

## Chapitre 10

### Aspects méthodologiques de l'enquête

---

#### 10.1 Plan de sondage

##### 10.1.1 Population visée et base de sondage

La population visée par la présente enquête est constituée des familles d'enfants âgés de moins de cinq ans au 30 septembre 2004 et résidant au Québec à l'exclusion :

- des familles vivant dans les régions sociosanitaires 17 (Nunavik) et 18 (Terres-Cries-de-la-Baie-James);
- des familles vivant dans les territoires non organisés (régions sans organisation municipale, administrées directement par la province ou par la circonscription électorale, caractérisées par la faible densité de leur population et l'étendue de leur territoire);
- des familles vivant dans les réserves amérindiennes.

Ces exclusions correspondent à environ 1 % des familles québécoises.

La base de sondage a été conçue à partir du fichier des bénéficiaires de la Régie de l'assurance maladie du Québec (RAMQ), dont la mise à jour est régulière. Ce fichier permet de cibler les familles dont au moins un de leurs enfants à moins de cinq ans, tout en fournissant l'information suivante : nombre d'enfants de moins de cinq ans, âge de l'aîné parmi les enfants de moins de cinq ans, sexe de l'aîné parmi les enfants de moins de cinq ans, âge de la mère, âge du père, nom des parents, prénom des parents, adresse des parents<sup>1</sup>, langue de correspondance et, dans certains cas, numéros de téléphone à domicile et au travail des parents. La base de sondage utilisée pour l'enquête correspond à la liste des familles ayant au moins un enfant âgé de moins de cinq ans au 30 septembre 2004 et dont les renseignements sur la famille apparaissent dans le fichier de la RAMQ en date du 23 août 2004, à l'exclusion des familles ayant un seul enfant de moins de un an. Les renseignements sur ces dernières familles proviennent du fichier de la RAMQ en date du 6 octobre 2004, afin de prendre en compte les inscriptions des enfants après le 23 août 2004. Le tableau 10.1 présente la répartition des familles visées par l'enquête selon la région administrative et la zone géographique (urbaine ou rurale<sup>2</sup>).

---

1. Lorsque connue, l'adresse de la mère est utilisée.

2. Le découpage de la zone géographique est expliqué à la section 10.1.2.

**Tableau 10.1**  
**Répartition des familles visées par l'enquête selon la région administrative et la zone urbaine ou rurale**

Région administrative	Zone géographique				Total	
	Urbaine		Rurale			
	n	%	n	%	n	%
Bas-Saint-Laurent (01)	3 618	1,5	3 186	5,5	6 804	2,3
Saguenay–Lac-Saint-Jean (02)	6 825	2,9	2 955	5,1	9 780	3,3
Capitale-Nationale (03) <sup>3</sup>	19 432	8,3	3 004	5,2	22 436	7,6
Mauricie (04)	5 771	2,5	1 923	3,3	7 694	2,6
Estrie (05)	6 751	2,9	4 711	8,1	11 462	3,9
Montréal (06)	78 084	33,2	0	0,0	78 084	26,6
Outaouais (07)	10 307	4,4	3 562	6,1	13 869	4,7
Abitibi-Témiscamingue (08)	2 747	1,2	2 880	5,0	5 627	1,9
Côte-Nord (09)	2 274	1,0	1 028	1,8	3 302	1,1
Nord-du-Québec (10)	627	0,3	82	0,1	709	0,2
Gaspésie–Îles-de-la-Madeleine (11)	741	0,3	2 170	3,7	2 911	1,0
Chaudière-Appalaches (12)	9 260	3,9	6 111	10,5	15 371	5,2
Laval (13)	15 313	6,5	0	0,0	15 313	5,2
Lanaudière (14)	11 260	4,8	4 680	8,1	15 940	5,4
Laurentides (15)	13 922	5,9	7 222	12,5	21 144	7,2
Montréal (16)	43 440	18,5	10 792	18,6	54 232	18,5
Centre-du-Québec (17)	4 993	2,1	3 648	6,3	8 641	2,9
<b>Total</b>	<b>235 365</b>	<b>100,0</b>	<b>57 954</b>	<b>100,0</b>	<b>293 319</b>	<b>100,0</b>

Source : Fichier des bénéficiaires de la Régie de l'assurance maladie du Québec, 11 novembre 2004.

Ce tableau indique que l'on compte 293 319 familles, dans l'ensemble du Québec, ayant des enfants de moins de cinq ans, inscrites au fichier des bénéficiaires de la RAMQ. La plupart de ces familles habitent en zone urbaine (235 365 en zone urbaine et 57 954 en zone rurale). La moitié des familles de la zone urbaine viennent de deux régions : celle de Montréal compte le tiers des familles (33,2 %) et celle de la Montérégie, 18,5 % des familles. La répartition des familles vivant en zone rurale est mieux distribuée entre les 15 régions qui en comportent une (Montréal et Laval n'en ont pas). On remarque cependant qu'un peu plus de 40 % des familles de la zone rurale résident en Montérégie (18,6 %), dans les Laurentides (12,5 %) ou dans la Chaudière-Appalaches (10,5 %). Il pourrait manquer un nombre infime de familles dans le tableau, soit celles qui ont eu un premier enfant avant le 30 septembre 2004 et pour lesquelles la RAMQ n'avait pas encore saisi la naissance du nouveau-né dans ses fichiers au moment de la compilation. Le tableau 10.2 présente la répartition des enfants visés par l'enquête selon la région administrative et la zone géographique (urbaine ou rurale).

3. La zone urbaine de la région de la Capitale-Nationale comprend la Communauté-Urbaine-de-Québec (CUQ) et les régions urbaines hors CUQ.

**Tableau 10.2**  
**Répartition des enfants âgés de moins de cinq ans au 30 septembre 2004 selon la région administrative et la zone urbaine ou rurale**

Région administrative	Zone géographique				Total	
	Urbaine		Rurale		n	%
	n	%	n	%		
Bas-Saint-Laurent (01)	4 442	1,5	4 065	5,5	8 507	2,3
Saguenay–Lac-Saint-Jean (02)	8 491	2,9	3 727	5,1	12 218	3,4
Capitale-Nationale (03) <sup>4</sup>	23 881	8,2	3 710	5,0	27 591	7,6
Mauricie (04)	7 126	2,5	2 420	3,3	9 546	2,6
Estrie (05)	8 521	2,9	6 110	8,3	14 631	4,0
Montréal (06)	95 103	32,8	0	0,0	95 103	26,1
Outaouais (07)	12 480	4,3	4 408	6,0	16 888	4,6
Abitibi-Témiscamingue (08)	3 409	1,2	3 660	5,0	7 069	1,9
Côte-Nord (09)	2 754	0,9	1 267	1,7	4 021	1,1
Nord-du-Québec (10)	799	0,3	100	0,1	899	0,2
Gaspésie–Îles-de-la-Madeleine (11)	869	0,3	2 574	3,5	3 443	0,9
Chaudière-Appalaches (12)	11 624	4,0	7 940	10,8	19 564	5,4
Laval (13)	18 832	6,5	0	0,0	18 832	5,2
Lanaudière (14)	13 929	4,8	5 913	8,0	19 842	5,5
Laurentides (15)	17 463	6,0	9 060	12,3	26 523	7,3
Montréal (16)	54 280	18,7	13 846	18,8	68 126	18,7
Centre-du-Québec (17)	6 308	2,2	4 731	6,4	11 039	3,0
<b>Total</b>	<b>290 311</b>	<b>100,0</b>	<b>73 531</b>	<b>100,0</b>	<b>363 842</b>	<b>100,0</b>

Source : Fichier des bénéficiaires de la Régie de l'assurance maladie du Québec, 11 novembre 2004.

Ce tableau indique que la répartition des 363 842 enfants selon la zone géographique suit le même modèle que celle des familles. Il y a plus d'enfants qui vivent en zone urbaine qu'en zone rurale (290 311 et 73 531). C'est dans les régions de Montréal et de la Montérégie que la moitié des enfants en zone urbaine ont leur domicile. En outre, un peu plus de 40 % des enfants en zone rurale habitent la Montérégie (18,8 %), les Laurentides (12,3 %) ou la Chaudière-Appalaches (10,8 %). Notons que ces données sont légèrement sous-estimées, car elles ne tiennent pas compte des enfants non inscrits dans le fichier des bénéficiaires au moment de l'extraction. De plus, les données agrégées de la RAMQ ne permettent pas d'établir de distinction entre les familles ayant trois enfants et les familles comptant plus de trois enfants âgés de moins de cinq ans au 30 septembre 2004.

