questionnaire avait été rempli pour l'EQDEM étaient utilisés pour la collecte auprès des parents.

Il s'agissait d'une collecte multimodale utilisant les modes Web et téléphonique. Les parents avaient ainsi la possibilité de répondre au questionnaire informatisé auto-administré via une page Web, hébergée sur le site sécurisé de l'ISQ (mode IWAO – interview Web assistée par ordinateur), ou encore de prendre part à une entrevue téléphonique assistée par ordinateur menée par un intervieweur (mode ITAO – interview téléphonique assistée par ordinateur). Comme pour les autres enquêtes réalisées à l'ISQ utilisant une collecte multimodale, le mode Web était d'abord offert aux parents des enfants de l'échantillon de l'EQPPEM; le mode téléphonique était utilisé lorsque les parents signalaient à l'ISQ ne pas vouloir remplir le questionnaire en ligne ou être dans l'impossibilité de le faire, ou après des relances téléphoniques ultimes infructueuses effectuées auprès des non-répondants. Avec cette approche, mentionnons qu'environ 7 répondants sur 10 (73 %) ont rempli le questionnaire en mode Web.

Les étapes de la stratégie de collecte adoptée pour l'EQPPEM sont les suivantes : envoi d'une lettre d'information aux parents des enfants pour lesquels un questionnaire a été rempli pour l'EQDEM pour les inviter à participer à l'EQPPEM et les inciter à répondre par Internet en utilisant l'identifiant attribué par l'ISQ¹⁸. Par la suite, une première lettre de relance visant à réitérer aux parents non répondants l'invitation à remplir le questionnaire via Internet a été envoyée. Après une période de temps jugée suffisante, des rappels téléphoniques ont débuté; si la personne mentionnait vouloir toujours remplir le questionnaire sur le Web, on demandait le courriel pour effectuer les relances suivantes. Sinon, les parents étaient invités à réaliser l'entrevue au téléphone s'ils étaient disponibles.

La collecte de donnée auprès des parents s'est échelonnée sur près de 12 semaines, soit du 19 avril au 3 juillet 2017.

Avant la collecte, les intervieweurs ont reçu une formation spécifique à l'enquête pour se familiariser avec le contexte, les objectifs, le contenu du questionnaire, les codes de résultat et le questionnaire informatisé associés à l'enquête. Les intervieweurs étaient chargés de fournir une assistance aux participants désirant remplir le questionnaire en ligne, procédaient à des relances téléphoniques ou réalisaient des entrevues téléphoniques.

Le nombre de questionnaires remplis a été suivi de près, globalement et pour chacun des découpages géographiques du plan de sondage. En outre, plusieurs stratégies ont été mises en place pour que la collaboration des parents des enfants sélectionnés soit favorisée et pour que la gestion de la collecte se trouve facilitée. Par exemple, en tout temps, les parents disposaient d'un numéro de téléphone et d'une adresse courriel réservés à l'enquête.

Quant aux parents pour lesquels le MEES n'avait pas de numéro de téléphone (environ 5 % des personnes échantillonnées), des recherches ont été effectuées afin que des numéros valides soient trouvés. Les parents pour qui le numéro de téléphone était resté inconnu après la recherche ont reçu une lettre leur demandant de contacter l'ISQ.

Notons que l'on incluait les parents biologiques, adoptifs, les tuteurs et les tutrices parmi les répondants clés potentiels pour recueillir l'information concernant l'enfant échantillonné.

18. Ceux n'ayant pas accès à Internet ou ne désirant pas faire l'entrevue en mode Web avaient dès lors la possibilité de contacter l'ISQ par téléphone pour répondre au questionnaire administré par un intervieweur.

Soulignons enfin que l'on a tout mis en place pour alléger le fardeau de réponse des parents. Par exemple, les questionnaires informatisés Web et téléphonique comportaient des passages complexes préprogrammés pour que le parent ne voie que les questions qui correspondent à sa situation. Plusieurs tests ont été réalisés avant le début de la collecte pour que l'on s'assure du bon fonctionnement des applications informatiques IWAO et ITAO. Le questionnaire aux parents de l'EQPPEM¹⁹ comportait sept sections (tableau 2.4).

Tableau 2.4

Contenu du questionnaire adressé aux parents des enfants sélectionnés pour l'EQPPEM

Section A – Description de la famille
Section E – Parcours préscolaire en milieu de garde
Section F – Autres activités ou programmes éducatifs (y compris les activités de lecture et de numérotation en contexte familial)
Section G – Santé de l'enfant
Section H – Contexte familial (santé des parents, stress familial et soutien social)
Section I – Résidence et perception de la sécurité du quartier
Section J – Caractéristiques socioéconomiques

Note : Les sections B à D du questionnaire étaient exclusivement destinées aux parents des enfants de l'échantillon de l'ÉÉPPP.

Toutes les questions ont été révisées par une linguiste et traduites en anglais pour qu'il soit possible de réaliser des entrevues dans les deux langues. Pour le contenu particulier de l'EQPPEM, la durée moyenne du questionnaire lors des appels téléphoniques a été de 16 minutes. La durée du questionnaire Web n'a pas pu être déterminée, puisqu'il est impossible de considérer les pauses; seul le temps de connexion peut être comptabilisé. Étant donné cette limite importante, la durée moyenne des questionnaires Web n'a pas été calculée.

2.2.3 Résultats de collecte

À l'instar des taux de réponse généralement présentés par l'ISQ, les taux présentés pour l'EQPPEM 2017 sont pondérés. Ils tiennent compte, entre autres facteurs, de la non-proportionnalité régionale et locale de l'échantillon de l'enquête.

Concernant les hypothèses de départ de l'EQPPEM, le taux de réponse obtenu à l'EQDEM (enseignants) est supérieur à celui anticipé. Le caractère obligatoire de l'EQDEM a certainement favorisé ce bon résultat. La non-réponse est en général associée au fait que les enseignants ne peuvent fournir de réponse pour les élèves qui n'étaient pas dans leur classe depuis au moins un mois ou encore au fait que certains parents ont refusé de faire participer leur enfant à l'EQDEM. Alors que le scénario initial prévoyait l'envoi d'un peu moins de 16 100 unités de l'échantillon de départ à la collecte auprès des parents, ce sont plutôt les parents de près de 17 000 enfants qui ont reçu l'invitation à répondre au questionnaire sur le parcours préscolaire. En fait, le taux d'admissibilité combiné au taux de réponse à l'EQDEM a atteint 93,7 % au lieu du 89,1 % anticipé.

Au total, ce sont 11 827 parents d'enfants de maternelle qui ont répondu à l'EQPPEM, ce qui correspond à un taux de réponse pondéré global de 66,3 %. Ce taux prend en considération la non-réponse à l'EQDEM et la non-réponse à l'EQPPEM.

19. Le questionnaire de l'EQPPEM est disponible sur le site Web de l'ISQ à l'adresse www.stat.gouv.qc.ca/.

Le tableau 2.6 présente le nombre de répondants et le taux de réponse pondéré obtenu par RSS ainsi que par RLS (regroupement de RLS) pour les deux régions où un découpage géographique plus fin était requis; le lieu de résidence de l'enfant est utilisé pour ce tableau. Le nombre de répondants présents dans la base de données disponible à l'Infocentre est inférieur à 11 827; cette base de données compte plutôt 11 545 enfants puisque 282 enfants répondants, qui correspondent à des EHDAA fréquentant des classes ordinaires où moins de 50 % des élèves sont considérés comme des EHDAA, sont exclus de l'ensemble de données analysées. La dernière colonne du tableau 2.5 présente la distribution de cet ensemble selon la région sociosanitaire, le RLS ou le regroupement de RLS choisi.

Tableau 2.5

Nombre de répondants et taux de réponse pondéré à l'enquête, nombre de répondants dans la base de données déposée à l'Infocentre selon le découpage géographique de l'adresse de résidence de l'enfant, Québec, 2017

Région sociosanitaire (RSS)/ RLS ou regroupement de RLS	Nombre de répondants n (EHDAA inclus)	Taux de réponse pondéré ¹ (%) (EHDAA inclus)	Nombre de répondants n (EHDAA exclus)
01 Bas-Saint-Laurent	551	73,4	538
02 Saguenay-Lac-Saint-Jean	658	70,6	644
03 Capitale-Nationale	1 225	68,2	1 192
RLS Québec-Sud	485	68,9	465
RLS Québec-Nord	634	69,1	623
Autres RLS	106	62,2	104
04 Mauricie et Centre-du-Québec	1 010	65,5	985
05 Estrie	907	61,6	885
06 Montréal	1 023	61,2	1 014
07 Outaouais	503	63,4	483
08 Abitibi-Témiscamingue	463	64,9	448
09 Côte-Nord	470	61,6	457
10 Nord-du-Québec	111	63,6	108
11 Gaspésie-Îles-de-la-Madeleine	427	63,2	408
12 Chaudière-Appalaches	1 824	72,8	1 780
RLS Alphonse-Desjardins	734	73,0	712
RLS de Beauce	525	72,0	517
Autres RLS	565	72,8	551
13 Laval	491	60,3	478
14 Lanaudière	562	64,0	546
15 Laurentides	543	65,9	539
16 Montérégie	1 059	67,6	1 040
Ensemble du Québec	11 827	66,3	11 545

1. Ce taux de réponse tient compte de la réponse à l'EQDEM et à l'EQPPEM; il s'agit du rapport entre le nombre pondéré de répondants et le nombre pondéré d'enfants admissibles.

Source : Institut de la statistique du Québec, *Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017*.

2.3 Traitement des données

Cette section fournit de l'information sur le traitement des données de l'EQPPEM. On y présente d'abord les éléments de validation des données de l'enquête, la méthode de pondération retenue et l'analyse de la non-réponse partielle. Ensuite, on traite de certains aspects de l'analyse, en particulier la présentation des estimations, les tests statistiques et la précision des estimations.

2.3.1 Validation et saisie

Le recours à la collecte assistée par ordinateur, que ce soit Web ou téléphonique, permet une saisie automatique des réponses fournies par les parents. La très grande majorité des questions sont en outre autocodées, c'est-à-dire que l'intervieweur ou le parent n'a qu'à cocher le choix de réponse approprié. Pour certaines questions comportant plusieurs possibilités de réponse, des menus déroulants offrant des choix de réponse prédéterminés ont été conçus.

De plus, les collectes téléphoniques et Web étant entièrement informatisées, plusieurs validations de base sont donc réalisées au fur et à mesure que le questionnaire est rempli, notamment la vérification du respect des choix de réponse pour les questions fermées ou celle de l'adéquation des sauts de section pour les blocs de questions ne s'adressant qu'à un sous-groupe de répondants. Bien que les questionnaires informatisés aient été amplement testés avant le début de la collecte, tous les codes de réponses et les sauts de section, de même que la cohérence des réponses d'un même questionnaire, ont quand même été validés a posteriori.

Par ailleurs, plusieurs questions comportent un choix de réponse ouvert représenté par la catégorie « Autre ». C'est le cas par exemple de la question QJ4, soit « Quel est le plus haut diplôme que vous avez obtenu? » ou encore de la question Q3, soit « Quel est votre lien avec "prénom"? ». Un certain nombre de cas « Autre » a été reclassé à la suite de l'examen de la réponse fournie par le parent. En effet, malgré les consignes données aux parents, certains ont mal codé leur réponse. Par exemple, il était indiqué que si le parent avait une garderie de type familial à la maison, il devait cocher « Dans la maison de quelqu'un » et non « Autre » pour les questions QE3A, QE4A, QE5A, QE6A et QE7A. Ces cas « Autres » ont donc été reclassés dans cette catégorie.

Enfin, les réponses fournies par les parents aux questions sur la fréquentation d'une maternelle 4 ans (QF1) ou sur la participation au programme Passe-Partout (QF2) en 2015-2016 dans une école publique ont été comparées aux informations provenant du fichier administratif du MEES. Pour près de 10 % des enfants, il y avait incohérence entre l'information provenant du système scolaire et celle donnée par le parent. Après examen, les informations de la base de sondage confirmées par les enseignants seraient plus exactes que les réponses fournies par les parents. L'ordre des questions présentées aux parents, le manque de description précise des programmes ou la méconnaissance des parents auraient eu comme effet de créer une surdéclaration de la fréquentation de la maternelle 4 ans temps partiel et temps plein. Ainsi, les résultats des questions QF1 et QF2 ont été retirés du fichier de données pour tous les répondants; une variable dérivée a été construite à partir de la base de sondage et des renseignements fournis par les enseignants, ce qui a permis de pallier ce retrait.

2.3.2 Pondération

La pondération consiste à attribuer à chaque enfant pour qui un questionnaire a été rempli par le parent, une valeur, c'est-à-dire un poids statistique, qui correspond au nombre d'enfants qu'il « représente » dans la population visée. La pondération est essentielle pour qu'une inférence adéquate soit faite à cette population, puisque cette dernière n'a pas été observée dans sa totalité. En effet, l'EQPPEM est menée à l'aide d'un échantillon et, comme toutes les enquêtes, elle n'est pas exempte de non-réponse. Le poids statistique doit donc tenir compte tant de la probabilité de sélection des enfants, prédéterminée par le plan d'échantillonnage de l'EQPPEM, que de la non-réponse à l'EQDEM et à l'EQPPEM des enfants échantillonnés.

Il faut dire que dans l'EQPPEM, les objectifs régionaux et infrarégionaux de même que la combinaison avec l'échantillon de l'ÉÉPPP entraînent une variation considérable des probabilités de sélection et font que la répartition de l'échantillon selon les RA et les RLS ne représente pas bien celle de la population visée. D'où la nécessité de tenir compte de la pondération pour la production des estimations.

De plus, la non-réponse a touché davantage certains groupes d'enfants de l'échantillon, ce qui a nécessité un ajustement de la pondération visant à réduire un biais potentiel de non-réponse dans les estimations. Cette non-réponse peut survenir à différents moments du processus de collecte puisque les commissions scolaires et les écoles privées peuvent refuser le jumelage des données recueillies auprès des enseignants avec celles fournies par les parents; de leur côté, les parents peuvent également refuser ce jumelage ou simplement décider de ne pas participer à l'EQDEM ou à l'EQPPEM. Il s'agit dans tous ces cas d'une non-réponse à l'EQPPEM.

