

Étude longitudinale du développement des enfants du Québec (ÉLDEQ)

Pondération longitudinale : volets 1998 à 2002

Direction Santé Québec
Institut de la Statistique du Québec
Octobre 2003

Répondants au volet 2002

Parmi les 2 120 familles répondantes au volet 1998 admissibles au suivi longitudinal, 1 944 ont répondu à nouveau à l'enquête au volet 2002. De ce nombre, 1 894 avaient également répondu à tous les volets depuis 1998, les 50 autres familles ont participé à quatre des cinq volets (tableau I). Bien que l'unité d'échantillonnage soit l'enfant cible, on parle ici de famille répondante au sens de famille au sein de laquelle vit l'enfant cible.

Tableau I – Nombre de répondants au QIRI (instrument principal)

Volet 1998	Volet 1999	Volet 2000	Volet 2001	Volet 2002	Nombre de répondants
oui	oui	oui	Oui	oui	1 894
oui	non	oui	Oui	oui	9
oui	oui	non	Oui	oui	15
oui	oui	oui	Non	oui	26 (n=1 944 en 2002)
oui	oui	oui	Oui	non	30
oui	oui	non	Oui	non	2
oui	oui	oui	Non	non	35
oui	oui	non	Non	non	43
oui	non	oui	Non	non	3
oui	non	non	Non	non	63 (n=2 120 en 1998)
non	non	non	non	non	697 (n=2 817 échantillon initial)

La pondération longitudinale doit tenir compte de la non-réponse globale aux volets 1998 à 2002. Il a été décidé de travailler à partir des 2 120 répondants du premier volet et d'ajuster les poids initiaux transversaux du volet 1998 pour tenir compte de la non-réponse survenue à l'un ou l'autre des volets 1999 à 2001, en supposant un mécanisme de non-réponse stable de 1999 à 2002. Cette façon de faire ne permet pas d'utiliser les données les plus récentes sur les ménages, cependant les variables utilisées en général sont stables dans le temps.

L'ajustement de la pondération pour la non-réponse est basé sur la création de classes de pondération. Comme aux volets précédents, la technique employée est la modélisation par segmentation fondée sur l'algorithme CHAID (« *Chi-square automatic interaction detection* »). Les classes de pondération sont créées sous forme d'arborescence; elles ne résultent donc pas nécessairement du croisement de toutes les variables considérées pour la modélisation. Pour une famille donnée, l'ajustement de la pondération consiste à diviser le poids de référence par la proportion de familles répondantes observée au sein de la classe à laquelle elle appartient.

Les variables considérées pour la modélisation sont essentiellement de nature socioéconomique. Des analyses exploratoires bivariées ont d'abord été faites afin de sélectionner un ensemble de variables pour la modélisation par segmentation.

Admissibilité à l'enquête au volet 2002

Parmi les 2 120 répondants au volet initial, on compte 24 familles ayant déménagé définitivement hors du Québec et une famille dont l'enfant cible est décédé entre les volets 1998 et 2002. Les 24 familles déménagées hors du Québec ne sont plus visées par l'enquête en ce sens qu'elles ne font plus partie de la population sur laquelle porte l'inférence. Ces familles sont donc considérées inadmissibles. La population visée est par conséquent composée des enfants qui sont demeurés au Québec entre les volets 1998 et 2002 ou qui ont quitté la province mais pourraient y revenir.

Pour ce qui est de la famille dont l'enfant cible est décédé, aux fins de la pondération et puisque les analyses portent uniquement sur les enfants survivants, cette famille sera considérée inadmissible. Les familles n'ayant pu être retracées, ayant refusé de répondre ou ayant été dans l'impossibilité de le faire sont toutes considérées admissibles à l'enquête. Bien que l'on sache que parmi les familles n'ayant pu être retracées, il pourrait y avoir des familles déménagées définitivement hors du Québec, leur nombre est trop petit pour que l'on en tienne compte dans la pondération. Sur cette base, l'échantillon admissible à l'enquête au volet 2002 est composé de 2 095 familles.

Taux de réponse

Le tableau II présente le nombre de répondants obtenus pour les différents instruments ou sections d'instruments de l'enquête à chacun des volets ainsi que pour l'ensemble des volets 1998 à 2002.

Tableau II - Nombre de répondants par instrument ou section d'instrument

	Volet 1998	Volet 1999	Volet 2000	Volet 2001	Volet 2002	Volets 1998 à 2002 ¹
QIRI et QPRI (fichiers SOCIO, ENFAN, PCM)	2 120	2 045	1 997	1 950	1 944	1 894
QIRI (section MÈRE) ²	2 119	2 042	1 990	1 936	1 930	1 874
QIRI (section PÈRE) ³	1 950	1 855	1 773	1 723	1 699	1 535
QIRI (section FRATRIE)	s/o	1 196	s/o	1 520	s/o	s/o
QAAM (fichier MÈRE)	2 051	1 966	1 934	1 897	1 877	1 719
QAAP ⁴ (fichier QAAP)	1 775	1 673 ⁵	1 616	1 519	1 475	1 204
QAAF (fichier FRAT)	s/o	s/o	s/o	1 488	s/o	s/o
QAAPABS ⁶ (fichier QPABS)	44	69	91	112	128	6

1. Nombre de répondants aux cinq volets.

2. Les nombres de répondantes sont égaux aux nombres de mères/conjointes présentes parmi les répondantes au QIRI, peu importe qu'il s'agisse ou non de la même personne à travers les années.