### 10.1.2 Plan de sondage et stratification

Le plan de sondage de l'enquête est un plan stratifié non proportionnel à deux degrés. Le premier degré consiste en un échantillon stratifié non proportionnel de familles et le deuxième degré comprend tous les

4. La zone urbaine de la région de la Capitale-Nationale comprend la Communauté-Urbaine-de-Québec (CUQ) et les régions urbaines hors CUQ.

enfants de moins de cinq ans des familles échantillonnées. Les variables de stratification utilisées sont au nombre de trois :

1. La subdivision de la région administrative de la résidence des familles :

- Bas-Saint-Laurent (01) – zone urbaine;
- Bas-Saint-Laurent (01) – zone rurale;
- Saguenay–Lac-Saint-Jean (02) – zone urbaine;
- Saguenay–Lac-Saint-Jean (02) – zone rurale;
- Capitale-Nationale (03) à l'exclusion de la Communauté-Urbaine-de-Québec (CUQ)<sup>5</sup> – zone urbaine;
- Capitale-Nationale (03) à l'exclusion de la Communauté-Urbaine-de-Québec (CUQ)<sup>5</sup> zone rurale;
- Communauté-Urbaine-de-Québec (CUQ)<sup>5</sup>;
- Mauricie (04) – zone urbaine;
- Mauricie (04) – zone rurale;
- Estrie (05) – zone urbaine;
- Estrie (05) – zone rurale;
- Montréal (06) – zone 1;
- Montréal (06) – zone 2;
- Montréal (06) – zone 3;
- Outaouais (07) – zone urbaine;
- Outaouais (07) – zone rurale;
- Abitibi-Témiscamingue (08) – zone urbaine;
- Abitibi-Témiscamingue (08) – zone rurale;
- Côte-Nord (09) – zone urbaine;
- Côte-Nord (09) – zone rurale;
- Nord-du-Québec (10) – zone urbaine;
- Nord-du-Québec (10) – zone rurale;
- Gaspésie–Îles-de-la-Madeleine (11) – zone urbaine;
- Gaspésie–Îles-de-la-Madeleine (11) – zone rurale;
- Chaudière-Appalaches (12) – zone urbaine;
- Chaudière-Appalaches (12) – zone rurale;
- Laval (13);
- Lanaudière (14) – zone urbaine;
- Lanaudière (14) – zone rurale;

---

5. Comme défini géographiquement au moment de l'enquête de 2000-2001.

- Laurentides (15) – zone urbaine;
- Laurentides (15) – zone rurale;
- Montérégie (16) à l'exclusion de la municipalité régionale de comté (MRC) de Champlain<sup>6</sup> – zone urbaine;
- Montérégie (16) à l'exclusion de la municipalité régionale de comté (MRC) de Champlain<sup>6</sup> – zone rurale;
- Municipalité régionale de comté (MRC) de Champlain<sup>6</sup>;
- Centre-du-Québec (17) – zone urbaine;
- Centre-du-Québec (17) – zone rurale.

2. Le nombre d'enfants de moins de cinq ans :

- un enfant;
- deux enfants;
- trois enfants ou plus.

3. L'âge de l'aîné parmi les enfants de moins de cinq ans :

- moins de un an;
- de un an à moins de deux ans;
- de deux ans à moins de trois ans;
- de trois ans à moins de quatre ans;
- de quatre ans à moins de cinq ans.

Le découpage du Québec selon les 36 subdivisions des régions administratives de résidence décrites précédemment est plus détaillé que celui dont on s'est servi pour les enquêtes de 1998<sup>7</sup> et de 2000-2001<sup>8</sup>. La nécessité d'obtenir des données relatives à des régions géographiques plus désagrégées a justifié ce nouveau découpage. Ainsi, les régions administratives ont été divisées en deux selon la zone géographique, soit urbaine ou rurale (à l'exception de Laval et Montréal qui ne comportent pas de zone rurale selon la méthode utilisée ici). Les régions administratives de la Capitale-Nationale, de Montréal et de la Montérégie ont, quant à elles, été subdivisées en trois. Le découpage géographique effectué lors des enquêtes précédentes (par taille de municipalité) a tout de même été reconstitué afin de s'assurer

---

6. Comme défini géographiquement au moment de l'enquête de 2000-2001

7. Bureau de la statistique du Québec (1999). Enquête sur les besoins des familles en matière de services de garde. Rapport d'analyse descriptive, mars, 312 p.

8. Institut de la statistique du Québec (2001). Rapport d'enquête sur les besoins des familles en matière de services de garde. Rapport d'analyse descriptive, septembre, 473 p.

que les résultats de l'application du modèle – permettant d'estimer les places à contribution réduite requises en service de garde – soient comparables d'une enquête à l'autre.

Le découpage des régions administratives selon la zone urbaine ou rurale a été effectué sur la base des codes postaux et à partir des municipalités. Ainsi, le Fichier de conversion des codes postaux (FCCP) produit par Statistique Canada, en collaboration avec la Société canadienne des postes, a été utilisé pour établir le lien entre le code postal, le secteur de dénombrement (plus petite unité géographique pour laquelle existent des données et qui est définie aux fins de collecte des données du recensement) et la zone urbaine ou rurale de chaque famille.

Code postal → Secteur de dénombrement → Zone urbaine/rurale

Ainsi, dans un premier temps, on a utilisé le FCCP basé sur la géographie du Recensement de 1996, avec la mise à jour des codes postaux en 2001. Ensuite, pour tenir compte de l'apparition de nouveaux codes postaux, on s'est servi du fichier basé sur la géographie du Recensement de 2001, avec la mise à jour des codes postaux en juin 2003. Pour ce qui est des codes postaux apparus entre juin 2003 et juillet 2004 (moment du tirage de l'échantillon), on a dû procéder, en raison de l'absence de la zone géographique, à l'imputation de la zone urbaine ou rurale selon certaines hypothèses.

Le choix de cette approche appelle quelques réserves. Ainsi, il peut arriver qu'un code postal soit rattaché à plus d'un secteur de dénombrement (c'est ce dernier qui est considéré comme urbain ou rural). On conserve alors dans le fichier le meilleur appariement entre ces deux variables. Dans d'autres cas, le code postal du fichier de la RAMQ ne correspond pas au lieu de résidence des familles sélectionnées. En effet, celles-ci peuvent utiliser une boîte ou un casier postal pour recevoir leur courrier. Toutefois, cette distinction ne devrait pas causer trop d'incohérence en matière d'attribution d'une zone géographique, puisqu'il est fort probable que la boîte ou le casier postal se situe dans la même zone. De plus, au Québec, très peu de codes postaux chevauchent deux régions administratives. Lorsque cela se produit, on retient le lien le plus probable (c'est-à-dire la région administrative pour laquelle le code postal compte le plus de population).

Afin de s'assurer que les résultats de l'application du modèle permettant d'estimer les places à 7 \$ requises en service de garde soient comparables d'une enquête à l'autre, il fallait conserver la même géographie qu'au moment de l'enquête de 2000-2001, ce qui explique notamment la création de la

subdivision Communauté-Urbaine-de-Québec (CUQ) dans la Capitale-Nationale et de la Municipalité régionale de comté (MRC) de Champlain dans la Montérégie.

La région administrative de Montréal a été subdivisée en trois zones distinctes. La délimitation de ces zones a été réalisée à partir du classement des territoires de CLSC, selon l'indice croissant de défavorisation produit par le Conseil scolaire de l'Île-de-Montréal; les caractéristiques socioéconomiques des familles habitant sur les territoires de CLSC de la région ont été utilisées pour délimiter les trois zones.

Établie par le ministère de la Santé et des Services sociaux (MSSS) et mise à jour en avril 2004, la relation entre le code postal et le territoire de CLSC réclame aussi une réserve. Ainsi, le code postal peut parfois apparaître dans deux territoires de CLSC. Dans ce cas, on a attribué au code postal le CLSC qui dessert la population la plus importante.