Enfin, la pondération des répondants de l'EQPPEM comporte une dernière étape qui permet de corriger les défauts de couverture de la base de sondage dont il a été question à la section 1.2 et d'améliorer la qualité des estimations produites à l'échelle des régions de résidence des enfants et de celles concernant le niveau de développement des enfants à la maternelle en 2016-2017.

La stratégie de pondération qui a été élaborée par l'ISQ pour l'EQPPEM tient ainsi compte de la probabilité de sélection des enfants des échantillons constitués pour l'enquête, de l'ajustement de la non-réponse totale à l'EQDEM et à l'EQPPEM et d'un ajustement à certains totaux de population connus ou estimés concernant les enfants de maternelle 5 ans de la population visée.

La première étape consiste à calculer le poids de départ de chaque enfant qui correspond à l'inverse de sa probabilité de sélection. Chacun des 18 062 enfants échantillonnés pour l'EQPPEM reçoit donc un poids équivalent à l'inverse de sa probabilité de sélection dans la strate à laquelle il appartient. Les pondérations de départ des enfants de l'échantillon varient donc d'une strate à l'autre²⁰. Pour chacune, le poids de base correspond au résultat de la division du nombre d'enfants dans la base de sondage par le nombre d'enfants sélectionnés dans l'échantillon.

20. En tout, 725 strates ont été construites pour l'EQPPEM. Elles correspondent au croisement des 684 strates utilisées pour l'ÉÉPPP avec les découpages géographiques particuliers de l'EQPPEM, c.-à-d. 704 strates; s'ajoutent à ce nombre 21 strates pour les échantillons régionaux et infrarégionaux sélectionnés en plus de l'échantillon de l'ÉÉPPP.

En ce qui a trait à l'EQDEM, bien qu'il y ait une légère non-réponse à l'échelle des écoles et des enseignants, elle est principalement observée à l'échelle des enfants et correspond au refus de participation de la part des parents. Dans quelques rares cas, l'école a refusé le jumelage des renseignements fournis par l'enseignant avec ceux de l'EQPPEM. Pour la pondération, il a été convenu de traiter cette non-réponse en une seule étape d'ajustement étant donné les taux de réponse très élevés obtenus (environ 95 %). Ainsi, les enfants « répondants », c'est-à-dire ceux pour lesquels un questionnaire rempli dans le cadre de l'EQDEM est disponible pour un jumelage avec celui de l'EQPPEM, sont pondérés et « représentent » ainsi ceux pour qui ce n'est pas le cas. Pour ce faire, on effectue une analyse pour définir les variables ou caractéristiques selon lesquelles la probabilité d'avoir un questionnaire « répondant » pourrait varier. On utilise la méthode du score²¹ pour déterminer quels sont les groupes homogènes de pondération, c'est-à-dire des sous-groupes de l'échantillon à l'intérieur desquels la propension à répondre est assez uniforme²². L'ajustement s'effectue ensuite à l'intérieur de chacun de ces groupes : le poids de base des enfants admissibles à l'EQPPEM pour lesquels un questionnaire EQDEM est disponible est multiplié par l'inverse du taux de réponse (pondéré) dans le groupe homogène de pondération. La première étape de l'ajustement de la non-réponse est ainsi terminée.

Ensuite, un ajustement pour la non-réponse à l'EQPPEM est effectué : on emploie la méthode du score de nouveau pour constituer des groupes d'enfants homogènes en ce qui a trait à la propension à répondre au questionnaire adressé aux parents. Les probabilités de répondre à ce questionnaire sont encore ici estimées au moyen d'un modèle de régression logistique²³.

Enfin, la dernière étape de la pondération est la poststratification. Cet ajustement permet de s'assurer que la somme des poids de tous les répondants de l'enquête correspond bien à la population des enfants de maternelle visée par l'EQPPEM, notamment pour certaines caractéristiques associées aux mesures d'intérêt. Comme mentionné en introduction du rapport, l'analyse de la relation entre le parcours préscolaire des enfants, leurs conditions de vie et les différents aspects de leur développement à la maternelle 5 ans fait partie des objectifs de l'EQPPEM. L'EQDEM, qui est basée sur un recensement des enfants inscrits en maternelle en 2016-2017, offre des estimations très précises du niveau de développement de ces enfants à leur entrée à l'école. En effet, des totaux estimés peuvent être produits à partir de l'information obtenue des enseignants au sujet des 83 335 enfants répondants de l'EQDEM, qui représentent près de 95 % de la population visée. Ainsi, les estimations de l'EQDEM résultant du croisement d'un indicateur de développement divisé en trois catégories (vulnérable dans aucun domaine, vulnérable dans un domaine et vulnérable dans deux domaines ou plus²⁴) avec une variable de regroupements de régions/RLS de la résidence des enfants ont servi à la poststratification des répondants de l'EQPPEM.

2.3.3 Analyse de la non-réponse partielle

Outre la non-réponse totale compensée par la pondération, il existe un autre type de non-réponse, soit celle associée à une question précise, appelée non-réponse partielle. Le taux de non-réponse partielle se définit comme le rapport entre le nombre pondéré d'enfants pour lesquels aucune réponse n'a été fournie à la question (ou bien la réponse fournie à la question est invalide) et le nombre pondéré d'enfants admissibles à la question.

-
21. Pour plus de renseignements sur la méthode du score, veuillez consulter Haziza et Beaumont, 2007 ou Eltinge et Yansaneh, 1997.
 22. Les variables retenues pour la construction des groupes homogènes sont : le nombre de parents inscrits dans le fichier du MEES, le lieu de naissance des parents (Québec ou hors Québec), le réseau de l'école des enfants (privé, public) et la combinaison de la région administrative et du RLS/regroupement de RLS de la résidence des enfants.
 23. Le modèle retenu comporte plusieurs indicateurs : le lieu de résidence de l'enfant tel que défini par la région administrative ou encore, pour certains, par le réseau local de services; le nombre de parents inscrits dans le fichier du MEES; la langue maternelle de l'enfant; un indicateur calculé à partir du rang de l'indice de défavorisation associé à l'école (IMSE); et un indicateur de développement de l'enfant divisé en trois catégories (vulnérable dans aucun domaine, vulnérable dans un domaine et vulnérable dans deux domaines ou plus) produit à partir de l'information fournie par l'enseignant.
 24. Pour plus d'information sur le concept de vulnérabilité, sur les domaines de développement et sur les indicateurs qui y sont associés, le lecteur peut consulter le rapport de l'EQDEM 2017, disponible sur le site Web de l'ISQ à l'adresse suivante : <http://www.stat.gouv.qc.ca/statistiques/sante/enfants-ados/developpement-enfants-maternelle-2017.html>.

La non-réponse partielle peut entraîner des biais dans les estimations si les non-répondants partiels présentent des caractéristiques différentes de celles des répondants. Plus la proportion de non-réponse partielle est élevée, plus le risque de biais est grand. En général, lorsqu'on analyse des données de l'EQPPEM, il est raisonnable de faire l'hypothèse que, pour une non-réponse partielle inférieure à 5 % dans l'ensemble du Québec, les risques de biais sont plutôt faibles. De façon générale, la non-réponse partielle est peu élevée dans l'EQPPEM; le risque que des biais importants soient associés à la non-réponse partielle est donc assez faible. Compte tenu de la précision des estimations, l'impact de la non-réponse partielle sur les conclusions des analyses est en général négligeable.

L'ISQ a procédé à un examen particulier des quelques questions ou indicateurs présentant des taux de non-réponse partielle supérieur à 5 % à l'échelle du Québec. Les prochains paragraphes décrivent les observations qui en découlent.

L'examen des taux de non-réponse partielle a montré qu'une seule question dépasse le seuil critique de 5 %, soit celle portant sur les raisons principales pour ne pas avoir fait garder son enfant de façon régulière avant la maternelle (QE11), pour laquelle le taux de non-réponse partielle est de 6,9 %. Un examen des caractéristiques des non-répondants comparativement à celles des répondants a été réalisé afin que les risques potentiels de biais dû à cette non-réponse soient évalués. Cet examen révèle que, comparativement aux répondants, les parents qui n'ont pas répondu à cette question sont proportionnellement surreprésentés chez les parents d'enfants nés à l'extérieur du Canada, chez les parents qui ont l'anglais (avec ou sans autres langues, sauf le français) comme langue maternelle ou qui résident dans la RMR de Montréal (données non présentées). Les non-répondants sont par ailleurs moins nombreux en proportion que les répondants chez les bénéficiaires de l'assurance-emploi et chez les résidents de petites villes ou de milieux ruraux. Ainsi, l'EQPPEM sous-estimerait très légèrement la proportion d'enfants qui n'ont pas été gardés parce que l'un de ses parents a fait le choix de demeurer à la maison et surestimerait un peu la proportion d'enfants n'ayant pas été gardés parce qu'il n'y avait pas de place disponible en service de garde. Dans les deux cas, compte-tenu de la précision des estimations, le biais dû à la non-réponse partielle n'a aucune incidence sur les conclusions des analyses effectuées.

Quelques indicateurs rendant compte de l'intensité de l'utilisation des services éducatifs présentent également des taux de non-réponse partielle avoisinant le seuil critique de 5 %. Il s'agit de la durée cumulative en services de garde éducatifs (CUMUL_SGER), de l'âge au début de l'utilisation d'un service éducatif (AG_DEB_SERV_EDU_R) et de la dose d'exposition moyenne en services éducatifs (DOSE_EDU_R2). L'examen des biais potentiels liés à la non-réponse partielle a montré qu'il y a proportionnellement plus de non-répondants parmi les familles à faible revenu et parmi les enfants vivant au sein de familles monoparentales ou recomposées. Ce faisant, l'EQPPEM sous-estimerait très légèrement la proportion d'enfants les moins intensément exposés aux services éducatifs, c'est-à-dire ceux qui ont bénéficié de tels services 12 mois ou moins, ceux qui ont commencé à les utiliser à partir de 4 ans seulement ou ceux pour qui la dose moyenne d'exposition à ces services est jugée faible (24 mois ou moins). Dans chaque cas, le biais lié à la non-réponse partielle serait inférieur à un point de pourcentage et n'aurait aucune incidence sur les conclusions des analyses, une fois la précision des estimations prise en compte.

Par ailleurs, l'examen des taux de non-réponse partielle pour certains indicateurs montre que certains d'entre eux pourraient présenter des biais, bien que les taux de non-réponse partielle soient un peu inférieurs au seuil de 5 %. Les indicateurs dichotomiques de garde régulière en services de garde éducatifs (GARD_SGE) et d'utilisation d'un service éducatif (SERVICE_EDU) pendant au moins 3 mois, de même que ceux présentant le profil longitudinal de modes de garde ou de services éducatifs utilisés entre 0 et 5 ans (PROFIL_LONG et PROFIL_EDU2_8CAT) comportent tous un peu moins de 5 % de non-réponse partielle. Cependant, puisque tous les non-répondants partiels sont tous des enfants gardés, ces indicateurs présentent une surestimation de la proportion d'enfants non gardés. On recommande donc aux analystes d'utiliser l'indicateur dichotomique simple de garde régulière pendant au moins 3 mois (QE1R) pour produire cette estimation (7,7 % selon cet indicateur c. 8,1 % selon PROFIL_LONG). Il importe aussi de

mentionner que ces indicateurs sont tout de même appropriés pour la comparaison des proportions d'enfants selon le mode de garde ou le service éducatif utilisé ou encore pour l'examen comparatif des caractéristiques des groupes d'enfants définis par les différentes valeurs des indicateurs. Par exemple, l'indicateur de profil longitudinal des modes de garde permet de comparer les caractéristiques des enfants ayant fréquenté exclusivement des CPE entre 0 et 5 ans avec celles des enfants qui n'ont fréquenté aucun service de garde.

Par ailleurs, pour les populations estimées (Pe), aussi appelées effectifs, un taux de non-réponse partielle, aussi minime soit-il, entraîne un biais équivalent dans l'estimation produite. À l'Infocentre, toutes les estimations de populations ont donc été produites avec une correction pour la non-réponse partielle.

Enfin, les indicateurs utilisant le revenu familial avant impôt présentent une non-réponse partielle importante qui a été traitée par l'ISQ. La section qui suit détaille ce traitement.

2.3.4 Imputation du revenu familial

Lorsque le risque de biais associé à la non-réponse partielle est jugé préoccupant, l'imputation des données manquantes est une solution souvent mise de l'avant. Pour une variable présentant de la non-réponse partielle, l'imputation consiste à attribuer une valeur à un individu en remplacement d'une donnée manquante ou invalide au moyen des renseignements disponibles à son sujet²⁵.

Pour l'EQPPEM, l'ISQ a procédé à l'imputation des données des non-répondants partiels pour le revenu familial, lequel sert à la construction de trois indicateurs :

- l'indicateur de revenu en quatre catégories basé sur la mesure de faible revenu (MFR);
- l'indicateur de faible revenu basé sur la mesure de faible revenu du ménage (MFR);
- l'indicateur de faible revenu basé sur le seuil de faible revenu du ménage (SFR).

Le revenu familial est établi à partir des questions sur le revenu total provenant de toutes les sources, avant impôt et autres retenues, de tous les membres du ménage pour les 12 derniers mois (QJ1 et QJ2 et sous-questions QJ1 [A à E] et QJ2 [A à E]). Le répondant avait la possibilité de fournir cette information soit sous forme de valeur exacte, bien qu'approximative, soit sous forme de catégorie en indiquant la tranche de revenu. Parfois, la déclaration sous forme de catégorie de revenus n'est pas assez précise pour permettre l'attribution d'une valeur aux trois indicateurs d'intérêt. Lorsque ces cas sont ajoutés à ceux pour lesquels aucune réponse n'a été fournie au sujet du revenu du ménage, le taux pondéré de non-réponse partielle atteint 6,84 %.

L'imputation déductive ou par donneur (« Hot Deck ») a permis de réduire à 1,82 % la non-réponse partielle associée aux indicateurs de revenu, les cas pour lesquels aucune information sur le revenu n'est disponible étant demeurés « non-répondants » pour les trois indicateurs calculés à partir du revenu. Grâce à l'imputation du revenu, les analyses utilisant l'un ou l'autre de ces indicateurs comme variables de croisement auront ainsi une meilleure assise.