On a pris en compte les deux autres variables de stratification (nombre d'enfants de moins de cinq ans et âge de l'aîné parmi les enfants de moins de cinq ans), car elles sont potentiellement liées au choix du mode de garde. La base de sondage utilisée permet de tenir compte des trois variables de stratification.

L'annexe C présente la répartition de la population et de l'échantillon entre les 319 strates créées par le croisement des trois variables de stratification.

De façon détaillée, voici la composition des trois zones de Montréal selon les territoires de CLSC :

<b>Zone 1 Indice de défavorisation le plus bas</b>	<b>Zone 2 Zone intermédiaire</b>	<b>Zone 3 Indice de défavorisation le plus élevé</b>
Lac-Saint-Louis Mont-Royal Dollard-des-Ormeaux Pierrefonds	Plateau Mont-Royal Pointe-aux-Trembles Côte-Saint-Luc Rivière-des-Prairies Métro Vieux-Lachine Mercier-Est LaSalle Verdun Anjou Ahuntsic Montréal/Centre-Ville Saint-Louis-du-Parc Notre-Dame-de-Grâce/ Montréal-Ouest Mercier-Ouest Rosemont Saint-Léonard Saint-Laurent La Petite Patrie Villeray Bordeaux/Cartierville	Côte-Saint-Paul Côte-des-Neiges Montréal-Nord Snowdon Hochelaga-Maisonneuve Saint-Michel Montréal/Centre-Sud Parc Extension Pointe Saint-Charles Saint-Henri

### 10.1.3 Détermination de la taille et répartition de l'échantillon

La taille de l'échantillon a été fixée de façon à obtenir une marge d'erreur maximale de 1 % quant aux proportions des préférences des familles pour chaque groupe d'âge, dans tout le Québec, et d'environ 5 % pour ce qui est de chacune des subdivisions des régions administratives. La détermination de la taille de l'échantillon dépend du taux de réponse anticipé de 65 %<sup>9</sup> et d'un effet de plan attribuable à la non-proportionnalité de la répartition de l'échantillon.

Ces objectifs de précision étaient suffisants pour s'assurer d'obtenir un minimum de 600 familles échantillonnées par taille de municipalité<sup>10</sup>, ce qui était essentiel pour répondre aux exigences du modèle élaboré par le ministère de la Famille, des Aînés et de la Condition féminine (MFACF).

Ainsi, on a privilégié une stratification non proportionnelle relativement aux subdivisions des régions administratives de résidence. L'évaluation de l'effet du plan de sondage permet d'estimer l'impact de la non-proportionnalité, afin d'évaluer la taille de l'échantillon nécessaire pour obtenir la précision souhaitée. L'effet de plan illustre donc l'efficacité d'une taille d'échantillon au regard de la précision des résultats. À titre d'exemple, un effet de plan égal à deux veut dire que la même précision pourrait être obtenue avec un échantillon aléatoire simple deux fois moins grand. L'effet de plan attribuable à cette non-proportionnalité dans la présente enquête est estimé à 2,1 quant aux estimations faites sur la base de l'ensemble des familles ayant au moins un enfant âgé de moins de cinq ans au 30 septembre 2004.

Par contre, on a favorisé une stratification proportionnelle en ce qui regarde l'âge de l'aîné, afin que les enfants de familles échantillonnées appartiennent, en nombre relativement équivalent, aux cinq groupes d'âge mentionnés précédemment.

En conséquence, la taille initiale de l'échantillon a été fixée à 20 594 familles.

---

9. Aux fins de calcul, le taux de 65 % a été utilisé comme hypothèse pour l'ensemble des strates.

10. Découpage géographique utilisé lors des enquêtes précédentes.

## 10.2 Questionnaire et résultats de la collecte

### 10.2.1 Questionnaire et prétest

Le questionnaire utilisé pour la présente enquête s'inspire fortement de celui qui a servi à l'enquête sur les besoins des familles en matière de services de garde éducatifs de 2000-2001. La version du questionnaire prétesté a été élaborée par les représentants du ministère de la Famille, des Aînés et de la Condition féminine (MFACF), en étroite collaboration avec l'Institut de la statistique du Québec (ISQ). L'annexe A présente la version française et la version anglaise du questionnaire.

Le questionnaire comporte les huit sections suivantes :

Section A Information générale sur la garde des enfants

Section B Utilisation régulière des services de garde en raison du travail ou des études des parents

Section C Utilisation régulière des services de garde en raison d'un autre motif que le travail ou les études des parents

Section D Utilisation irrégulière des services de garde en raison du travail ou des études des parents

Section E Utilisation irrégulière des services de garde en raison d'un autre motif que le travail ou les études des parents

Section F Amélioration des services de garde

Section G Préférences des parents en matière de services de garde

Section H Caractéristiques sociodémographiques

Le prétest a été effectué du 25 mai au 4 juin 2004 inclusivement. Un échantillon constitué de 200 familles d'enfants âgés de moins de cinq ans au 31 mai 2004 a été sélectionné aléatoirement dans quelques régions du Québec (Montréal, Outaouais et Chaudière-Appalaches) à partir du fichier des bénéficiaires de la Régie de l'assurance maladie du Québec (RAMQ).

L'objectif fondamental du prétest consistait à évaluer les aspects suivants :

- la clarté des questions et leur compréhension;
- les taux de réponse et de collaboration susceptibles d'être atteints lors de l'enquête;
- la durée de l'entrevue téléphonique.

Les entrevues téléphoniques ont été réalisées à l'aide d'un logiciel ITAO (interview téléphonique assistée par ordinateur). Préalablement à la collecte téléphonique, les répondants avaient été prévenus par une lettre. Trois lettres différentes ont été préparées en fonction des renseignements disponibles pour joindre le répondant sélectionné. Dans le cas du prétest, les répondants sélectionnés pour l'enquête ont tous reçu une première lettre expliquant les objectifs de l'enquête et les prévenant qu'ils seraient contactés par téléphone dans les jours suivants. Dans le cas des familles dont on ignorait le numéro de téléphone, une lettre leur a été expédiée une semaine plus tard pour les inciter à prendre contact avec l'ISQ pour remplir le questionnaire. Enfin, cinq jours avant la fin des activités de collecte, une lettre de rappel a été envoyée pour les inciter, une deuxième fois, à prendre contact avec l'ISQ. Les lettres servant à établir un premier contact avec les répondants étaient cosignées par la sous-ministre adjointe à la direction générale des politiques familiales du ministère de l'Emploi, de la Solidarité sociale et de la Famille (MESSF) et du directeur général de l'ISQ.

En ce qui a trait au prétest, il est important de noter que le fichier de la RAMQ ne contenait aucun numéro de téléphone. Une recherche a donc été effectuée à partir des coordonnées postales de la mère et du père (ou de l'un des deux) pour trouver les numéros de téléphone. Ainsi, les recherches ont permis d'obtenir 80 % des numéros de téléphone dont 5 % se sont révélés faux.

Au total, 123 questionnaires ont été remplis lors du prétest. Le taux de réponse se situe à 61,8 %, tandis que le taux de collaboration atteint 94,6 %. La durée moyenne d'une entrevue complète est d'environ 24 minutes; celle-ci varie en fonction du nombre d'enfants dans la famille, du type de services de garde utilisés et de la région de résidence de la famille. Le temps moyen d'entrevue à Montréal a été d'environ deux minutes de plus, compte tenu de la moins bonne maîtrise de la langue française ou anglaise d'un certain nombre de répondants. Une moyenne de 5 appels a été faite par dossier complet, le maximum de tentatives étant de 21. Le prétest a démontré que la collecte téléphonique, précédée d'une lettre de présentation pour décrire les objectifs du mandat, se prêtait très bien à l'enquête.

Dans l'ensemble, le questionnaire s'administre bien; toutefois, quelques modifications y ont été apportées et quelques questions retirées. Des exemples ou des précisions ont été ajoutés à certaines questions afin

de les clarifier. De même, l'enchaînement des questions de la section G sur les préférences des parents en matière de services de garde a été modifié, afin d'accélérer l'administration du questionnaire et de réduire l'impatience causée par le caractère répétitif des questions.