Le lecteur est invité à consulter le rapport de l'EQPPEM (Lavoie et autres, 2018) pour obtenir des renseignements sur la construction des indicateurs de revenu basés sur la MFR (voir aussi Statistique Canada, 2016). Quant à l'indicateur de faible revenu basé sur le seuil de faible revenu (SFR), il consiste essentiellement à comparer le revenu familial du ménage aux seuils SFR, qui « sont des limites de revenu en deçà duquel une famille est susceptible de consacrer une part plus importante de son revenu à l'achat de nécessités comme la nourriture, le logement et l'habillement qu'une famille

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

moyenne » (Statistique Canada, 2014). Il faut dire que les seuils SFR tiennent compte de la taille du ménage et de la communauté au sein de laquelle vit le ménage.

La section 4.2.1 aborde certaines limites associées à l'utilisation des deux indicateurs de faible revenu disponibles à l'Infocentre comme variables de croisement.

2.3.5 Précision des estimations

La plupart des analyses effectuées à partir des données de l'EQPPEM sont descriptives et bivariées. On estime la proportion des enfants qui possèdent une caractéristique en divisant la somme des poids des enfants possédant cette caractéristique par la somme des poids de l'ensemble des enfants pour lesquels cette caractéristique a été mesurée, qu'ils la possèdent ou non.

Un exemple de calcul pour l'estimation d'un effectif (P_e), c'est-à-dire l'estimation du nombre d'enfants qui possèdent une caractéristique, est présenté à l'annexe 3 du guide méthodologique de l'édition 2012 de l'EQDEM. On procède de la même façon pour l'estimation d'un effectif dans l'EQPPEM.

Dans toute enquête statistique, les estimations produites comportent des erreurs dues à l'échantillonnage en raison du fait que seule une partie des unités de la population est sélectionnée pour participer à l'enquête et qu'il y a de la non-réponse globale. Ces erreurs se répercutent sur la précision des estimations produites. La complexité du plan d'échantillonnage, qui n'est pas un plan aléatoire simple, et la non-réponse globale à l'EQDEM et à l'EQPPEM influencent la précision des estimations. Il est nécessaire de mesurer la précision de chaque estimation (proportion ou moyenne) et d'en tenir compte dans l'interprétation des résultats inférés à la population visée.

Pour l'EQPPEM, l'estimation de la variance et les tests statistiques sont effectués à l'aide de poids d'autoamorçage, aussi appelés « poids *bootstrap* ». Une série de 500 poids d'autoamorçage a été créée, ce qui permet de tenir compte adéquatement non seulement du plan de sondage complexe, mais aussi de tous les ajustements de non-réponse et de calage apportés à la pondération dans l'estimation de la variance et les tests statistiques. Ces poids ont par la suite été utilisés avec le logiciel SUDAAN²⁶. Pour des informations supplémentaires sur les poids d'autoamorçage, le lecteur est invité à consulter l'annexe 1 du guide méthodologique de l'EQDEM 2017.

Le coefficient de variation (CV) est l'indicateur de précision relative recommandé par l'ISQ pour l'analyse des données de l'EQPPEM. Les estimations dont le CV est supérieur à 15 % sont accompagnées d'un astérisque dans les tableaux et les figures du rapport de l'EQPPEM. Le tableau suivant résume le système de classification utilisé pour la présentation des estimations dans les tableaux en fonction de leur coefficient de variation.

26. La méthodologie pour réaliser les tests statistiques (voir section 2.4.6) et calculer les intervalles de confiance a été améliorée entre l'EQDEM 2012 et l'EQDEM 2017. La nouvelle façon de faire nécessite 500 poids d'autoamorçage.

Tableau 2.6

Système de classification de la précision des estimations en vigueur à l'Institut de la statistique du Québec

Précision de l'estimation (CV)	Signe distinctif	Réserve à mentionner lors de la diffusion
CV ≤ 15 %	(aucun signe)	Estimation diffusée sans mention
15 % < CV ≤ 25 %	*	Estimation à interpréter avec prudence
CV > 25 %	**	Estimation à utiliser avec circonspection; fournie à titre indicatif seulement

Source : Institut de la statistique du Québec.

Rappelons que le CV se calcule comme suit :

$$CV = \frac{\sqrt{\text{variance de l'estimation}}}{\text{estimation}}$$

Soulignons que chaque estimation ponctuelle de l'EQPPEM diffusée par l'ISQ sur son site Web est présentée avec l'estimation par intervalle correspondante. Les intervalles diffusés ont un niveau de confiance de 95 %. Pour un paramètre donné, plus l'intervalle de confiance à 95 % est court, meilleure est la précision de l'estimation qui est produite avec les données de l'EQPPEM.

2.3.6 Analyses et tests statistiques

Dans les produits de l'ISQ, en présence d'un croisement entre deux variables du fichier de l'EQPPEM, un test global d'indépendance du khi-deux²⁷ est d'abord effectué. Ce test permet de déceler une association possible entre la variable analysée et la variable de croisement, correspondant généralement à une caractéristique sociodémographique. Pour l'analyse des données de l'EQPPEM, les tests dont le seuil de signification observé est inférieur au seuil fixé à 5 % montrent une association significative.

Lorsque la variable (d'analyse ou de croisement) compte plus de deux catégories et que le test global d'indépendance est significatif, des tests d'égalité de proportions sont menés afin de déterminer quelles sont les proportions qui diffèrent significativement²⁸. Chacun des tests est basé sur une statistique de Wald calculée à partir de la différence des transformations *logit*²⁹ des proportions estimées. La variance et la covariance de la différence des *logits* sont calculées à l'aide des poids d'autoamorçage. Cette méthode de construction des tests requiert les 500 pondérations d'autoamorçage et peut être utilisée pour toutes les comparaisons de proportions. Notons qu'un écart important entre deux proportions n'est pas nécessairement statistiquement significatif. Tout dépend de la variabilité associée aux estimations. La section 3.1 présente plus de renseignements sur les tests statistiques disponibles sur le portail de l'Infocentre de santé publique. Le lecteur désirent obtenir de l'information générale sur les tests statistiques, leur utilité

27. Plus précisément, c'est la statistique F (version log-linéaire) correspondant à un ajustement de Satterthwaite du test du khi-deux qui est utilisée.

28. Dans le cas où la variable de croisement est la région administrative, et en présence d'un test global significatif, des tests d'égalité de proportions ont été faits entre chacune des régions et son complémentaire provincial (le reste du Québec).

29. La transformation « *logit* » = $\log(p/(1-p))$ est utilisée comme transformation normalisatrice dans la construction d'intervalles de confiance afin que leurs taux de couverture soient améliorés, en particulier pour de petites proportions (Korn et Graubard, 1999).

et ce qu'ils apportent aux analyses est invité à consulter l'annexe 8 du guide méthodologique de l'EQDEM 2012.

2.4 Portée et limites de l'enquête

Tout a été mis en place pour assurer la qualité et la représentativité de l'EQPPEM. La disponibilité d'une base de sondage construite avec des fichiers récents détenus par le MEES a, par exemple, permis de préciser la définition des enfants de la population visée. En plus d'offrir une excellente couverture, cette base comporte plusieurs renseignements facilitant le contact avec les parents des enfants des échantillons. D'ailleurs, la possibilité de réaliser une stratification adéquate des enfants de la population visée selon la région administrative ou le réseau local de services dans lequel se trouve l'école fréquentée par l'enfant assure l'élaboration d'un plan d'échantillonnage efficace pour les objectifs de l'EQPPEM.

L'échantillon disponible pour l'EQPPEM est de taille considérable à l'échelle provinciale. Le nombre de répondants disponible à l'échelle régionale (RA ou RSS) offre également un potentiel intéressant pour la production d'estimations de qualité pour des phénomènes peu fréquents touchant aussi peu que 10 % de la population.

La collecte informatisée permet aussi d'optimiser la qualité des données recueillies auprès des parents. Elle assure également un fardeau de réponse limité aux répondants. Il peut cependant y avoir certaines erreurs de mémoire pour les mesures portant sur des périodes de référence plus longues, comme celles associées au congé pris par les parents au moment de la naissance ou de l'adoption de l'enfant. Dans ce cas précis, puisque la venue ou l'adoption d'un enfant est un événement marquant dans la vie, on peut penser que les risques de ces erreurs sont moins importants pour l'EQPPEM.

Le traitement statistique des données de l'enquête a aussi fait l'objet d'une attention particulière. Ainsi, les pondérations qui doivent être utilisées dans l'analyse des données ont été calculées de façon à ce que la fiabilité de l'inférence à la population visée soit assurée. Les mesures de redressement effectuées permettent également de réduire au minimum les biais potentiels dus à la non-réponse. Il faut noter que tous les renseignements de la base de sondage et ceux fournis par les enseignants sont utilisés pour l'ajustement de la non-réponse associée aux questionnaires adressés aux parents.

Comme c'est le cas pour toutes les enquêtes populationnelles, bien que toutes les précautions possibles aient été prises pour que la qualité des données soit assurée et pour que les biais soient minimisés, certaines limites doivent être considérées. Ainsi, il est impossible de garantir l'exactitude des réponses fournies par les parents qui peuvent, par exemple, avoir de la difficulté à se souvenir de choses passées depuis plusieurs mois, voire quelques années, comme c'est le cas pour certains aspects particuliers (nombre d'heures de garde, par exemple) du parcours préscolaire de leur enfant. Il est aussi possible que les parents ignorent certaines informations. Par exemple, des parents adoptifs peuvent ne pas connaître tous les renseignements qui concernent les premières années de vie de leur enfant. Les réponses des parents à certaines questions, par exemple celles portant sur les activités parentales d'initiation à la littérature et la numératie, pourraient également être entachées d'un biais de désirabilité sociale. À cela s'ajoute le facteur « enseignant » en ce qui a trait à la mesure du développement des enfants étudiés, c'est-à-dire le fait que deux enseignants ne jugent pas nécessairement de la même façon un élève; de manière générale, certains enseignants peuvent avoir été plus sévères que d'autres lorsqu'ils ont répondu au questionnaire de l'IMDPE utilisé dans l'EQDEM.

Il faut aussi rappeler que les données disponibles à l'Infocentre n'incluent pas les enfants considérés EHDAA. Cela doit être gardé en mémoire, en particulier lors de l'examen du lien entre le développement des enfants et leurs caractéristiques, leur contexte familial ou leur parcours préscolaire.

Enfin, une enquête transversale³⁰ comme l'EQPPEM permet de déceler des liens entre deux variables de même que des différences entre des sous-groupes de population. Ce genre d'enquête ne permet toutefois pas d'établir de lien de causalité entre les caractéristiques étudiées. D'ailleurs, les analyses pouvant être réalisées à l'Infocentre, tout comme celles présentées dans le rapport de l'EQPPEM produit par l'ISQ, s'appuient essentiellement sur des méthodes bivariées d'analyse statistique. L'interprétation de certains résultats doit donc être faite avec prudence. Seules des analyses multivariées contrôlraient de manière appropriée l'effet de confusion possible de certains facteurs exogènes; de telles analyses permettraient, par exemple, de vérifier si certaines variables sociodémographiques expliquent lien entre certaines variables du parcours préscolaire et la vulnérabilité à la maternelle suggéré par les analyses bivariées. L'approche retenue pour l'analyse permet néanmoins d'explorer les données recueillies de façon utile et fournit ainsi une bonne description du parcours préscolaire des enfants visés par l'EQPPEM.

30. Ce type d'enquête mesure un phénomène au cours d'une période donnée.

III. Aspects méthodologiques spécifiques de l'Infocentre

3.1 Mesures de précision supplémentaires

Comme mentionné précédemment, le CV est l'indicateur de précision relative recommandé par l'ISQ pour l'analyse des données de l'EQPPEM 2017 et la seule mesure de précision utilisée dans le rapport diffusé par l'ISQ. Par contre, dans le recueil statistique publié sur son site Web, des intervalles de confiance sont présentés, tout comme dans les tableaux produits à l'Infocentre, dans lesquels une autre mesure de précision est disponible : l'erreur-type.

3.1.1 Intervalle de confiance

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

Il existe plusieurs types d'intervalles de confiance. Ceux-ci se distinguent par le choix de la loi servant à l'approximation de la distribution de l'estimateur. Les méthodes standard utilisent habituellement la loi normale ou la loi binomiale. Il existe également des méthodes qui permettent de s'affranchir de l'hypothèse selon laquelle les données suivent une loi connue; la méthode basée sur le calcul des centiles de la distribution empirique *bootstrap* est l'une d'elles. Pour plus de détails sur ces méthodes, veuillez consulter le guide méthodologique de l'EQDEM 2012 (ISQ et collab., 2013).

Ces intervalles de confiance, que nous avons appelés « intervalles d'autoamorçage », ont été utilisés dans d'autres enquêtes à l'Infocentre (notamment l'EQDEM 2012), dans le but avoué de rendre possible la production systématique de tests d'égalité des proportions dans les tableaux produits à l'Infocentre sans avoir recours à la méthode comparant deux intervalles de confiance, une méthode reconnue comme étant très conservatrice. Or, le fait de ne pas avoir de fondement théorique connu pour le test d'égalité de deux proportions basé sur cette méthode a récemment poussé l'ISQ à modifier la méthode de construction des intervalles de confiance et des tests statistiques pour ses enquêtes.

Par conséquent, pour les données de l'EQPPEM 2017, les intervalles de confiance produits sont basés sur la méthode *logit*-Wald, qui utilise une transformation *logit* de la proportion et génère un intervalle asymétrique. Ces intervalles sont donc cohérents avec le test d'égalité de deux proportions, qui utilise également une transformation *logit* (voir les sections 2.3.6 et 3.4.2). Cet intervalle de niveau $1 - \alpha$ est obtenu de la façon suivante :

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) \pm z_{1-\alpha/2} \sqrt{\hat{v}\left(\ln\left(\frac{\hat{p}}{1-\hat{p}}\right)\right)}$$

En utilisant la méthode Delta, on peut démontrer que

$$\hat{v}\left(\ln\left(\frac{\hat{p}}{1-\hat{p}}\right)\right) = \left(\frac{1}{\hat{p}(1-\hat{p})}\right)^2 \hat{v}(\hat{p}) ,$$

où $\hat{v}(\hat{p}) = \frac{\sum_{b=1}^B (\hat{p}_b - \hat{p})^2}{B}$ et \hat{p}_b est l'estimation de la proportion avec la b^{e} réplique d'autoamorçage, \hat{p} la moyenne de la proportion obtenue avec chaque réplique et B le nombre de répliques d'autoamorçage.