### **10.2.2 Déroulement de la collecte**

La collecte s'est déroulée du 14 septembre au 12 novembre 2004. Les lettres décrites précédemment dans la sous-section portant sur le questionnaire et le prétest ont également été utilisées pour l'enquête. Ces lettres apparaissent à l'annexe A.

Les entrevues téléphoniques ont été réalisées à l'aide d'un logiciel ITAO (interview téléphonique assistée par ordinateur) et effectuées tous les jours de la semaine, tant le jour que le soir, entre 8 h 30 et 21 h la semaine, et entre 9 h 30 et 17 h la fin de semaine. Au total, 4,2 tentatives, en moyenne, ont été effectuées pour joindre la personne sélectionnée au sein du ménage pour mener l'entrevue.

Tout comme lors du prétest, l'échantillon utilisé pour la collecte provenait du fichier des bénéficiaires de la RAMQ. Par contre, les numéros de téléphone des parents apparaissaient, lorsque disponibles, dans le fichier. Ainsi, au départ, de l'information sur le numéro de téléphone de l'un ou des deux parents était accessible relativement à environ 81 % des familles. Une recherche a donc été effectuée dans le cas de 19 % des familles ainsi que pour celles dont le numéro de téléphone s'est avéré erroné en cours d'enquête.

Bien que les entrevues se soient bien déroulées, des explications ont dû être fournies aux répondants pour les accompagner tout au long du questionnaire, les aider à bien saisir les nuances entre la garde régulière et la garde irrégulière ainsi qu'entre les motifs de garde. Quelques expressions comme « enclins à » ou « prévisible » se sont avérées un peu plus difficiles à saisir pour les parents moins scolarisés ou ceux dont la langue première n'est pas le français ou l'anglais.

### **10.2.3 Résultats de la collecte**

Pour mener cette enquête, on a constitué un échantillon de 20 594 familles comptant un ou plusieurs enfants âgés de moins de cinq ans au 30 septembre 2004 et résidant au Québec. Au total, 14 389

questionnaires ont été remplis. Le taux de réponse global<sup>11</sup> à la collecte s'établit à 70,1 %. Ce taux découle du ratio du nombre total d'entrevues complètes par rapport à l'échantillon admissible. Quant au taux de collaboration, il est de 94 %. Il correspond au rapport entre les entrevues complètes et le nombre de familles avec lesquelles un contact a été établi. Le tableau 10.3 présente les résultats de la collecte.

**Tableau 10.3**  
**Résultats globaux de la collecte**

	<b>Total</b>
Échantillon initial (1)	20 594
Familles non admissibles	76
Pas d'enfants de moins de cinq ans ou enfants à la maternelle	19
Hors Québec	52
Autres	5
Échantillon admissible (2)	20 518
Entrevues complètes (3)	14 389
Refus (4)	925
<b>Taux de réponse (3) / (2)</b>	<b>70,1 %</b>
<b>Taux de collaboration (3) / ((3) + (4))</b>	<b>94,0 %</b>

Les modifications apportées à la façon de poser les questions de la section G portant sur les préférences des parents en matière de services de garde, ainsi que l'habileté et l'expérience acquises par les intervieweurs, ont permis de faire ramener à 17,55 minutes le temps moyen nécessaire pour conduire une entrevue, comparativement aux 24 minutes du prétest.

### 10.3 Pondération

La pondération a pour but d'associer à une famille répondante le nombre d'unités (le poids) qu'elle représente dans la population, ce qui permet de rapporter les données des répondants à la population visée.

11. Il s'agit du taux non pondéré de réponse. Le taux pondéré de réponse, qui tient compte de la probabilité d'être sélectionné dans l'échantillon et de l'ajustement pour la non-réponse, est présenté dans la section sur la pondération.

L'enquête comporte trois étapes de pondération : le calcul des poids initiaux correspondant à l'inverse de la probabilité de sélection d'une famille dans sa strate, l'ajustement des poids pour tenir compte de la non-réponse et la poststratification.

### 10.3.1 Probabilité de sélection

Comme l'étude repose sur un échantillon probabiliste, elle permet d'évaluer, pour chaque famille de la population, la probabilité de faire partie de l'échantillon. L'inverse de la probabilité de sélection est utilisé comme pondération initiale. Cette première étape de pondération permet de tenir compte de la non-proportionnalité de l'échantillon de l'étude par rapport à la distribution de la population.

La probabilité de sélection d'une famille appartenant à la strate  $i, j, k$  est donnée par :

$$\pi(ijk) = \frac{n_{ijk}}{N_{ijk}}$$

où

$i$  = la région de résidence (1 à 36);

$j$  = le nombre d'enfants de moins de cinq ans (1 à 3);

$k$  = l'âge de l'aîné parmi les enfants de moins de cinq ans (1 à 5);

$n_{ijk}$  = le nombre de familles échantillonnées de la strate  $i, j, k$ ;

$N_{ijk}$  = le nombre de familles dans la population de la strate  $i, j, k$ .

Ainsi, la pondération initiale est donnée par l'inverse de la probabilité de sélection :

$$P_{0f}(ijk) = \frac{1}{\pi(ijk)}, \forall i, j, k.$$

Étant donné que tous les enfants de moins de cinq ans participaient à l'enquête par l'intermédiaire de leurs familles, la pondération initiale des enfants est identique à celle des familles :

$$P_{0e}(ijk) = P_{0f}(ijk), \forall i, j, k$$

### 10.3.2 Non-réponse

Le taux de réponse est un élément important quant à la qualité des résultats d'une enquête, car il indique la présence potentielle de biais dans les résultats. En effet, dans toute enquête, il est possible que les non-répondants possèdent des caractéristiques différentes de celles des répondants. Ainsi, plus la non-réponse est élevée, plus considérables sont les risques que des biais soient introduits dans les estimations inférées à l'ensemble de la population à partir des réponses reçues.

L'ajustement pour la non-réponse consiste à redresser l'échantillon des répondants par une modification de la pondération afin de le rendre, dans la mesure du possible, semblable à l'échantillon tiré initialement. Cette technique nécessite de l'information complémentaire sur les répondants et les non-répondants. Pour que l'ajustement effectué soit efficace, il est important que l'information auxiliaire dont on dispose soit liée aux variables mesurées par l'enquête, à défaut de quoi l'incidence de l'ajustement sera négligeable sur la réduction d'un biais potentiel.

Ainsi, l'ajustement s'appuie sur la création de groupes homogènes à l'aide de variables provenant du fichier des bénéficiaires de la RAMQ. Ces variables, dites « administratives », sont l'âge de la mère, la langue de correspondance et la taille de la municipalité. Le processus d'ajustement a conduit à la formation de 13 groupes homogènes dans tout l'échantillon. Le tableau 10.4 illustre la variation du taux de réponse en fonction des variables en question.

L'ajustement pour la non-réponse chez les familles s'exprime par un facteur de pondération, obtenu par l'inverse du taux de réponse  $T_g$  pour chaque groupe homogène de pondération  $g$ .