L'intervalle de confiance de la fonction *logit* (\hat{p}) est donc délimité par les bornes inférieure et supérieure suivantes :

$$I_{logit} = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) - z_{1-\alpha/2} \left(\frac{1}{\hat{p}(1-\hat{p})}\right) \sqrt{\hat{v}(\hat{p})}$$

$$S_{logit} = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) + z_{1-\alpha/2} \left(\frac{1}{\hat{p}(1-\hat{p})}\right) \sqrt{\hat{v}(\hat{p})}$$

Par transformation inverse, les bornes de l'intervalle de confiance pour \hat{p} sont définies comme suit :

$$I_{\hat{p}} = \frac{\exp(I_{logit})}{1 + \exp(I_{logit})}$$

$$S_{\hat{p}} = \frac{\exp(S_{logit})}{1 + \exp(S_{logit})}$$

Le niveau de confiance recommandé pour les intervalles produits à l'aide des données de l'EQPPEM 2017 est de 95 % pour toutes les analyses réalisées, qu'elles soient nationales, régionales ou locales (voir section 3.4.3).

3.1.2 Erreur-type

Une autre mesure de précision disponible dans les tableaux produits sur le portail de l'Infocentre est l'erreur-type. Celle-ci correspond à la racine carrée de la variance de l'estimateur (proportion ou effectif) et est estimée à l'aide des 500 poids d'autoamorçage fournis dans le fichier de données de l'enquête (voir la section 2.3.5).

3.2 Non-réponse partielle pour les indicateurs de l'Infocentre

3.2.1 Taux de non-réponse partielle pour les estimations à l'échelle régionale et locale

Pour les estimations régionales ou locales tirées de l'EQPPEM 2017, la non-réponse partielle n'a pas fait l'objet d'un examen détaillé. Cette tâche revient donc à l'utilisateur. Ce dernier devra évaluer lui-même si l'incidence de la non-réponse sur les estimations de proportions peut être considérée comme négligeable. Ce n'est pas le cas lorsqu'un écart important est observé dans les taux de non-réponse partielle entre des sous-groupes et que les écarts observés dans les estimations de proportions au sein de ces sous-groupes sont également importants. Aux fins de l'analyse, l'utilisateur peut faire l'hypothèse que l'incidence de la non-réponse partielle à l'échelle régionale est similaire à celle observée à l'échelle nationale, ou encore établir un diagnostic qualitatif de l'incidence de la non-réponse suivant la procédure décrite à l'annexe 2 du guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires de l'*Enquête québécoise sur le développement des enfants à la maternelle* de 2012 (ISQ, 2013). De plus, il est en général raisonnable de faire l'hypothèse qu'une non-réponse partielle inférieure à 10 % à l'échelle régionale ou locale entraîne de faibles risques de biais dans les estimations.

3.2.2 Traitement pour l'estimation des proportions et des effectifs

Des conseils sur les différentes façons de considérer la catégorie de non-réponse partielle pour l'estimation sont présentés à la section 2.3.4 du guide méthodologique de l'EQDEM 2012. À la section 2.3.5 de ce guide sont également données des explications sur le traitement de la non-réponse partielle dans l'estimation du nombre d'élèves possédant une caractéristique donnée.

3.3 Ajustement des proportions selon l'âge ou le sexe

Pour l'EQPPEM, la standardisation des estimations de proportions selon l'âge ou le sexe n'est pas offerte pour l'exploitation des données à l'Infocentre. La comparaison des distributions régionales (régions sociosanitaires) et infrarégionales (RLS ou regroupement de RLS pour les régions 03 et 12) des enfants selon

l'âge à l'admission à la maternelle ou selon le sexe n'a pas permis de détecter de différence significative entre les découpages géographiques considérés. Cela indique que la standardisation n'aurait pas (ou vraiment très peu) d'incidence sur les estimations régionales/sous-régionales produites. L'ajustement ne fait donc pas partie des options offertes aux utilisateurs des données de l'EQPPEM.

3.4 Tests statistiques et seuils de signification disponibles à l'Infocentre

3.4.1 Test global d'indépendance

Le test statistique d'indépendance du khi-deux employé à l'Infocentre pour l'édition 2017 de l'EQPPEM prend en considération le plan de sondage de l'enquête par le biais de l'ajustement de Satterthwaite (aussi appelé Rao-Scott de deuxième degré). En raison de l'utilisation du logiciel SAS à l'Infocentre, une formule légèrement différente pour la statistique du test est appliquée et celle-ci suit approximativement une loi du khi-deux et non la statistique F telle que proposée par l'ISQ. Les deux statistiques s'équivalent et mènent à des seuils de signification observés similaires lorsque le nombre de degrés de liberté est grand, ce qui est généralement le cas avec l'EQPPEM. Pour connaître les conditions de validité du test, veuillez consulter l'annexe 9 du guide méthodologique de l'EQDEM 2012.

3.4.2 Test d'égalité de deux proportions

Des tests d'égalité de deux proportions peuvent ensuite être réalisés. Ceux-ci permettent de mieux comprendre les associations révélées par les tests du khi-deux significatifs. La comparaison est basée sur la construction d'une statistique de Wald (W) à partir de la différence des transformations « *logit* » des proportions estimées.

$$W = \frac{\text{logit}(\hat{p}_1) - \text{logit}(\hat{p}_2)}{\sqrt{\hat{v}(\text{logit}(\hat{p}_1) - \text{logit}(\hat{p}_2))}} \sim \text{Normale}$$

où $\text{logit}(\hat{p}_i) = \log(\hat{p}_i / (1 - \hat{p}_i))$.

L'estimation de la variance de la différence de deux *logits* de l'équation précédente peut être obtenue directement avec SUDAAN à l'aide des poids d'autoamorçage; c'est l'approche utilisée par l'ISQ dans ses analyses.

Cependant, le test programmé à l'Infocentre pour étudier l'association entre un indicateur et une variable de croisement s'appuie sur une légère variante du test précédent, mais s'inspire toujours de la transformation *logit*. Ainsi, étant donné l'utilisation du logiciel SAS par l'Infocentre, l'adaptation de la macro SAS existante pour la différence entre deux ratios (ou proportions) s'avère une option intéressante pour le calcul de l'estimation de la variance de la différence de deux *logits*. Le dénominateur du test de Wald devient alors :

$$\hat{v}(\text{logit}(\hat{p}_1) - \text{logit}(\hat{p}_2)) = \frac{\sum_{b=1}^B (\hat{X}_b - \hat{X})^2}{B}$$

où \hat{X}_b correspond à l'estimation de la différence de $\text{logit}(\hat{p}_1) - \text{logit}(\hat{p}_2)$ obtenue avec le b^e poids d'autoamorçage, \hat{X} à la moyenne des différences de *logit* obtenues à l'aide des poids d'autoamorçage et B au nombre de répliques d'autoamorçage.

Des vérifications ont été effectuées dans le cadre de l'Enquête québécoise sur la santé de la population 2014-2015 afin de s'assurer que le test utilisé à l'Infocentre donnait des résultats similaires à ceux recommandés par l'ISQ. Tout porte à croire que ces deux approches aboutissent à des conclusions identiques avec des seuils observés du même ordre de grandeur, sauf lorsque le seuil observé est très près du seuil de signification.

3.4.3 Seuils de signification disponibles à l'Infocentre

Au moment de construire un test statistique d'égalité de deux proportions, par exemple, on fixe habituellement le seuil de signification, soit la probabilité de rejeter l'hypothèse d'égalité alors que les proportions sont en fait égales. Ce risque est donc contrôlé. D'un autre angle, on se préoccupe aussi de la puissance du test, c'est-à-dire la probabilité de détecter une différence entre les proportions lorsque cette différence existe dans la population visée. En règle générale, pour un seuil de signification fixé, la puissance augmente avec le nombre de répondants et baisse à mesure que l'ordre de grandeur de la différence à détecter diminue. À l'opposé, pour un nombre de répondants et un écart à détecter fixés, la puissance sera plus grande lorsque le seuil de signification est plus élevé.

Compte tenu du nombre de répondants dans l'EQPPEM, l'ISQ recommande d'analyser les données de l'enquête en utilisant le seuil de signification de 5 % pour l'ensemble des analyses de données. Ce seuil statistique est celui qu'a retenu l'ISQ pour les analyses présentées dans le rapport de l'enquête (Lavoie et autres, 2018). Cette recommandation s'applique à la fois au test d'indépendance du khi-deux et aux tests d'égalité des proportions.

Conséquemment à cette recommandation, les intervalles de confiance seront construits au niveau de confiance de 95 %, qu'ils soient associés à des caractéristiques provinciales, régionales ou locales.

IV. Exploitation des données de l'EQPPEM 2017 sur le portail de l'Infocentre de santé publique

La présente section aborde des éléments à considérer pour une saine exploitation des données de l'EQPPEM 2017. Plus précisément, la section 4.1 recense les possibilités, en matière de production statistique, qu'offre le portail de l'Infocentre de santé publique pour les données de l'EQPPEM. La section 4.2 rassemble, quant à elle, les principales particularités des indicateurs de l'EQPPEM et des résultats diffusés par l'Infocentre de santé publique. On y présente notamment les choix faits pour certains indicateurs et les incidences sur l'utilisation et l'interprétation de ces indicateurs. Il est à noter que cette section n'a pas pour objet de couvrir l'ensemble des choix faits pour l'EQPPEM, mais seulement de donner de l'information supplémentaire sur certains de ceux-ci.

4.1 Production de tableaux et comparaison de proportions sur le portail de l'Infocentre de santé publique

Le portail de l'Infocentre de santé publique propose plusieurs options d'analyse pour les indicateurs du PNS dont la source est l'EQPPEM 2017. La première partie de cette section présente les différents paramètres d'une requête type ainsi que les valeurs que peuvent prendre ces paramètres. La deuxième partie fournit quelques exemples de requêtes et de résultats pour chacun des niveaux géographiques permis (ensemble du Québec, régional et local). Des recommandations importantes à l'intention de l'utilisateur y sont également formulées quant aux tests à utiliser pour la comparaison d'indicateurs du PNS et quant à la production de tableaux. Il est d'ailleurs à souligner que ces recommandations ont toutes été retenues et appliquées sur le portail.

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

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

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

Bien entendu, la première étape dans l'exécution d'une requête est le choix d'un indicateur ou d'une famille d'indicateurs parmi ceux disponibles sur le portail. En cliquant sur le nom de l'indicateur retenu, l'utilisateur a accès à la requête; les paramètres suivants doivent alors être définis.

Paramètre « Niveau géographique »

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

Paramètre « Région sociosanitaire »

Ce paramètre permet de choisir la région sociosanitaire à analyser lorsque le niveau géographique « Région sociosanitaire de résidence » est sélectionné. Il y a 16 régions.

Paramètre « Sous-population »

Ce paramètre permet de définir une seule sous-population d'analyse. Les choix offerts sont : « Sexe », « Âge » et « Totale ». Dans le cas où, par exemple, le paramètre « Sexe » est sélectionné et que l'utilisateur spécifie « par sexe », ce dernier obtiendra des résultats fournis séparément pour les garçons et pour les filles, en plus du résultat pour les sexes combinés. Lorsqu'aucune sous-population en particulier n'est visée, il suffit de choisir la valeur « Totale » pour ce paramètre.

Paramètre « Première variable de croisement »

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

Paramètre « Comparaison régionale »

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

Paramètre « Comparaison locale »

Ce paramètre permet à l'utilisateur de définir quel type d'analyse à l'échelle des RLS il souhaite effectuer, et il est affiché seulement lorsque le niveau géographique spécifié est « Local (RLS par région) ». Les choix possibles sont : « Aucune », « RLS comparés avec le reste de la région sociosanitaire de résidence » et « RLS comparés avec le reste du Québec ». Lorsqu'une comparaison locale est sélectionnée, le paramètre « Variable de croisement » n'est plus disponible.

Rappelons qu'il est recommandé de comparer un RLS au reste du Québec uniquement lorsque les RLS de la région sociosanitaire correspondante diffèrent significativement entre eux. Pour ce faire, on peut utiliser le test global « RLS comparés avec le reste de la région sociosanitaire de résidence ». Ainsi, en présence d'un résultat significatif à ce test global, le résultat du test de comparaison du RLS avec le reste du Québec peut être interprété.

Paramètre « Deuxième variable de croisement »

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

Paramètre « Test statistique »

Ce paramètre permet de demander l'exécution de tests statistiques. Tout d'abord, un test global du khi-deux est produit³¹. Ensuite, si le résultat est significatif, toutes les comparaisons deux à deux sont réalisées³². Comme l'indique le nom du seul choix actif, soit « Test de comparaison de l'indicateur selon la variable de croisement, stratifié selon la sous-population », l'autre choix étant « Aucun », ce paramètre permet de produire séparément, pour chacune des sous-populations définies par la variable de sous-population, les tests mentionnés précédemment.

L'utilisateur de l'Infocentre de santé publique remarquera rapidement que les tests statistiques effectués sur les données de l'EQPPEM 2017 accessibles sur le portail prennent toujours la même forme. Le diagramme 4.1 illustre le cheminement que doit suivre un utilisateur afin de bien comprendre les tests qui sont effectués lors d'une requête. En effet, un test global du khi-deux est produit lorsqu'une comparaison est demandée par l'utilisateur, afin d'évaluer s'il existe une relation significative entre l'indicateur et la variable de croisement, pour la sous-population étudiée³³. Dans le cas où le test global montre une relation significative (valeur p du test inférieure au seuil de signification α ³⁴), des tests d'égalité de proportions deux à deux sont effectués et un tableau présentant les résultats est affiché à la suite du tableau des estimations.

Dans le cas d'un indicateur à deux catégories seulement, analysé selon une variable de croisement qui ne possède également que deux catégories, le test d'égalité de deux proportions est utilisé comme test global (section 3.4.2). Le test global du khi-deux (section 3.4.1) pourrait aussi s'appliquer et donnerait en général un résultat similaire. C'est pour cette raison que seul le test d'égalité de deux proportions, décrit à la section 3.4.2, est effectué (comme illustré au diagramme 4.1). Le tableau présentant le résultat de ce test est donc affiché directement à la suite de celui qui contient les estimations.