**Tableau 10.4**  
**Taux de réponse pondéré des trois variables administratives et taux de réponse pondéré global**

Variable	Catégorie	Taux de réponse %
Taille de la municipalité	Île de Montréal – zone 1	63,1
	Île de Montréal – zone 2	66,7
	Île de Montréal – zone 3	60,2
	Rive-Sud de Montréal (MRC de Champlain)	68,7
	Laval-Île Jésus	69,7
	Communauté-urbaine-de-Québec (CUQ)	68,2
	Très petites municipalités	70,7
	Petites municipalités	70,8
	Moyennes municipalités	71,4
	Grandes municipalités	71,1
	Très grandes municipalités	69,3
Âge de la mère	Moins de 20 ans	54,8
	20 ans ≤ âge < 25 ans	56,3
	25 ans ≤ âge < 30 ans	67,7
	30 ans ≤ âge < 35 ans	71,2
	35 ans ≤ âge < 40 ans	71,9
	40 ans ou plus	72,7
Langue de correspondance	Français	70,4
	Anglais	60,1
<b>Taux de réponse pondéré globale</b>		<b>68,9</b>

Le taux de réponse  $T_g$  est défini par la somme pondérée des unités répondantes sur la somme pondérée des unités admissibles :

$$T_g = \frac{\sum_{i,j,k \in g} P_{0f}(ijk) \cdot R(ijk)}{\sum_{i,j,k \in g} P_{0f}(ijk) \cdot A(ijk)}$$

$$R(ijk) = \begin{cases} 1 & \text{si l'unité est répondante} \\ 0 & \text{sinon} \end{cases}$$

où

$$A(ijk) = \begin{cases} 1 & \text{si l'unité est admissible} \\ 0 & \text{sinon} \end{cases}$$

Ainsi, chaque famille répondante, de la strate  $i, j, k$ , et du groupe  $g$ , se voit attribuer un poids  $P_{1f}$  égal à :

$$P_{1f}(ijk g) = \frac{P_{0f}(ijk)}{T_g}$$

L'ajustement pour la non-réponse chez les enfants est identique à celui des familles :

$$P_{1e}(ijk g) = P_{1f}(ijk g).$$

### 10.3.3 Poststratification

La poststratification consiste à redresser l'échantillon des répondants en modifiant la pondération, afin de le rendre similaire à la population visée. Ainsi, on découpe la population en groupes (aussi appelés « poststrates ») dans lesquels les réponses au questionnaire sont homogènes. Tout comme la stratification, la poststratification augmente la précision des estimations; elle contribue simultanément à diminuer le biais de la non-réponse et à corriger pour la sous-couverture.

La variable qui a servi à l'ajustement est celle qui donne la distribution du nombre de familles selon le nombre d'enfants âgés de moins de cinq ans au 30 septembre 2004 et l'âge de l'aîné des enfants de moins de cinq ans. Les données utilisées proviennent d'une compilation spéciale effectuée par la RAMQ à partir du fichier des bénéficiaires. Cette compilation a été effectuée le 11 novembre 2004.

La correction pour la poststratification est donc appliquée aux familles ayant des enfants âgés de moins de cinq ans au 30 septembre 2004 selon le facteur de correction suivant :

$$T_t = \frac{W(t)}{\sum_{i,j,k,g \in t} P_{1f}(ijk g)}$$

où

$W(t)$  représente la taille de la population visée pour la poststrate  $t$ ;

$\sum_{i,j,k,g \in t} P_{1f}(ijk g)$  représente la somme pondérée des unités répondantes sur chaque poststrate

suite à l'ajustement pour la non-réponse.

Le poids ajusté pour les familles devient donc :

$$P_{2f} = P_{1f}(ijk g) \cdot T_t, \forall i, j, k, g \in t$$

La correction pour la poststratification chez les enfants est basée sur celle qui a été effectuée pour les familles; elle est appliquée à tous les enfants d'un même ménage.

Le poids ajusté pour les enfants devient donc :

$$P_{2e} = P_{1e}(ijk g) \cdot T_t, \forall i, j, k, g \in t$$

## 10.4 Types d'estimations réalisées

Comme le questionnaire portait en grande partie sur des données qualitatives, les estimations se font majoritairement au niveau des proportions. Ainsi, la proportion de familles (ou d'enfants) possédant une caractéristique est estimée en attribuant à chaque répondant (parent ou enfant) la valeur 1 s'il possède la caractéristique mesurée, ou 0 dans le cas contraire. La proportion est donc un ratio de deux sommes pondérées : le numérateur est la somme pondérée des répondants ayant la caractéristique et le dénominateur est la somme pondérée de tous les répondants. Dans le cas de certaines questions ciblées, on a procédé à l'estimation d'un total ou d'une moyenne. Le total est donc une somme pondérée des valeurs fournies par les répondants, tandis que la moyenne est un ratio de deux sommes pondérées : le numérateur est une somme pondérée des valeurs fournies par le répondant et le dénominateur est la somme pondérée de tous les répondants.

### 10.4.1 Présentation des estimations

Au cours de cette enquête, on a procédé à deux types d'analyses : des analyses bivariées et des analyses multivariées.

#### 10.4.1.1 Analyses bivariées

Les analyses bivariées sont celles où chaque question est mise en relation avec un seul facteur d'intérêt à la fois. Plusieurs variables de croisement ont été utilisées pour produire les tableaux : l'âge des enfants, le sexe des enfants, la région administrative, la subdivision des régions administratives (pour les besoins de ces analyses, la CUQ a été réunie à la zone urbaine de la Capitale-Nationale et la MRC de Champlain a été amalgamée à la zone urbaine de la Montérégie) et quelques variables sociodémographiques (le

revenu, l'organisation de la famille, le lieu de naissance, la principale occupation, le diplôme, la zone géographique et l'atypisme de l'emploi). Les catégories des variables de croisement peuvent être adaptées en raison de critères de confidentialité.

Les estimations produites pour le présent rapport sont présentées sous forme de tableaux croisés dans lesquels les colonnes représentent les modalités d'une variable d'analyse et les lignes montrent la répartition de cette question selon un facteur d'intérêt (par exemple, l'âge des enfants ou la région administrative). Les marges d'erreur (voir la section sur les erreurs dues à l'échantillonnage) ont été calculées, selon le plan de sondage, à l'aide des logiciels SAS et SUDAAN. Pour tous les tableaux croisés, on a effectué un test du Chi-carré avec l'ajustement de Satterthwaite. Cet ajustement tient compte de la corrélation entre les unités. Le seuil de signification statistique a été fixé à 5 % pour tous les tests. Dans chacun des tableaux, la population de référence visée est identifiée par l'énoncé « Base » sous le titre du tableau. Le tableau 10.5 illustre le gabarit utilisé.

La colonne « Total » contient les renseignements relatifs à l'ensemble des modalités de la variable d'analyse. Les autres colonnes font référence à chaque modalité que prend cette variable. Chacune d'entre elles est subdivisée en cinq autres colonnes et a comme signification :

- % : la proportion estimée dans la population qui possède une certaine caractéristique;
- BI : la borne inférieure de l'intervalle de confiance associée à l'estimation de la proportion;
- BS : la borne supérieure de l'intervalle de confiance associée à l'estimation de la proportion;
- n : le nombre de répondants associé à l'estimation de la proportion;
- ct : la cote de précision associée à l'estimation de la proportion. Sous le tableau apparaît une légende de la cote de précision associée à la proportion estimée.

La ligne « Ensemble du Québec » représente l'ensemble de la population, toutes les modalités de la variable de croisement confondues. Les autres lignes font référence à chaque modalité que prend la variable de croisement.

**Tableau 10.5**  
**EXEMPLE : Extrait du tableau 1.1**

*QUESTION 9*

*QUELLE SITUATION CORRESPOND LE MIEUX À L'ORGANISATION ACTUELLE DE VOTRE FAMILLE ?*

*(BASE : TOUTES LES FAMILLES)*

	Famille monoparentale					Couple avec enf. union actuelle					Total					
	%	BI	BS	n	ct	%	BI	BS	N	ct	%	BI	BS	n	ct	
<b>RÉGIONS ADMINISTRATIVES</b>																
<b>Ensemble du Québec</b>	<b>11.2</b>	10.4	12.0	1403	A	<b>80.3</b>	79.3	81.2	11613	A	...	<b>100.0</b>	0.0	.	14358	A
<b>Bas-Saint-Laurent (01)</b>	<b>6.5</b>	4.9	8.4	50	C	<b>85.2</b>	82.8	87.5	702	A	...	<b>100.0</b>	0.0	.	823	A
<b>Saguenay-Lac-Saint-Jean (02)</b>	<b>8.4</b>	6.4	10.7	63	C	<b>84.1</b>	81.5	86.8	707	A	...	<b>100.0</b>	0.0	.	836	A
<b>Capitale-Nationale (03)</b>	<b>10.1</b>	7.7	12.4	92	C	<b>82.1</b>	79.2	85.0	1309	A	...	<b>100.0</b>	0.0	.	1230	A
<b>Mauricie (04)</b>	<b>17.4</b>	14.4	20.3	116	B	<b>69.6</b>	65.7	73.4	613	A	...	<b>100.0</b>	0.0	.	823	A
: : :	:	:	:	:	:	:	:	:	:	:	...	:	:	:	:	:
: : :	:	:	:	:	:	:	:	:	:	:	...	:	:	:	:	:
<b>Nord-du-Québec (10)</b>	<b>8.2</b>	6.1	10.8	27	C	<b>81.2</b>	77.6	84.4	282	A	...	<b>100.0</b>	0.0	.	345	A
<b>Gaspésie--Îles-de-la-Madeleine (11)</b>	<b>13.4</b>	10.8	16.0	90	B	<b>78.2</b>	75.1	81.3	542	A	...	<b>100.0</b>	0.0	.	689	A
<b>Chaudière-Appalaches (12)</b>	<b>8.1</b>	6.4	10.2	65	C	<b>84.4</b>	82.0	86.9	758	A	...	<b>100.0</b>	0.0	.	893	A
: : :	:	:	:	:	:	:	:	:	:	:	...	:	:	:	:	:

*LE SEUIL OBSERVÉ DU CROISEMENT (CHI-CARRÉ DE SATTERTHWAITE) EST: 0.00*  
*\*\*\* CE CROISEMENT PRÉSENTE UN LIEN SIGNIFICATIF \*\*\**

*COTES: A: EXCELLENTE PRÉCISION (CV <=5%) B: TRÈS BONNE PRÉCISION (5% < CV <=10%) C: BONNE PRÉCISION (10% < CV <=15%)*  
*D: PRÉCISION PASSABLE (15% < CV <=25%) E: FAIBLE PRÉCISION, À UTILISER AVEC CIRCONSPÉCTION (CV >25%)*  
*BORNES: BI: BORNE INFÉRIEURE BS: BORNE SUPÉRIEURE*

Plusieurs données ont été recueillies pour chacun des enfants âgés de moins de cinq ans au 30 septembre 2004 au sein de la famille. Les données de l'enquête permettent donc de fournir de l'information en se référant à l'univers des familles ou à celui des enfants.

Les estimations effectuées pour cette enquête atteignent pleinement les objectifs de précision fixés initialement. En effet, la précision sur les erreurs maximales en ce qui a trait aux préférences des familles pour chaque classe d'âge (0 à moins d'un 1 an, 1 an, 2 ans, 3 ans et 4 ans) et pour chacune des subdivisions des régions administratives sont d'environ 1 % et 5 % respectivement.

#### 10.4.1.2 Rapport d'analyse descriptive bivariée

La méthodologie utilisée pour la rédaction de ce rapport d'analyse descriptive bivariée est la suivante : généralement, les résultats de chacune des questions sont présentés pour l'ensemble de la population.

Par la suite, si le résultat du test du Chi-carré dénote une association entre la variable d'analyse et la variable de croisement, les résultats sont examinés plus attentivement. Pour les modalités d'intérêt de la variable d'analyse, on dégage ensuite les modalités de la variable de croisement qui se démarquent (par une incidence plus forte ou plus faible) des autres modalités en comparant les intervalles de confiance. Pour une caractéristique d'une variable d'analyse donnée, une différence est exprimée entre deux modalités de la variable de croisement si les intervalles de confiance respectifs à chacune des modalités comparées sont disjoints.

Ainsi, il est possible que le résultat d'une variable de croisement avec une variable d'intérêt ne soit pas discuté. En effet, il se peut que le test ne présente pas de lien significatif, que la précision des estimations soit passable ou faible, que l'information soit sans intérêt pour l'analyse ou que, bien que le test présente un lien significatif entre la variable de croisement et la variable d'intérêt, le chevauchement des intervalles de confiance ne permet pas d'observer de différences entre les modalités de la variable de croisement.

Dans le texte, une estimation suivie d'un astérisque (\*) indique une précision passable et suivie de deux astérisque (\*\*) une précision faible. Dans ce dernier cas, l'estimation est à utiliser avec circonspection et ne devrait être utilisée qu'à titre indicatif. Dans les tableaux, ces estimations sont exprimées sous la forme de cote D ou E respectivement. La signification de ces cotes est expliquée de façon plus détaillée à la section 10.5.1.

Un des objectifs de ce rapport étant de vérifier les différences existant entre les régions quant à la garde des enfants, plusieurs tableaux ont été construits en utilisant la variable de région administrative comme variable de croisement. L'extrait du tableau 1.1 présenté précédemment (tableau 10.5) en est un exemple. D'ailleurs, afin d'illustrer les différents concepts théoriques expliqués dans les paragraphes précédents, voici comment a été effectué l'analyse de ce tableau. Dans un premier temps, afin de vérifier s'il y a une association entre la variable de croisement (la région administrative) et la variable d'analyse (Question 9 sur l'organisation familiale), le résultat du test du Chi-carré est examiné. Comme le seuil observé du test (0,00) est inférieur à 0,05, une relation est observée entre les deux variables. Par la suite, comme une association a été détectée entre les deux variables, un examen est fait afin de voir s'il y a des différences entre les régions pour chacune des modalités de la variable d'analyse. Il y a différence si les intervalles de confiance autour des intervalles comparées ne se chevauchent pas. Par exemple, comme l'intervalle de confiance [14,4 %, 20,3 %] autour de l'estimation du pourcentage de familles monoparentales de la Mauricie (17,4 %) ne chevauchent pas les intervalles de confiance autour des estimations de familles monoparentales du Bas-Saint-Laurent [4,9 %, 8,4 %], du Saguenay-Lac-Saint-Jean [6,4 %, 10,7 %], de la Capitale-Nationale [7,7 %, 12,4 %], du Nord-du-Québec [6,1 %, 10,8 %] et de

la Chaudière-Appalaches [6,4 %, 10,2 %], on peut conclure que la proportion de familles monoparentales est plus grande en Mauricie que dans les régions mentionnées. Par contre, comme l'intervalle de confiance [10,8 %, 16,0 %] autour de l'estimation de familles monoparentales de la Gaspésie-Îles-de-la-Madeleine chevauche celui de la Mauricie, on ne peut conclure à une différence entre les deux estimations. De même, on ne pourra conclure de différence entre la proportion de familles monoparentales de la Gaspésie-Îles-de-la-Madeleine et celle du Nord-du-Québec (intervalle de confiance [6,1 %, 10,8 %] ), car les intervalles se touchent par l'extrémité (10,8 %).

Comme mentionné précédemment, il peut arriver lors de l'analyse qu'on détecte une association entre les deux variables (seuil observé du test inférieur à 0,05) mais qu'on ne soit pas en mesure d'observer de différences au niveau régional (la variable de croisement). Cela peut arriver si le seuil du test est près de 0,05 ou si l'analyse porte sur un petit nombre d'individus concernés. À ce moment, il sera indiqué dans le texte qu'il y a association entre la variable de croisement et la variable d'analyse mais qu'il serait hasardeux de dégager des différences compte tenu du recouvrement des intervalles de confiance. Il existe par contre une exception à cette règle et c'est lorsque les variables d'analyse et de croisement ont deux modalités chacune. À ce moment là, comme le test dénote une association et qu'on ne peut comparer qu'une modalité avec une autre, il y a forcément différence entre ces deux modalités.

#### **10.4.1.3 Analyses multivariées**

Les analyses multivariées ont pour but de définir les facteurs liés à différents aspects de l'utilisation des services de garde des enfants âgés de moins de cinq ans au 30 septembre 2004. Il s'agit de déterminer, parmi un ensemble de facteurs potentiels, lesquels sont les plus importants pour expliquer le fait d'être utilisateur ou non.

Contrairement à l'analyse bivariée, l'analyse multivariée permet de considérer simultanément plusieurs variables explicatives qui peuvent être corrélées entre elles. Il est ainsi possible d'étudier la relation entre l'utilisation des services de garde et une variable explicative en enlevant l'effet d'une ou de plusieurs autres variables explicatives sur cette relation. Par exemple, si la relation observée entre l'utilisation régulière des services de garde et le revenu de la mère est principalement due au fait que les mères qui travaillent ont un revenu plus élevé, le fait d'inclure le revenu et l'occupation de la mère simultanément dans le modèle permettra d'étudier la relation « nette » entre l'utilisation des services de garde et le revenu de la mère, une fois prise en compte l'occupation de la mère.