Dans certaines circonstances, qui sont décrites aux sections 6.3.1 et 6.3.2 du guide méthodologique de l'EQDEM 2012, seuls les tests globaux du khi-deux seront produits par la requête. En présence d'un résultat significatif, l'utilisateur est invité à utiliser les intervalles de confiance de chacune des proportions, qui sont présentés dans les tableaux, afin de déterminer la source de la différence observée. Il peut donc lui-même comparer les intervalles de confiance fournis dans le tableau pour tenter de savoir si l'écart observé est significatif, ou encore réduire la taille de la requête³⁵.

31. Dans le cas d'un indicateur à deux catégories et d'une variable de croisement aussi à deux catégories, seules les comparaisons deux à deux sont produites. Voir la section 3.4.2 pour plus de renseignements.

32. Il est possible que malgré un test global significatif, les comparaisons deux à deux ne soient pas effectuées. Pour plus de renseignements, veuillez consulter la section 4.2.

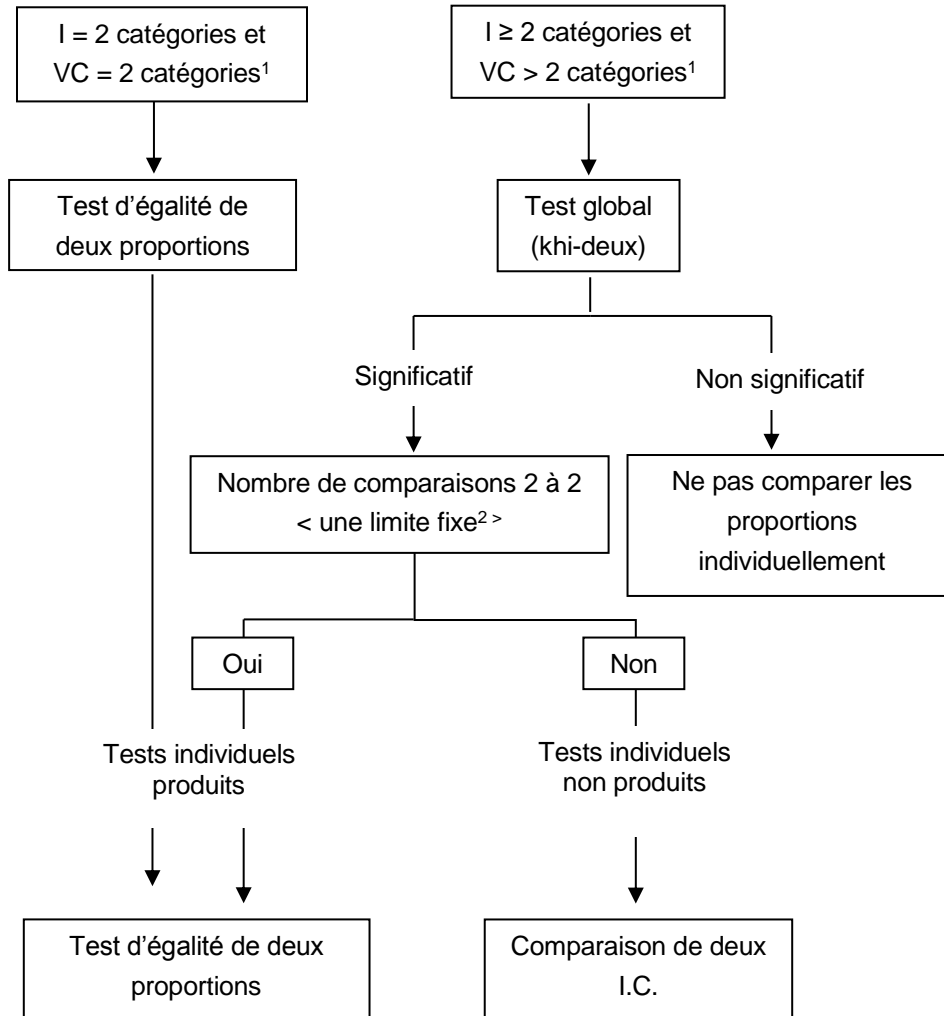
33. Il y aura autant de tests globaux du khi-deux qu'il y a de sous-populations formées par la requête.

34. Pour l'EQPPEM 2017, une seule valeur pour le seuil de signification est permise, soit 5 %. Pour de l'information sur le seuil recommandé, veuillez consulter la section 3.4.3.

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

Diagramme 4.1

Tests statistiques recommandés et produits à l'Infocentre de santé publique pour la comparaison des proportions d'un indicateur de l'EQPPEM 2017 selon une variable de croisement^{1,2}



1. I = Indicateur et VC = Variable de croisement.

2. Si l'indicateur (I) a deux catégories, les tests individuels sont produits seulement si le nombre total de comparaisons deux à deux est inférieur à 30, ou encore à 32 lorsque la région sociosanitaire est sélectionnée comme variable de croisement. Si l'indicateur (I) a plus de deux catégories, les tests individuels sont produits seulement si le nombre total de comparaisons deux à deux est inférieur à 60, ou encore à 64 lorsque la région sociosanitaire est sélectionnée comme variable de croisement. Le nombre total de comparaisons deux à deux correspond à la sommation des comparaisons deux à deux (entre la variable de croisement et l'indicateur) des modalités de la sous-population pour lesquelles le test du khi-deux est significatif.

Paramètre « Diffusion »

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

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

Pour les indicateurs du PNS dont la source est l'EQPPEM 2017, l'Infocentre de santé publique offre trois niveaux géographiques d'analyse : une production pour l'ensemble du Québec, une à l'échelle régionale et une à l'échelle locale. C'est le paramètre « Niveau géographique » de la fenêtre de saisie qui permet de faire le choix.

Pour chacun de ces trois niveaux géographiques, la façon d'analyser un indicateur est quasiment identique. Il faut choisir la « sous-population », la « première variable de croisement », la « deuxième variable de croisement » (dans certaines situations seulement) et la sortie ou non de « tests statistiques ». Comme expliqué à la section précédente, seuls le nombre permis de variables de croisement et les tests permis peuvent différer selon le niveau géographique choisi.

Les prochains paragraphes abordent, à l'aide d'exemples, chacun de ces éléments. Le premier groupe d'exemples montre les différentes options pour obtenir des estimations de proportions portant sur l'ensemble des enfants fréquentant la maternelle 5 ans du Québec visés par l'enquête. Le deuxième groupe concerne les analyses pour une région sociosanitaire de résidence qui permettent de comparer la région sociosanitaire avec le reste du Québec ou avec une autre région. Le troisième groupe permet des analyses encore plus fines, soit à l'échelle des RLS, ce qui permet des comparaisons entre une RLS avec le reste de la région sociosanitaire de résidence.

Premier groupe d'exemples : Ensemble du Québec

Pour ce premier groupe d'exemples, c'est la proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle selon leur sexe qui est analysée. Deux options sont offertes : la première consiste à produire les proportions selon le sexe comme variable de croisement (tableau 4.1); la seconde consiste à choisir le sexe comme variable de sous-population (tableau 4.2). Seul l'en-tête des tableaux d'estimations demandés diffère, mais la différence entre les deux options réside principalement dans les analyses et les tests disponibles. Il incombe donc à l'utilisateur de produire le tableau qui correspond à ses besoins. Par exemple, au tableau 4.1, il est possible d'obtenir un test statistique de comparaison de l'indicateur selon la variable de croisement (sexe), tandis qu'au tableau 4.2, puisqu'il n'y a aucune variable de croisement sélectionnée, il n'est pas possible d'effectuer un tel test.

Pour des tableaux portant sur l'ensemble du Québec, il est aussi possible de choisir la valeur « Région sociosanitaire de résidence » comme première variable de croisement. Ce choix permet de faire une analyse des régions sous un angle national. De cette façon, et lorsque le test global est significatif, les prévalences régionales de l'indicateur sont comparées à celles du reste du Québec, et ce, pour chacune des régions sociosanitaires. Le tableau 4.3 reproduit le visuel de la sortie de l'Infocentre de santé publique pour une telle requête sans les tests.

Seulement pour le niveau géographique national³⁶, il est possible dans certaines situations de choisir une deuxième variable de croisement. Le cas échéant, les estimations sont produites pour le croisement des deux variables choisies, séparément pour chacune des sous-populations définies, mais aucun test statistique n'est effectué. Le tableau 4.4 donne un exemple de sortie de l'Infocentre de santé publique.

Tableau 4.1

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle pour l'ensemble du Québec, EQPPEM 2017

Variable de croisement	Statistiques				
Sexe	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	92,7	0,5	0,45	(91,8 - 93,5)	0,0
Féminin	91,9	0,5	0,49	(90,9 - 92,8)	0,0
Total	92,3	0,4	0,33	(91,6 - 92,9)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

Tableau 4.2

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle pour l'ensemble du Québec, EQPPEM 2017

Sous-population	Statistiques				
Sexe	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	92,7	0,5	0,45	(91,8 - 93,5)	0,0
Féminin	91,9	0,5	0,49	(90,9 - 92,8)	0,0
Total	92,3	0,4	0,33	(91,6 - 92,9)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

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

Tableau 4.3

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle pour l'ensemble du Québec, EQPPEM 2017

Variable de croisement	Statistiques				
	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
01 Bas-Saint-Laurent	92,7	1,2	1,09	(90,2 - 94,5)	0,0
02 Saguenay–Lac-Saint-Jean	92,5	1,2	1,14	(89,9 - 94,4)	0,0
03 Capitale-Nationale	94,6	0,8	0,72	(93,0 - 95,9)	0,0
04 Mauricie et Centre-du-Québec	90,2	1,1	1,01	(88,1 - 92,0)	0,0
05 Estrie	91,7	1,1	0,99	(89,6 - 93,5)	0,0
06 Montréal	91,8	0,9	0,84	(90,0 - 93,3)	0,0
07 Outaouais	91,4	1,4	1,31	(88,5 - 93,7)	0,0
08 Abitibi-Témiscamingue	88,9	1,8	1,59	(85,4 - 91,7)	0,0
09 Côte-Nord	92,6	1,3	1,22	(89,8 - 94,6)	0,0
10 Nord-du-Québec	94,6	2,8	2,63	(86,5 - 98,0)	0,0
11 Gaspésie–Îles-de-la-Madeleine	94,5	1,2	1,09	(91,9 - 96,3)	0,0
12 Chaudière-Appalaches	92,5	0,8	0,73	(90,9 - 93,8)	0,0
13 Laval	93,3	1,4	1,28	(90,3 - 95,4)	0,0
14 Lanaudière	92,7	1,2	1,14	(90,2 - 94,7)	0,0
15 Laurentides	92,2	1,4	1,26	(89,3 - 94,3)	0,0
16 Montérégie	92,7	1,1	0,98	(90,5 - 94,4)	0,0
Ensemble du Québec	92,3	0,4	0,33	(91,6 - 92,9)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, décembre 2018

Tableau 4.4

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle pour l'ensemble du Québec, EQPPEM 2017

Variables de croisement		Statistiques				
Structure géographique	Nombre de déménagements dans les cinq dernières années	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
RMR de Montréal	Aucun déménagement	93,7	0,7	0,70	(92,2 - 94,9)	0,0
	Un déménagement	91,7	1,1	1,04	(89,4 - 93,5)	0,0
	Deux déménagements	90,2	1,9	1,75	(86,2 - 93,1)	0,0
	Trois déménagements ou plus	91,8	2,4	2,17	(86,4 - 95,2)	0,0
Autres RMR	Aucun déménagement	94,7	0,6	0,58	(93,4 - 95,7)	0,0
	Un déménagement	93,1	1,2	1,09	(90,6 - 94,9)	0,0
	Deux déménagements	91,7	2,1	1,94	(87,0 - 94,8)	0,0
	Trois déménagements ou plus	92,1	2,3	2,08	(86,9 - 95,3)	0,0
Agglomération de recensement	Aucun déménagement	94,5	0,9	0,84	(92,6 - 95,9)	0,0
	Un déménagement	91,0	1,9	1,69	(87,1 - 93,8)	0,0
	Deux déménagements	84,2	4,8	4,06	(74,5 - 90,6)	0,0
	Trois déménagements ou plus	86,4	3,8	3,32	(78,5 - 91,7)	0,0
Petites villes et monde rural	Aucun déménagement	90,4	0,8	0,75	(88,8 - 91,8)	0,0
	Un déménagement	89,0	1,6	1,46	(85,8 - 91,5)	0,0
	Deux déménagements	91,5	2,1	1,96	(86,8 - 94,7)	0,0
	Trois déménagements ou plus	89,5	3,5	3,16	(81,5 - 94,3)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, décembre 2018

Deuxième groupe d'exemples : régional (Région sociosanitaire)

Pour les analyses à l'échelle régionale, le paramètre « Comparaison régionale » permet de définir le type d'analyse souhaitée. Les choix sont : « Aucune », « Région sociosanitaire de résidence comparée avec le reste du Québec » et « Région sociosanitaire de résidence comparée avec une autre région sociosanitaire ». Dans certaines situations, une seule variable de croisement peut être sélectionnée. Le tableau 4.5, obtenu à l'aide du choix « Aucune » pour le paramètre « Comparaison régionale », présente la proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle selon le sexe de l'enfant pour la région de Montréal, en plus des résultats pour l'ensemble du Québec.

Pour vérifier si cette proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle est différente de celle du reste du Québec, l'utilisateur peut obtenir l'information dans la colonne « Résultat du test » au tableau 4.6, à l'aide du choix « Région sociosanitaire de résidence comparée avec le reste du Québec » pour le paramètre « Comparaison régionale », de même que la proportion pour le reste du Québec.

Pour vérifier si cette proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle est différente de celle d'une autre région, il peut obtenir l'information dans la colonne « Résultat du test » au tableau 4.7, à l'aide du choix « Région sociosanitaire de résidence comparée avec une autre région sociosanitaire » pour le paramètre « Comparaison régionale ».

Tableau 4.5

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle, EQPPEM 2017

Sous-population	Territoire	Statistiques				
Sexe	Région sociosanitaire de résidence	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Masculin	06 Montréal	92,1	1,3	1,17	(89,4 - 94,1)	0,0
	Ensemble du Québec	92,7	0,5	0,45	(91,8 - 93,5)	0,0
Féminin	06 Montréal	91,4	1,4	1,25	(88,6 - 93,6)	0,0
	Ensemble du Québec	91,9	0,5	0,49	(90,9 - 92,8)	0,0
Total	06 Montréal	91,8	0,9	0,84	(90,0 - 93,3)	0,0
	Ensemble du Québec	92,3	0,4	0,33	(91,6 - 92,9)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

Tableau 4.6

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle, EQPPEM 2017

Territoire	Statistiques					
Région sociosanitaire de résidence	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test ¹
06 Montréal	91,8	0,9	0,84	(90,0 - 93,3)	0,0	ns
Reste du Québec	92,5	0,4	0,34	(91,8 - 93,1)	0,0	
Ensemble du Québec	92,3	0,4	0,33	(91,6 - 92,9)	0,0	

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, décembre 2018

Tableau 4.7

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle, EQPPEM 2017

Territoire	Statistiques					
Région sociosanitaire de résidence	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle	Résultat du test ¹
06 Montréal	91,8	0,9	0,84	(90,0 - 93,3)	0,0	sig
03 Capitale-Nationale	94,6	0,8	0,72	(93,0 - 95,9)	0,0	

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

Troisième groupe d'exemples : local (RLS par région)

De façon similaire à ce qui a été présenté pour le dernier niveau d'analyse, un paramètre « Comparaison locale » doit être défini pour les analyses à l'échelle locale. Les choix possibles sont : « Aucune », « RLS comparés avec le reste de la région sociosanitaire de résidence » et « RLS comparés avec le reste du Québec ». Encore une fois, dans certaines situations, une seule variable de croisement peut être sélectionnée.

Le tableau 4.8 présente le visuel de la sortie d'une requête à l'échelle locale, où aucune sous-population ni variable de croisement ne sont demandées, et pour laquelle la comparaison de chacun des RLS avec le reste de la région est requise (à l'aide du choix « RLS comparés avec le reste de la région sociosanitaire de résidence » pour le paramètre « Comparaison locale »), le territoire de RLS étant considéré comme une variable de croisement.

Tableau 4.8

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle, région sociosanitaire de résidence – 12 Chaudière-Appalaches, EQPPEM 2017

Territoire (variable de croisement)	Statistiques				
	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
RLS Alphonse-Desjardins	91,3	2,0	1,85	(86,9 - 94,3)	0,0
RLS de Beauce	93,4	1,2	1,10	(90,9 - 95,2)	0,0
Autres RLS regroupés	92,4	1,0	0,92	(90,4 - 94,0)	0,0
Total de la région 12 Chaudière-Appalaches	92,5	0,8	0,73	(90,9 - 93,8)	0,0

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, décembre 2018

Test global d'indépendance entre l'indicateur et les RLS, région sociosanitaire de résidence 12 Chaudière-Appalaches

Statistique
Valeur p ¹
0,6353

Notes

1 : Valeur p obtenue par le test global d'indépendance du khi-deux sur les proportions brutes.

· Le tableau des comparaisons de chaque RLS avec le reste de leur région sociosanitaire de résidence est présenté seulement si le test global d'indépendance du khi-deux est significatif au seuil de 5 % et si la valeur p n'est pas masquée.

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

Lors de l'analyse des données sur le portail de l'Infocentre de santé publique, certaines situations méritent une attention particulière. Chaque sous-section présente un cas précis d'indicateur pour lequel des décisions ont été prises dans le cadre de la préparation des guides méthodologiques précédents ou par le groupe de travail de l'EQPPEM 2017, et précise les conséquences ou les mises en garde qui en découlent pour les analyses.

4.2.1 Nombre trop important de tests demandés pour une requête

Lorsque le nombre de comparaisons deux à deux est supérieur à 60, ou encore à 64 lorsque la région sociosanitaire de résidence est sélectionnée comme variable de croisement, celles-ci ne sont pas produites lors de l'exécution de la requête et seuls les tests globaux du khi-deux le sont, un pour chacune des sous-populations définies par le paramètre « Sous-population », et ce, même si un test du khi-deux est significatif.

Lorsqu'une telle situation se présente, une note mentionnant que le nombre de comparaisons deux à deux est trop élevé et que celles-ci ne sont pas effectuées est affichée au bas du tableau. Un exemple de cette note est présenté au tableau 4.9. L'utilisateur est alors invité à réduire l'ampleur de sa demande ou, pour les tests du khi-deux significatifs, à utiliser les intervalles de confiance de chacune des proportions, qui sont présentés dans les tableaux, afin de déterminer la source de la différence observée. Pour plus d'information, la section 5.1.2 du guide méthodologique de l'EQDEM 2012 peut être consultée.

Tableau 4.9

Répartition des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde à au moins une période entre la naissance et la maternelle selon la durée cumulative en milieu de garde pour l'ensemble du Québec, EQPPEM 2017 (extrait d'un tableau complet)

Variable de croisement	Indicateur	Statistiques				
		Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
01 Bas-Saint-Laurent	12 mois ou moins	* 3,2	24,2	0,77	(2,0 - 5,1)	1,1
	13 à 24 mois	* 5,0	21,4	1,07	(3,3 - 7,6)	1,1
	25 à 36 mois	* 8,6	15,2	1,31	(6,4 - 11,6)	1,1
	37 à 48 mois	30,3	6,8	2,06	(26,4 - 34,5)	1,1
	Plus de 48 mois	52,9	4,3	2,27	(48,4 - 57,3)	1,1
02 Saguenay-Lac-Saint-Jean	12 mois ou moins	* 3,9	21,7	0,84	(2,5 - 5,9)	2,0
	13 à 24 mois	* 4,0	21,2	0,86	(2,7 - 6,1)	2,0
	25 à 36 mois	9,5	13,6	1,29	(7,3 - 12,3)	2,0
	37 à 48 mois	36,3	5,8	2,10	(32,3 - 40,5)	2,0
	Plus de 48 mois	46,2	5,0	2,32	(41,7 - 50,8)	2,0
03 Capitale-Nationale	12 mois ou moins	* 3,9	15,8	0,62	(2,9 - 5,3)	1,1
	13 à 24 mois	* 5,3	15,2	0,80	(3,9 - 7,1)	1,1
	25 à 36 mois	7,4	11,9	0,88	(5,9 - 9,3)	1,1
	37 à 48 mois	39,8	4,0	1,61	(36,7 - 43,0)	1,1
	Plus de 48 mois	43,7	3,5	1,52	(40,7 - 46,7)	1,1
04 Mauricie et Centre-du-Québec	12 mois ou moins	* 3,9	17,5	0,68	(2,7 - 5,4)	1,0
	13 à 24 mois	7,3	12,4	0,91	(5,7 - 9,3)	1,0
	25 à 36 mois	10,5	10,2	1,07	(8,6 - 12,8)	1,0
	37 à 48 mois	34,0	4,9	1,67	(30,8 - 37,4)	1,0
	Plus de 48 mois	44,3	3,9	1,73	(40,9 - 47,7)	1,0

Test global d'indépendance entre l'indicateur et la région sociosanitaire de résidence

Statistique
Valeur p ¹
<0,0001

Notes

1 : Valeur p obtenue par le test global d'indépendance du khi-deux sur les proportions brutes.

- Étant donné que le nombre de comparaisons deux à deux découlant de cette requête est supérieur à 64, elles ne sont pas présentées. Pour les résultats globaux significatifs au seuil de 5 %, il est suggéré d'effectuer des tests t d'égalité ou de comparer les intervalles de confiance deux à deux afin d'identifier une source de différence.

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

4.2.2 Non-réponse partielle

Proportion de non-réponse partielle supérieure au seuil critique

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

Les questions pour lesquelles le taux de non-réponse partielle dépasse le seuil critique doivent en général être analysées plus en détail au regard des biais. Les non-répondants peuvent présenter des caractéristiques différentes des répondants, ce qui peut avoir un effet non négligeable sur les estimations présentées (sous-estimation ou surestimation)³⁷.

Chacun des tableaux produits sur le portail de l'Infocentre de santé publique affiche les pourcentages de non-réponse partielle de l'indicateur analysé. Afin d'aider les utilisateurs, une note est placée au bas du tableau lorsque cette non-réponse dépasse le seuil critique correspondant au niveau géographique considéré. Le tableau 4.10 présente un exemple de tableau sans croisement affichant chacune de ces notes.

37. Pour savoir comment procéder au diagnostic et à l'analyse qualitative de la non-réponse partielle, le lecteur est invité à consulter l'annexe 2 du guide méthodologique de l'EQDEM 2012.

Tableau 4.10

Proportion des enfants de maternelle 5 ans n'ayant pas été gardés régulièrement à au moins une période entre la naissance et la maternelle parce qu'un parent était sans emploi et demeurait à la maison, région sociosanitaire de résidence – 06 Montréal, EQPPEM 2017

Territoire	Statistiques				
	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
06 Montréal	* 31,9	17,9	5,73	(21,8 - 44,0)	** 16,2
Ensemble du Québec	29,3	7,2	2,12	(25,3 - 33,6)	ª 6,9

Notes

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

* : Coefficient de variation supérieur à 15 % et inférieur ou égal à 25 %. La valeur de la proportion doit donc être interprétée avec prudence.

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

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, novembre 2018

Bien que la non-réponse partielle d'un indicateur soit faible, il est possible que celle de la variable utilisée comme variable de croisement soit non négligeable. La proportion de non-réponse partielle affichée dans un tableau avec croisement correspond à la non-réponse de l'indicateur pour chacune des catégories de la variable de croisement. La taille des domaines d'étude formés par ces catégories est en général beaucoup plus faible que celle de l'indicateur sans croisement, puisque chacun des domaines est un sous-ensemble de la population globale sur laquelle porte l'indicateur. Par conséquent, la précision des estimations est également plus faible.

Par ailleurs, pour les tableaux avec croisement, une note au bas du tableau indique le pourcentage de non-réponse partielle de la variable de croisement (tableau 4.11).

Tableau 4.11

Proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle pour l'ensemble du Québec, EQPPEM 2017

Variable de croisement	Statistiques					
	Nombre ¹	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
Plus haut diplôme obtenu par le père (ou le conjoint)						
Aucun diplôme	5 540	85,3	1,8	1,55	(82,0 - 88,1)	0,0
Diplôme de niveau secondaire (incluant le DEP)	21 350	91,4	0,7	0,60	(90,1 - 92,5)	0,0
Diplôme de niveau collégial (incluant l'AEC)	14 860	93,2	0,7	0,69	(91,8 - 94,5)	0,0
Diplôme universitaire	27 050	93,5	0,7	0,61	(92,2 - 94,6)	0,0

Notes

1 : Population estimée (arrondie à la dizaine) des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle. Puisque la non-réponse partielle a été répartie selon la proportion observée pour chaque catégorie de la variable de croisement (ou de la sous-population ou du territoire), la somme des estimations de population obtenues ne correspond pas nécessairement à l'estimation de population pour les catégories regroupées.

- Les données des enfants handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDA) sont exclues du fichier.

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

- La non-réponse partielle de la variable de croisement « Plus haut diplôme obtenu par le père (ou le conjoint) », calculée sur les répondants visés par cet indicateur, est de 13,2 %.

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

Source : Infocentre de santé publique à l'Institut national de santé publique du Québec, décembre 2018

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

Il a été mentionné à la section 2.3.3 que les estimations d'effectifs (populations estimées) diffusées sur le portail de l'Infocentre de santé publique sont ajustées afin de tenir compte de la non-réponse partielle. Aucun traitement de la non-réponse partielle de la variable de croisement n'est toutefois effectué, c'est-à-dire que la non-réponse partielle de la variable de croisement n'est pas répartie au sein de ses différentes catégories. Seule la non-réponse de l'indicateur l'est pour le sous-groupe des répondants à la variable de croisement, de sorte que les estimations des effectifs qui apparaissent dans les tableaux avec croisement sous-estiment les effectifs réels de la population³⁸. Afin de faire état de cette limite, une note est affichée, le cas échéant, au bas du tableau. Le tableau 4.11 est un exemple de tableau affichant cette note.

4.2.3 Diffusion à l'échelle locale pour deux régions sociosanitaires de résidence

Comme il a été brièvement mentionné à la section 2.4, le niveau géographique utilisé pour la pondération correspond à celui des objectifs de l'enquête, soit l'échelle locale pour les régions avec suréchantillon (pour la diffusion de statistiques par RLS : Capitale-Nationale et Chaudière-Appalaches) et l'échelle régionale pour les autres régions sociosanitaires. Il a donc été décidé de se limiter aux objectifs de l'enquête, en raison principalement du fait qu'un ajustement plus fin pour les régions sans suréchantillon aurait eu pour effet de diminuer la qualité de l'ajustement de la pondération au niveau de l'ensemble du Québec. En fait, l'ensemble des choix méthodologiques faits pour l'EQPPEM 2017 n'avaient pas pour objectif de fournir des

38. L'utilisateur qui souhaite obtenir des estimations non biaisées des effectifs d'un indicateur selon une variable de croisement peut procéder de la façon suivante : produire les tableaux de l'indicateur avec et sans le croisement et multiplier la proportion de la catégorie du croisement souhaité par l'effectif du tableau sans croisement pour cette même catégorie. Par exemple, si l'on souhaite obtenir le nombre d'enfants de maternelle 5 ans qui ont été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle parmi ceux dont le plus haut diplôme obtenu par le père est un diplôme universitaire, il faut prendre la proportion des enfants qui ont été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle parmi ceux dont le plus haut diplôme obtenu par le père est un diplôme universitaire dans le tableau croisé et multiplier cette proportion par l'effectif des enfants qui ont été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle du tableau sans croisement.

données de qualité à l'échelle locale pour les régions qui n'ont pas fait d'achat d'unités supplémentaires. Par conséquent, il a été décidé de ne pas diffuser d'estimations à l'échelle locale pour ces régions sur le portail de l'Infocentre de santé publique.

4.2.4 Limites de certaines variables de croisement

L'Infocentre de santé publique offre à ses utilisateurs deux approches pour définir une sous-population d'enfants vivant au sein de ménages à faible revenu :

- une sous-population de familles à faible revenu est définie à partir de l'indicateur de faible revenu basé sur la mesure de faible revenu (MFR);
- une autre sous-population de familles à faible revenu est définie à l'aide de l'indicateur de faible revenu basé sur les seuils de faible revenu (SFR).