Par ailleurs, l'analyse multivariée permet, le cas échéant, de tenir compte de l'effet combiné de deux ou de plusieurs facteurs interagissant entre eux. Par exemple, si la relation entre l'utilisation régulière des services de garde et l'occupation de la mère varie selon l'occupation du père, on doit tenir compte de ces deux facteurs simultanément pour bien décrire le phénomène étudié. On est alors en présence d'une interaction entre l'occupation de la mère et l'occupation du père. Seule l'analyse multivariée permet d'étudier les interactions entre les variables explicatives.

Le modèle de régression logistique<sup>12</sup> est utilisé pour la recherche des facteurs liés à différents aspects de l'utilisation des services de garde des enfants de moins de cinq ans. Ce modèle statistique est approprié, puisque chacune des variables d'intérêt (aussi appelées « variables dépendantes ») prend deux modalités, soit l'utilisation et la non-utilisation des services de garde.

Le logiciel statistique SUDAAN a été utilisé pour effectuer la modélisation. Ce logiciel tient compte du plan de sondage dans l'estimation des paramètres ainsi que de leur variance. En particulier, SUDAAN tient compte de l'effet de grappe dû au fait que l'unité d'échantillonnage est une famille, tandis que l'unité d'analyse est un enfant.

Les objectifs particuliers ainsi que les résultats des analyses multivariées effectuées pour cette enquête apparaissent au chapitre 9. Afin de faciliter leur interprétation, des tableaux schématiques y sont présentés pour chacun des facteurs liés significativement à l'utilisation des services de garde. Ces tableaux permettent d'identifier les sous-populations d'enfants qui diffèrent entre elles quant à leur probabilité d'utilisation des services de garde<sup>13</sup>. Les symboles exprimés dans ce type de tableau indiquent une gradation des probabilités d'utilisation des services de garde (de -- à +). Chaque colonne présente les résultats des comparaisons deux à deux. Deux catégories affichant un même symbole ne se distinguent pas de façon significative l'une de l'autre quant à la probabilité d'utilisation des services de garde. Parfois, plus d'un symbole est utilisé pour exprimer cette probabilité relativement à une catégorie donnée. Dans un tel cas, on ne peut dire si la probabilité relative d'utilisation des services de garde de cette catégorie d'enfants est inférieure ou supérieure aux catégories représentées par l'un ou l'autre de ces symboles.

---

12. D. W. HOSMER et S. LEMESHOW (1989). *Applied Logistic Regression*, New York, John Wiley & Sons, 307 p.

13. Les catégories résiduelles comprenant les « autres possibilités » et qui présentent peu d'intérêt du point de vue des comparaisons deux à deux n'apparaissent pas dans les tableaux schématiques.

Le tableau 10.6 illustre la manière dont sont présentées, au chapitre 9, les probabilités relatives d'utilisation des services de garde. On peut y lire, par exemple, que la probabilité d'utilisation des services de garde est significativement plus élevée chez les enfants de famille monoparentale ou de famille biparentale dont le père est aux études (symbole unique + + pour ces deux catégories) que chez les autres enfants. Chez les familles biparentales restantes, la probabilité d'utilisation est plus élevée lorsque le père travaille selon un statut atypique d'emploi que lorsqu'il travaille selon un statut non atypique (symboles + et – uniques pour ces deux catégories). On n'a toutefois pas détecté d'écart significatif entre les enfants de famille biparentale dont le père ne travaille pas et n'étudie pas et les enfants de famille biparentale dont le père travaille. En effet, la catégorie « biparentale – père ni travail ni études » est représentée par les symboles – et + simultanément, et ces symboles sont les mêmes que ceux des deux catégories de familles biparentales où le père travaille.

**Tableau 10.6**  
**Exemple illustrant les probabilités relatives d'utilisation des services de garde selon une variable explicative donnée**

Biparentale – père travail non atypique	–	
Biparentale – père ni travail ni études	–	+
Biparentale – père travail atypique		+
Monoparentale – mère ou père	+ +	
Biparentale – père aux études		

#### 10.4.2 Traitement de la confidentialité

L'ISQ est tenu de protéger la confidentialité des renseignements fournis par les répondants. Les tableaux d'estimations ont été vérifiés afin de s'assurer de contrôler le risque de divulgation d'information au sujet des répondants. La règle utilisée pour déterminer les tableaux qui présentent un risque de divulgation est celle qui est liée au seuil, laquelle s'assure d'un nombre minimum de répondants dans une cellule d'un tableau d'estimations. Par la suite, des techniques de masquage permettant de réduire le risque de divulgation ont été employées dans un certain nombre de tableaux.

### 10.5 Évaluation statistique de l'enquête

Dans toute enquête statistique, les estimations produites sont entachées d'erreurs. D'une part, il y a les erreurs liées à l'échantillonnage, c'est-à-dire celles qui sont dues au fait qu'on n'interroge qu'une partie de la population observée. D'autre part, il y a les erreurs qui ne sont pas dues à l'échantillonnage, parmi lesquelles se trouvent celles qu'amènent un mauvais taux de réponse, une faiblesse de la base de sondage, des questions difficiles à interpréter, des erreurs de saisie, etc.

### 10.5.1 Erreurs dues à l'échantillonnage

Les erreurs dues à l'échantillonnage sont mesurées à l'aide de la marge d'erreur et du coefficient de variation.

La marge d'erreur (m.e.) associée à une estimation est en fait une mesure de la précision statistique de cette estimation. À un niveau de confiance de 95 %, elle se définit comme suit :

$$\text{m.e.} = 1,96 \times \sqrt{(\text{variance de l'estimation})}$$

À partir de la marge d'erreur, il est possible de définir un intervalle de confiance<sup>14</sup> (I.C.) à 95 %, associé à l'estimation :

$$\text{I.C.} = \text{estimation} \pm \text{m.e.}$$

Cet intervalle illustre l'étendue des valeurs possibles que peut prendre la variable étudiée dans la population observée. On l'interprète de la façon suivante : « Si l'échantillonnage était reproduit un très grand nombre de fois, chaque échantillon produisant son propre intervalle de confiance, alors 95 % des intervalles contiendraient la vraie valeur du paramètre estimé par l'enquête. »<sup>15</sup>

Lorsque l'approximation par la loi Normale n'est pas possible, c'est-à-dire lorsque la proportion est trop petite, compte tenu de la taille efficace associée, l'intervalle de confiance basé sur la loi Normale a été remplacé par un intervalle de confiance asymétrique basé sur la loi Binomiale. Dans le cas des estimations de totaux, seule l'approximation par la loi Normale a été utilisée.

Les détails de la construction des bornes de l'intervalle Binomial ne sont pas présentés dans ce document. Cependant, les conditions d'application de la loi Normale pour obtenir une proportion et une taille efficace, c'est-à-dire le nombre de répondants ajusté par un facteur tel que l'effet du plan de sondage, sont présentées au tableau 10.7.

14. La marge d'erreur est justifiée uniquement lorsque l'approximation par la loi Normale s'applique. Dans le cas d'une distribution asymétrique, comme la distribution Binomiale, la marge d'erreur ne correspond pas à la demi-longueur de l'intervalle de confiance.

15. Cochran, William G. Sampling techniques. John Wiley & Sons, New York, 3rd Edition, p.12.

**Tableau 10.7**  
**Conditions d'application de la loi Normale**

Proportion	Taille efficace
$p < 0,05$ ou $p > 0,95$	10 000
$0,05 \leq p < 0,10$ ou $0,90 < p \leq 0,95$	1400
$0,10 \leq p < 0,20$ ou $0,80 < p \leq 0,90$	600
$0,20 \leq p < 0,30$ ou $0,70 < p \leq 0,80$	200
$0,30 \leq p < 0,40$ ou $0,60 < p \leq 0,70$	80
$0,40 \leq p < 0,50$ ou $0,50 < p \leq 0,60$	50
$p = 0,5$	30

Le coefficient de variation (C.V.) est une mesure qui permet de quantifier la précision de l'estimation :

$$C.V. = \frac{\text{marge d'erreur}}{(1,96 \times \text{estimation})}$$

Cette mesure contribue à faciliter l'interprétation quant à la précision d'une estimation. Plus le C.V. est élevé, moins précise est l'estimation et vice versa. Notamment, il faut être très prudent lors de l'utilisation des estimations dont le C.V. est très élevé (supérieur à 25 %), ce qui indique une très grande variabilité de l'estimation. Ces estimations ne devraient être utilisées qu'à titre indicatif. Le C.V. a également l'avantage de permettre de comparer la précision de différentes estimations entre elles. Le tableau 10.8 présente la relation entre le C.V. et la précision de l'estimation utilisée par l'ISQ.

Les estimations effectuées pour cette enquête atteignent pleinement les objectifs de précision fixés initialement. En effet, la précision sur les erreurs maximales en ce qui a trait aux préférences des familles pour chaque groupe d'âge et pour chacune des subdivisions des régions administratives sont d'environ 1 % et 5 % respectivement.

**Tableau 10.8**  
**Relation entre le C.V. et la précision de l'estimation**

C.V.	Cote	Précision de l'estimation
$C.V. \leq 5 \%$	A	Excellente
$5 \% < C.V. \leq 10 \%$	B	Très bonne
$10 \% < C.V. \leq 15 \%$	C	Bonne
$15 \% < C.V. \leq 25 \%$	D	Passable
$C.V. > 25 \%$	E	Faible <sup>1</sup>

1. À utiliser avec circonspection.

## 10.5.2 Erreurs non dues à l'échantillonnage

Comme mentionné précédemment, les erreurs non dues à l'échantillonnage peuvent avoir plusieurs sources : un mauvais taux de réponse (global ou partiel), des questions difficiles à interpréter, des erreurs de saisie (ou d'interprétation de l'intervieweur) ou des fausses réponses données (volontairement ou non) par les répondants.

### 10.5.2.1 Taux de réponse globale et non-réponse partielle

La plus importante erreur non due à l'échantillonnage est celle que cause la non-réponse. En effet, la non-réponse peut induire des biais dans les résultats.

La non-réponse est totale lorsque, pour une raison ou une autre, un individu choisi n'a pu être interviewé. Le taux de réponse globale est un indicateur de la qualité de l'enquête, notamment en regard des biais dus à la non-réponse qui pourraient être introduits dans les estimations. Le taux pondéré de réponse globale obtenu pour l'enquête est de 68,9 %, ce qui est supérieur à l'hypothèse initiale de 65,0 %. Comme on le spécifie à la section 10.3.2 sur la non-réponse, deux techniques ont été mises en œuvre pour diminuer le biais causé par le nombre de non-réponses, notamment l'ajustement pour la non-réponse et la poststratification. Par ailleurs, le taux pondéré de réponse globale est bon et il n'y a pas lieu de craindre a priori une influence négative sur la fiabilité des estimations obtenues. Les taux pondérés de réponse par subdivision des régions administratives de résidence et par groupe d'âge sont aussi satisfaisants (ils sont tous supérieurs à 60 %); il n'y a donc aucune crainte à avoir de ce côté.

La non-réponse est partielle lorsque seulement une partie du questionnaire a été remplie. Il est habituel de considérer qu'un taux de non-réponse partielle inférieur à 5 % ne devrait pas susciter d'inquiétude. Toutefois, lorsque ce taux est supérieur, des mises en garde doivent être faites quant à la présence de biais. Une analyse de la non-réponse partielle a donc été effectuée. Elle consiste à évaluer la non-réponse pour chaque question, à cibler les cas problématiques et à déterminer l'importance de la non-réponse partielle. Notons que le taux de non-réponse à chaque question est obtenu par le rapport entre le nombre d'individus n'ayant pas répondu à la question et le nombre d'individus devant répondre à cette question (répondants potentiels).

La totalité des questions et presque tous les indices créés ont un taux de non-réponse partielle ne dépassant pas 5 %. Les seules questions ayant un taux de non-réponse partielle supérieur à 5 % sont,

comme on pouvait s'y attendre, celles qui concernent le revenu : revenu annuel familial (5,9 %) et revenu annuel personnel (18,5 %) de la mère (sur l'ensemble des familles visées par l'enquête). Le fort taux de non-réponse partielle pour ce qui est du revenu personnel de la mère se justifie par le fait que la question sur le revenu personnel s'adressait uniquement à la personne répondant au questionnaire, soit la personne la mieux placée pour répondre pour son ou ses enfants. Comme cette personne pouvait être le père – comme tel a été le cas 14,1% du temps –, la non-réponse en regard de la mère est en grande partie explicable.

Si aucun biais systématique ne semble présent dans les estimations en ce qui regarde le revenu familial, il faut se montrer plus prudent dans l'interprétation des données utilisant le revenu personnel de la mère. La non-réponse à cette variable est associée à quelques variables dont la région administrative, la zone géographique, l'organisation familiale, le lieu de naissance des conjoints et la principale occupation des conjoints et de la mère. Comme ces variables sont également explicatives du revenu personnel de la mère, la présence d'un biais est possible. Par contre, l'incidence de ce biais est difficile à évaluer, étant donné les effets changeants de chacune des variables en cause. Quelques tendances laissent toutefois percevoir une légère sous-estimation des mères n'ayant aucun revenu et une légère surestimation des mères ayant un revenu élevé.

Dans le cas des analyses multivariées, la non-réponse partielle observée pour chacune des variables explicatives incluses dans le modèle est considérée dans son ensemble. La démarche de modélisation adoptée, lorsque cette non-réponse partielle est importante, est présentée au chapitre 9.

De façon générale, outre le revenu personnel de la mère, la non-réponse partielle ne semble pas un problème important à l'égard de la présence de biais potentiels.

#### **10.5.2.2 Interprétation des questions**

Comme mentionné précédemment, même si les entrevues se sont bien déroulées, des explications ont dû être fournies aux répondants pour les accompagner tout au long du questionnaire. Par contre, comme la formation reçue par les intervieweurs était uniforme, rien ne laisse présager la présence d'un biais dans les estimations.

Cependant, une mise en garde doit être faite quant à la question au sujet du nombre de listes d'attente pour un service de garde à 7 \$ sur lequel l'enfant est inscrit. En effet, le nom de certains enfants peut encore apparaître sur des listes d'attente pour un service de garde à 7 \$, même s'ils ont obtenu une place à 7 \$. Afin d'atténuer cet effet, des estimations du nombre d'enfants sur au moins une liste d'attente ont été aussi faites en se limitant aux enfants qui ne disposent pas d'une place à 7 \$ de façon régulière pour tout motif ou de façon irrégulière pour le travail ou les études.

### **10.5.2.3 Erreurs de saisie**

Comme mentionné précédemment, la collecte des données a été faite à l'aide d'un logiciel de type ITAO (interview téléphonique assistée par ordinateur) pour les entrevues téléphoniques. Ce logiciel permettait d'effectuer les sauts de questions de façon automatique et de laisser répondre les répondants admissibles à une question. De plus, certains répondants, dont le questionnaire comportait des incohérences, ont été rappelés pour corriger certains renseignements. Ainsi, il n'y a pas lieu de craindre que des erreurs de saisie aient pu amener des biais dans les résultats.

### **10.5.2.4 Réponses volontairement fausses**

Il n'existe pas de moyen absolu de valider la véracité de toutes les réponses reçues. Cependant, l'intérêt marqué des parents pour les services de garde offerts à leurs enfants suggère que les réponses fournies sont véridiques.

## **10.6 Appréciation globale**

Les données de l'enquête sur les besoins et les préférences des familles en matière de services de garde doivent être considérées comme très bonnes. Les marges d'erreur associées aux différentes mesures répondent aux objectifs initiaux.

Le taux de réponse globale de l'enquête est très satisfaisant et dépasse même l'hypothèse de départ, les taux de réponse par strate sont également bons et dépassent, pour la plupart, l'objectif de 65 %, et la non-réponse partielle est négligeable. Au total, la présence de biais potentiels non dus à l'échantillonnage a été réduite au maximum.

Pour ces raisons, le potentiel analytique des données de l'enquête est excellent et les résultats sont tout à fait inférables aux familles québécoises et à leurs enfants de moins de cinq ans résidant au Québec. Toutefois, il y a lieu de faire preuve de prudence dans l'usage des résultats accompagnés d'une cote élevée. Ces cotes sont représentées par un D ou un E dans les tableaux du tome II et par \* ou \*\* dans l'analyse descriptive.