Il faut dire que les deux approches ont servi à des analyses de données d'enquête portant sur les enfants de maternelle. Par exemple, la deuxième approche a été utilisée dans l'*Enquête montréalaise sur l'expérience préscolaire des enfants de maternelle* (EMEP), réalisée par l'ISQ en 2012. Par contre, suivant les observations de son expert en matière de faible revenu, l'ISQ a retenu la première des deux approches pour les analyses présentées dans le rapport de l'EQPPEM 2017 (Lavoie et autres, 2018). Les prochains paragraphes offrent un résumé de ces observations. Le lecteur y trouvera les principaux éléments devant être pris en considération au sujet de la portée et des limites des deux variables de croisement associées au faible revenu offertes aux utilisateurs des données de l'EQPPEM à l'Infocentre.

Au sujet des seuils SFR :

Les seuils SFR « sont des limites de revenu en deçà duquel une famille est susceptible de consacrer une part plus importante de son revenu à l'achat de nécessités comme la nourriture, le logement et l'habillement qu'une famille moyenne » (Statistique Canada, 2014).

L'unité utilisée pour les seuils SFR est « l'unité familiale », au sens de la « personne seule » ou de la « famille économique » de Statistique Canada. Dans une unité familiale, le regroupement est effectué sur la base des liens de sang, d'alliance, d'union libre ou d'adoption. Il faut signaler qu'une « personne seule » ne constitue pas nécessairement un ménage d'une personne. Par exemple, un ménage de deux personnes peut être constitué de deux « personnes seules », c'est-à-dire de deux personnes sans lien de sang, d'alliance, d'union libre ou d'adoption. De la même manière, un ménage peut compter plus d'une « famille économique », c'est-à-dire au moins deux regroupements formés de deux personnes ou plus.

Les désavantages des seuils SFR énumérés ci-après (Crespo, 2008) ont conduit l'Unité des conditions de vie de l'ISQ à abandonner toute production statistique sur le faible revenu basée sur ceux-ci, et ce, dès l'année de référence 2009.

- La dernière année de référence pour laquelle les seuils SFR ont été calculés remonte à 1992. Depuis cette date, les seuils ont seulement été indexés à l'Indice des prix à la consommation canadien. Or, cette indexation ne remplace pas la procédure de calcul, basée sur la relation empirique entre les dépenses et les revenus. L'anachronisme est sans doute aggravé par le fait qu'en 1992, la récession sévissait. De ce fait, la part moyenne du revenu consacrée aux trois postes de dépenses devait être plus élevée que pendant les années subséquentes où une croissance économique a été observée. La qualité de cet indicateur est grandement affectée par cette situation.
- Dans le calcul du pourcentage moyen des dépenses consacrées aux trois postes en question, on ne se limite pas aux besoins essentiels, comme il devrait en être pour toute mesure liée au faible revenu : les dépenses (en alimentation, en vêtement et en transport) peuvent inclure des biens dits de luxe.

- On suppose que les unités familiales qui consacrent le même pourcentage de leur revenu à ces postes ont le même « niveau de vie », ce qui constitue une hypothèse discutable.
- La diminution progressive du pourcentage moyen du revenu consacré à ces trois postes depuis l'élaboration des premiers seuils en 1959 a conduit à une hausse de la prévalence estimée du faible revenu.
- Un poste de dépenses important, le transport, n'entre pas dans le calcul des seuils SFR, alors que les dépenses de transport constituent le deuxième poste de dépenses, après le logement et avant l'alimentation chez la population québécoise. Outre le fait que les dépenses de transport sont un poste important, l'accès à un transport est une condition essentielle pour sortir du faible revenu.

Au sujet de la MFR :

« La mesure de faible revenu (MFR) est un pourcentage fixe (50 %) du revenu des ménages médian "ajusté", ce dernier terme traduisant la prise en compte des besoins du ménage » (Statistique Canada, 2013).

- Les seuils MFR ne permettent pas de faire des comparaisons sur de longues périodes. En effet, puisqu'ils sont basés sur le revenu médian et que ce revenu médian varie de façon conjoncturelle, ils tendent à sous-estimer le taux de faible revenu en période de récession, et à le surestimer en période d'expansion. Par exemple, si une personne a le même revenu pour une année de creux et pour une année de sommet, elle aura une probabilité moins élevée d'être à faible revenu l'année de creux, et plus élevée de l'être l'année de sommet. Cela s'explique par le fait que d'une année à l'autre, le seuil augmente, étant donné que le revenu médian augmente. Ce désavantage est intimement lié au fait que la MFR constitue uniquement une mesure relative; il s'atténue si les comparaisons sont effectuées sur une courte période, relativement homogène sur le plan économique.
- Les comparaisons interrégionales, à l'intérieur d'une province, sont discutables, puisque les seuils MFR ne tiennent pas compte des coûts de la vie spécifiques aux différentes régions de cette province. En effet, les seuils MFR ne sont pas régionalisés.

4.2.5 Limites de certains indicateurs

Pour l'analyse du lien entre les différents domaines de développement mesurés dans l'EQDEM et l'expérience préscolaire des enfants ou leurs conditions de vie, les utilisateurs de l'Infocentre auront la possibilité de choisir comme variable d'analyse l'un ou l'autre des indicateurs de vulnérabilité des enfants défini dans l'EQDEM 2017.

Contrairement aux différentes possibilités offertes pour les analyses de ces indicateurs avec la banque de données de l'EQDEM 2017 disponible à l'Infocentre, seuls les indicateurs dichotomiques de vulnérabilité dans un domaine de développement construits à partir du 10^e centile (enfants considérés comme vulnérables lorsque le score est inférieur ou égal au 10^e centile³⁹) pourront être mis en lien avec l'expérience préscolaire ou les conditions de vie mesurées dans le cadre de l'EQPPM.

39. Il s'agit du 10^e centile de la distribution pondérée de l'ensemble des scores des enfants d'une population de référence pour ce domaine. Dans le cas de l'EQDEM 2017, il a été convenu que les seuils utilisés seraient ceux de l'EQDEM 2012 afin de rendre possibles les comparaisons temporelles.

V. Protection des renseignements personnels

5.1 Confidentialité des tableaux destinés à la diffusion

Bien que le fichier de partage contenant les microdonnées de l'EQPPEM disponible à l'Infocentre de santé publique soit un fichier non masqué⁴⁰, en raison de l'accord qu'a donné le répondant pour la communication de ses renseignements à certains organismes, il existe, pour celui qui diffuse des données sous forme de tableaux, une obligation contractuelle de respecter leur caractère confidentiel. Il est important de savoir que la divulgation d'information ne survient pas seulement lorsque l'on donne le nom et l'adresse d'un répondant. Elle peut également survenir de façon indirecte lorsque l'on croise différentes caractéristiques du répondant. Statistique Canada, dans son document *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*⁴¹, donne une définition assez complète du risque de divulgation et de nombreux exemples de situations où il y a divulgation. L'encadré 5.1 présenté à la fin de cette section-ci est un extrait de ce document. Il fournit de l'information supplémentaire sur la divulgation d'information lors de la diffusion de tableaux de résultats, dans le contexte général d'une enquête.

Dans le cadre de l'EQPPEM, l'utilisateur doit veiller à protéger la confidentialité des données individuelles. L'ISQ s'est doté d'un ensemble de règles de confidentialité pour prévenir la divulgation de renseignements confidentiels résultant de la diffusion de tableaux. Leur application permet de déterminer quels tableaux ou résultats comportent un risque de divulgation et quels traitements de confidentialité doivent leur être appliqués.

L'Infocentre de santé publique et les autres détenteurs du fichier de partage sont soumis aux mêmes obligations que l'ISQ pour les données de l'EQPPEM. C'est pourquoi les utilisateurs de données de cette enquête doivent appliquer des règles équivalentes à celles établies par l'ISQ pour la diffusion⁴² de tableaux. Ces règles sont présentées à la section 5.2 dans le contexte de la diffusion de résultats obtenus de l'Infocentre de santé publique, mais pourraient tout aussi bien s'appliquer au contexte de la diffusion de résultats obtenus de l'ISQ par l'accès à distance ou de ceux produits par les autres partenaires de l'enquête.

40. Il va sans dire que les données du fichier de partage sont anonymisées.

41. CENTRES DE DONNÉES DE RECHERCHE DE STATISTIQUE CANADA (2005), *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*, Ottawa, Statistique Canada, p. 23-24.

42. Le terme « diffusion » comprend, entre autres choses, la publication, la sortie d'un résultat hors des locaux sécurisés de l'organisme, la transmission du résultat, qu'elle soit faite sur support papier ou informatique, ou verbalement, etc.

Encadré 5.1

Types de divulgation et exemples dans les données d'enquête

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

Types de divulgation

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

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

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

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

Quelques exemples de divulgation dans les données d'enquête

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

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

Source : CENTRES DE DONNÉES DE RECHERCHE DE STATISTIQUE CANADA (2005). *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*, Ottawa, Statistique Canada, p. 23-24.

5.2 Règles de masquage par suppression

Afin de réduire le risque d'identification des répondants, lors de la production des tableaux statistiques sur le portail de l'Infocentre de santé publique, les règles de masquage par suppression ont été appliquées pour les utilisateurs ayant un droit d'accès 40 ou 50.

Si, pour un indicateur, l'utilisateur ayant un droit d'accès 20 ou 30 est en présence d'un tableau ayant au moins une cellule jugée confidentielle (dans les tableaux produits à l'Infocentre, une telle situation est identifiée par le symbole « ^{aaa} » apparaissant à la gauche du « nombre »), il doit choisir l'une des solutions suivantes :

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

Tableau 5.1

Exemple fictif de sortie révélant qu'un utilisateur est en présence d'un tableau dont au moins une cellule est jugée confidentielle au niveau de l'élève (symbole « ^{aaa} »)

Territoire	Variable de croisement	Statistiques					
		Nombre ^a	Proportion brute (%)	C.V. (%)	Erreur-type (%)	I.C. à 95 %	% non-réponse partielle
RLS A	Enfant vit avec ses deux parents	1 830	* 44,9	16,9	1,44	(43,1 - 48,8)	1,1
	Séparation avant l'âge de 18 mois (incluant avant la naissance)	330	52,7	12,8	6,73	(39,7 - 65,4)	1,1
	Séparation entre 18 mois et moins de 36 mois	^{aaa} 0	* 52,6	16,3	10,85	(28,0 - 68,1)	1,1
	Séparation à 36 mois ou plus	180	* 33,5	20,7	6,93	(21,5 - 48,1)	0,0

1. Population estimée (arrondie à la dizaine) des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et 11 mois. Puisque la non-réponse partielle a été répartie selon la proportion observée pour chaque catégorie de la variable de croisement (ou de la sous-population ou du territoire), la somme des estimations de population obtenues ne correspond pas nécessairement à l'estimation de population pour les catégories regroupées.

aaa : Le nombre de répondants servant à estimer les effectifs populationnels et la proportion dans cette cellule, ou dans la cellule complémentaire, est inférieur à 5. En cas de diffusion de ces données, il est nécessaire de procéder au masquage de données selon les normes établies dans le guide méthodologique de l'EQPPEM et dans la fiche-indicateur.

Source : Données fictives.

Le tableau 5.2 présente les symboles utilisés dans l'application du masquage, et le tableau 5.3 montre un exemple de celui-ci pour un indicateur de l'EQPPEM.

Tableau 5.2

Symboles utilisés dans l'application du masquage

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

Tableau 5.3

Exemple de masquage spécifique présenté dans la fiche de l'indicateur sur la proportion des enfants de maternelle 5 ans ayant été gardés régulièrement dans tous types de milieux de garde entre la naissance et la maternelle

Situation	Oui	Non
1	≥ 5	≥ 5
2	< 5	≥ 5
3	≥ 5	< 5

VI. Recommandations pour l'interprétation des résultats

Des recommandations inspirées des travaux produits dans le contexte des enquêtes de santé de l'Institut de la statistique du Québec ont été formulées à la partie III du *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le développement des enfants à la maternelle 2012* (2013). Ces recommandations sont reconduites pour l'EQPPEM. À celles-ci s'ajoute une recommandation portant sur les analyses régionales.

La taille des échantillons régionaux a été établie de manière à permettre la production de statistiques ayant une précision acceptable à l'échelle régionale pour les indicateurs d'intérêt dans le cadre de l'EQPPEM, même pour l'examen de petites proportions. Par contre, pour analyser les associations entre le parcours préscolaire et l'environnement familial ou encore pour tirer des conclusions concernant le lien entre le développement de l'enfant et son parcours préscolaire, les utilisateurs des données doivent savoir qu'il est préférable de travailler au niveau provincial. En effet, la plus faible puissance statistique de tests réalisés à l'échelle régionale pourrait avoir pour effet qu'on ne détecte pas d'association entre une variable d'analyse et une variable de croisement, bien qu'on observe une relation tout à fait similaire à celle détectée au niveau de la province.

Références bibliographiques

- ANDRICH, D., et I. STYLES (2004). *Final Report on the Psychometric Analysis of the Early Development Instrument (EDI) Using the Rasch Model: A Technical Paper Commissioned for the Development of the Australian Early Development Instrument (AEDI)*, Perth, Australie-Occidentale, Murdoch University, 62 p.
- BOIVIN, M., et K. L. BIERMAN (2014). "School Readiness: Introduction to a Multifaceted and Developmental Construct", dans BOIVIN, M., et K. L. BIERMAN (Eds), *Promoting School Readiness and Early Learning. Implications of Developmental Research for Practice*, New York, The Guilford Press, p. 3-14.
- BRINKMAN, S., et autres (2006). *Construct and Concurrent Validity of the Australian Early Development Index*, Parkville, Royal Children's Hospital, 23 p.
- CAMIRAND, H., I. TRAORÉ et J. BAULNE (2016). *L'Enquête québécoise sur la santé de la population, 2014-2015 : pour en savoir plus sur la santé des Québécois. Résultats de la deuxième édition*, [En ligne], Québec, Institut de la statistique du Québec, 208 p. [<http://www.stat.gouv.qc.ca/statistiques/sante/etat-sante/sante-globale/sante-quebecois-2014-2015.pdf>]
- CENTRE DE DONNÉES DE RECHERCHE DE STATISTIQUE CANADA (2005). *Guide à l'intention des chercheurs ayant conclu une entente avec Statistique Canada*, Ottawa, Statistique Canada, 46 p.
- CRESCO, S. (2008). *Annuaire de statistiques sur l'inégalité de revenu et le faible revenu. Édition 2008*, [En ligne], Québec, Institut de la statistique du Québec, 190 p. [www.stat.gouv.qc.ca/statistiques/conditions-vie-societe/revenu/annuaire-inegalite.pdf].
- DESROSIERS, H., K. TÊTREAU et M. BOIVIN (2012). « Caractéristiques démographiques, socioéconomiques et résidentielles des enfants vulnérables à l'entrée à l'école », *Portraits et trajectoires. Série Étude longitudinale du développement des enfants du Québec – ÉLDEQ*, [En ligne], n° 14, mai, Institut de la statistique du Québec, p. 1-12. [www.stat.gouv.qc.ca/statistiques/sante/bulletins/portrait-201205.pdf].
- DIRECTION DE SANTÉ PUBLIQUE DE L'AGENCE DE LA SANTÉ ET DES SERVICES SOCIAUX DE MONTRÉAL (DSP-ASSSM) [Québec] (2008). *En route pour l'école! Enquête sur la maturité scolaire des enfants montréalais. Rapport régional 2008*, [En ligne], Montréal, Direction de santé publique, Agence de la santé et des services sociaux de Montréal, 134 p. [https://santemontreal.qc.ca/fileadmin/user_upload/Uploads/tx_asssmpublications/pdf/publications/978-2-89494-630-5.pdf]
- ELTINGE, J. L., et I. S. YANSANEH (1997). « Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey », *Techniques d'enquête*, vol. 23, n° 1, p. 37-45.
- FORGET-DUBOIS, N., et autres (2007). "Predicting Early School Achievement with the EDI: A Longitudinal Population-Based Study", *Early Education and Development*, vol. 18, n° 3, p. 405-426.
- FOURNIER, M. (2010). *Les qualités métriques de l'IMDPE dans le cadre de l'Enquête sur la maturité scolaire des enfants montréalais*, Montréal, Direction de santé publique, Agence de la santé et des services sociaux de Montréal. [Document de travail non publié].

- HAZIZA, D., et J.-F. BEAUMONT (2007). "On the Construction of Imputation Classes in Survey", *International Statistical Review*, [En ligne], vol. 75, n° 1, p. 25-43. [econpapers.repec.org/article/blaitatr/v_3a75_3ay_3a2007_3ai_3a1_3ap_3a25-43.htm].
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (ISQ), en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC (INSPQ) (2018). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires. Plan national de surveillance de l'état de santé de la population et de ses déterminants. Enquête québécoise sur la santé des jeunes du secondaire 2016-2017*, Québec, Gouvernement du Québec. [<http://www.stat.gouv.qc.ca/enquetes/sante/eqsjs2016-2017-guide-methodologique.pdf>].
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (ISQ), en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC (INSPQ) (2013). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance – Enquête québécoise sur le développement des enfants à la maternelle 2012*, Québec, Gouvernement du Québec. [<http://www.stat.gouv.qc.ca/enquetes/sante/eqdem-guide-methodologique.pdf>].
- JANUS, M., et autres (2007). *The Early Development Instrument: A Population-Based Measure for Communities. A Handbook on Development, Properties, and Use*, [En ligne], Hamilton, Offord Centre for Child Studies, McMaster University, 44 p. [www.gov.mb.ca/healthychild/edi/edi_handbook_2007.pdf].
- JANUS, M., et D. R. OFFORD (2007). "Development and Psychometric Properties of the Early Development Instrument (EDI): A Measure of Children's School Readiness", *Canadian Journal of Behavioural Science*, [En ligne], vol. 39, n° 1, p. 1-22. [edi.offordcentre.com/wp/wp-content/uploads/2015/10/Janus-Offord-2007.pdf].
- KORN, E. L., et B. I. GRAUBARD (1999). *Analysis of Health Surveys*, New York, John Wiley & Sons, 382 p.
- LAPOINTE, F., K. THIBODEAU et L. GINGRAS (2019). *Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017. Méthodologie de l'enquête*, [En ligne], Québec, Institut de la statistique du Québec, 26 p. [www.stat.gouv.qc.ca/statistiques/education/prescolaire-primaire/eappem_methodologie.pdf].
- LAVOIE, A., L. GINGRAS et N. AUDET (2019). *Enquête québécoise sur le parcours préscolaire des enfants de maternelle 2017. Portrait statistique pour le Québec et ses régions administratives*, [En ligne], Québec, Institut de la statistique du Québec, Tome 1, 154 p. [www.stat.gouv.qc.ca/statistiques/education/prescolaire-primaire/eappem_tome1.pdf].
- LEMELIN, J.-P., et M. BOIVIN (2007). « Mieux réussir dès la première année : l'importance de la préparation à l'école », *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010)*, [En ligne], vol. 4, fascicule 2, décembre, Institut de la statistique du Québec, p. 1-12. [www.jesuisjeserai.stat.gouv.qc.ca/publications/Fasc2Vol4.pdf].
- LOHR, S. (1999). *Sampling: Design and Analysis*, 1^{re} éd., Duxbury Press, États-Unis.
- MINISTÈRE DE L'ÉDUCATION ET DE L'ENSEIGNEMENT SUPÉRIEUR (MEES) [Québec] (2017). *Politique de la réussite éducative. Le plaisir d'apprendre, la chance de réussir*, [En ligne], Québec, Gouvernement du Québec, 33 p. [www.education.gouv.qc.ca/fileadmin/site_web/documents/PSG/politiques_orientations/politique_reussite_educative_10juillet_F_1.pdf].

- MINISTÈRE DE L'ÉDUCATION, DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE (MEESR) [Québec] (2015). *Indicateurs de l'éducation. Éducation préscolaire, enseignement primaire et secondaire. Édition 2014*, [En ligne], Québec, Gouvernement du Québec, 84 p. [www.education.gouv.qc.ca/fileadmin/site_web/documents/PSG/statistiques_info_decisionnell/indicateurs_2014_fr.pdf].
- MINISTÈRE DE L'ÉDUCATION, DU LOISIR ET DU SPORT (MELS) [Québec] (2007). *L'organisation des services éducatifs aux élèves à risque et aux élèves handicapés ou en difficulté d'adaptation ou d'apprentissage (EHDAA)*, [En ligne], Québec, Gouvernement du Québec, 25 p. [www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/adaptation_serv_compl/19-7065.pdf].
- MINISTÈRE DE L'ÉDUCATION, DU LOISIR ET DU SPORT (MELS) [Québec] (2006). *Programme de formation de l'école québécoise. Version approuvée. Éducation préscolaire, enseignement primaire*, [En ligne], Québec, Gouvernement du Québec, 354 p. [www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/formation_jeunes/prform2001.pdf].
- MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX (MSSS) [Québec] (2016). *Politique gouvernementale de prévention en santé. Un projet d'envergure pour améliorer la santé et la qualité de vie de la population*, [En ligne], Québec, Gouvernement du Québec, 112 p. [publications.msss.gouv.qc.ca/msss/fichiers/2016/16-297-08W.pdf].
- MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX (MSSS) [Québec] (2011). *Initiative concertée d'intervention pour le développement des jeunes enfants 2011-2014. Projet d'enquête et d'intervention*, Québec, Gouvernement du Québec, 11 p.
- RUST, K. F., et J. N. K. RAO (1996). "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research*, vol. 5, n° 3, p. 283-310.
- SIMARD, M., A. LAVOIE et N. AUDET (2018). *Enquête québécoise sur le développement des enfants à la maternelle 2017*, [En ligne], Québec, Institut de la statistique du Québec, 126 p. [www.stat.gouv.qc.ca/statistiques/sante/enfants-ados/developpement-enfants-maternelle-2017.pdf].
- SIMARD, M., M.-E. TREMBLAY, A. LAVOIE et N. AUDET (2013). *Enquête québécoise sur le développement des enfants à la maternelle 2012*, [En ligne], Québec, Institut de la statistique du Québec, 105 p. [www.stat.gouv.qc.ca/statistiques/sante/enfants-ados/developpement-enfants-maternelle-2012.pdf].
- STATISTIQUE CANADA (2014). *Les lignes de faible revenu, 2012-2013*, [En ligne], produit n° 75F0002M (n° 3) au catalogue de Statistique Canada, Ottawa, Statistique Canada, 30 p. [www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2014003-fra.pdf].
- STATISTIQUE CANADA (2016). *Les lignes de faible revenu : leur signification et leur calcul*, [En ligne], produit n° 750002M (n° 2) au catalogue de Statistique Canada, Ottawa, Statistique Canada, 11 p. [www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2016002-fra.pdf].

TREMBLAY, M.-E., et M. SIMARD (2018). *Méthodologie de l'Enquête québécoise sur le développement des enfants à la maternelle 2017*, Québec, Institut de la statistique du Québec, 40 p.
[\[http://www.stat.gouv.qc.ca/statistiques/sante/enfants-ados/eqdem-rapport-methodologique-2017.pdf\]](http://www.stat.gouv.qc.ca/statistiques/sante/enfants-ados/eqdem-rapport-methodologique-2017.pdf)

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

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

Cette annexe expose l'algorithme d'une requête effectuée sur le portail de l'Infocentre de santé publique.

Tableau A2.1

Tableaux récapitulatifs des requêtes paramétrables sur le portail de l'Infocentre de santé publique
Niveau géographique : Ensemble du Québec

Sous-population	Sexe	Groupe d'âge	Première variable de croisement	Deuxième variable de croisement	Test statistique	Type de proportion ²
Totale			- Aucune			
			- Autres variables ¹	- Aucune	- Aucun	
				- Autres variables	- Indicateur selon var. de croisement	
			- Sexe	- Aucune	- Aucun	
				- Autres variables	- Indicateur selon var. de croisement	
			- Groupe d'âge	- Aucune	- Aucun	
				- Autres variables	- Indicateur selon var. de croisement	
			Sexe	- Par sexe - Masculin - Féminin		- Aucune
- Groupe d'âge	- Aucune	- Aucun				
	- Autres variables	- Indicateur selon var. de croisement				
- Autres variables	- Aucune	- Aucun				
	- Autres variables	- Indicateur selon var. de croisement				
Groupe d'âge		Groupe d'âge				- Aucune
			- Sexe	- Aucune	- Aucun	
				- Autres variables	- Indicateur selon var. de croisement	
			- Autres variables	- Aucune	- Aucun	
				- Autres variables	- Indicateur selon var. de croisement	

1. Il s'agit des autres variables de la fiche, à l'exception du sexe et du groupe d'âge.

2. Le fait qu'on ne détecte pas de différence entre les RSS (ou les regroupements de RLS) pour la distribution des enfants selon l'âge à l'admission à la maternelle, ni pour la distribution des enfants selon le sexe indique que la standardisation n'aurait pas (ou vraiment très peu) d'effet sur les estimations régionales (ou infrarégionales) produites.

Tableau A2.2

Niveau géographique : Régional

Région socio-sanitaire	Sous- population	Sexe	Groupe d'âge	Comparaison régionale	Région de comparaison	Variable de croisement	Test statistique	Type de proportion ²	
Choix d'une région	Totale			- Aucune		- Aucune			
						- Sexe	- Aucun		
							- Indicateur selon var. de croisement		
						- Autres variables ¹	- Aucun		
				- Indicateur selon var. de croisement					
				- Groupe d'âge	- Aucun				
	- Indicateur selon var. de croisement								
	- Région par rapport au reste du Québec								
		- Région par rapport à une autre région			Choix région de comparaison				
	Sexe	- Par sexe - Masculin - Féminin			- Aucune		- Aucune		
							- Autres variables	- Aucun	
								- Indicateur selon var. de croisement	
- Groupe d'âge							- Aucun		
					- Indicateur selon var. de croisement				
- Région par rapport au reste du Québec									
	- Région par rapport à une autre région			Choix région de comparaison					
Groupe d'âge			Groupe d'âge	- Aucune		- Aucune			
						- Sexe	- Aucun		
				- Autres variables	- Indicateur selon var. de croisement				
					- Aucun				
- Région par rapport au reste du Québec									
	- Région par rapport à une autre région			Choix région de comparaison					

1. Il s'agit des autres variables de la fiche, à l'exception du sexe et du groupe d'âge.

2. Le fait qu'on ne détecte pas de différence entre les RSS (ou les regroupements de RLS) pour la distribution des enfants selon l'âge à l'admission à la maternelle, ni pour la distribution des enfants selon le sexe indique que la standardisation n'aurait pas (ou vraiment très peu) d'effet sur les estimations régionales (ou infrarégionales) produites.

Tableau A2.3

Niveau géographique : Local (RLS par région)

Région socio-sanitaire	Sous- population	Sexe	Groupe d'âge	Comparaison locale	Variable de croisement	Test statistique	Type de proportion ²	
Choix d'une région	Totale			- Aucune	- Aucune	- Aucun		
					- Sexe	- Indicateur selon var. de croisement		
					- Autres variables ¹	- Indicateur selon var. de croisement		
					- Groupe d'âge	- Indicateur selon var. de croisement		
				- RLS par rapport au reste de la région (ou RLS par rapport au reste du Québec ³)				
	Sexe	- Par sexe - Masculin - Féminin			- Aucune	- Aucune	- Aucun	
						- Autres variables	- Indicateur selon var. de croisement	
						- Groupe d'âge	- Indicateur selon var. de croisement	
					- RLS par rapport au reste de la région (ou RLS par rapport au reste du Québec ³)			
	Groupe d'âge			Groupe d'âge	- Aucune	- Aucune	- Aucun	
						- Sexe	- Indicateur selon var. de croisement	
						- Autres variables	- Indicateur selon var. de croisement	
- RLS par rapport au reste de la région (ou RLS par rapport au reste du Québec ³)								

1. Il s'agit des autres variables de la fiche, à l'exception du sexe et du groupe d'âge.
2. Le fait qu'on ne détecte pas de différence entre les RSS (ou les regroupements de RLS) pour la distribution des enfants selon l'âge à l'admission à la maternelle, ni pour la distribution des enfants selon le sexe indique que la standardisation n'aurait pas (ou vraiment très peu) d'effet sur les estimations régionales (ou infrarégionales) produites.
3. Rappelons qu'il est recommandé de comparer un RLS au reste du Québec uniquement lorsque les RLS de la région socio-sanitaire correspondante diffèrent significativement entre eux. Pour ce faire, on peut utiliser le test global « RLS comparés avec le reste de la région socio-sanitaire de résidence ». Ainsi, en présence d'un résultat significatif à ce test global, le résultat du test de comparaison du RLS avec le reste du Québec peut être interprété.