3. Les nombres de répondants sont égaux aux nombres de pères/conjoints présents parmi les répondants au QIRI, peu importe qu'il s'agisse ou non de la même personne à travers les années.

4. Le QAAP s'adresse aux pères/conjoints présents dans le ménage. Le fichier correspondant n'inclut que les répondants au QAAP.

5. Il a été décidé d'exclure du fichier QAAP initial (n=1 676) 3 conjoints non cohabitants (considérés inadmissibles au QAAP).

6. Le QAAPABS s'adresse aux pères biologiques non présents au sein du ménage mais ayant un contact avec l'enfant cible au moins une fois par mois. Le fichier correspondant n'inclut que les répondants au QAAPABS.

Le tableau III présente les taux de réponse pondérés pour l'instrument principal. Les taux de réponse longitudinaux s'obtiennent par le produit du taux de réponse pondéré du volet 1998 et de la proportion pondérée de répondants longitudinaux parmi les répondants du volet 1998 toujours admissibles à l'enquête au volet 2002. Le tableau IV présente quant à lui les taux de réponse longitudinaux obtenus aux instruments QAAM et QAAP. Le taux de réponse longitudinal au QAAM (QAAP) s'obtient par le produit du taux de réponse pondéré longitudinal au QIRI et de la proportion pondérée de répondants longitudinaux au QAAM (QAAP) parmi les répondants longitudinaux au QIRI et pour lesquels un parent/conjoint est présent au sein de la famille aux cinq volets.

Tableau III – Taux de réponse longitudinal pondéré au QIRI

	Proportion de répondants parmi les répondants au volet 1998	Taux de réponse relatif à l'échantillon initial
Volet 1998	s/o	75,3 %
Volets 1998 à 1999	96,5 %	72,6 %
Volets 1998 à 2000	95,0 %	70,9 %
Volets 1998 à 2001	93,0 %	69,2 %
Volets 1998 à 2002	92,8 %	69,0 %

Tableau IV – Taux de réponse pondéré au QAAM et au QAAP

	Proportion de répondants parmi les répondants au QIRI ¹					volets 1998 à 2002 ³	Taux de réponse longitudinal ² volets 1998 à 2002
	volet 1998	volet 1999	volet 2000	volet 2001	volet 2002		
QAAM	96,5 %	95,7 %	96,5 %	97,6 %	96,8 %	90,2 %	62,2 %
QAAP	90,1 %	88,9 %	90,3 %	87,4 %	85,4 %	76,8 %	53,0 %

1. Proportion de mères/conjointes (ou pères/conjoints) répondants au QAAM (QAAP) parmi les familles répondantes au QIRI et au sein desquelles un parent/conjoint est présent.

2. Taux de réponse aux volets 1998 à 2002 relativement à l'échantillon initial.

3. Proportion calculée relativement aux familles répondantes au QIRI aux volets 1998 à 2002 et au sein desquelles un parent/conjoint est présent aux cinq volets.

Analyse de la non-réponse au volet 2002 pour l'instrument QIRI

Pour tenir compte de la non-réponse au volet 2002, un ajustement a été fait à partir de la pondération transversale du QIRI du volet 1998 (PEPCM1). Cet ajustement est requis puisque les répondants au volet 2002 présentent des caractéristiques différentes des non-répondants. On minimise ainsi les risques de biais dus à la non-réponse dans les estimations qui seront produites. La nouvelle variable de pondération longitudinale pour le QIRI est appelée PEPCM5; elle est appropriée pour les analyses longitudinales impliquant des variables obtenues au QIRI aux volets 1998 à 2002. Plus précisément, la pondération PEPCM5 peut être utilisée pour analyser les variables du QIRI du volet 2002 en combinaison avec des variables du QIRI d'un ou plusieurs des volets précédents.

Bien que l'échantillon utilisé pour une analyse longitudinale impliquant des variables du QIRI des cinq premiers volets compte 1894 enfants, la pondération PEPCM5 porte sur les 1944 enfants ayant participé à l'enquête à quatre des cinq volets de 1998 à 2002. Les analyses impliquant des variables du QIRI des cinq volets doivent être pondérées à l'aide des poids PEPCM5, auquel cas

l'absence de données pour familles non répondantes à un des volets est considérée comme de la non-réponse partielle. Rappelons que l'inférence faite à partir des données pondérées des volets 1998 à 2002 porte sur les enfants nés au Québec en 1997-98 et qui n'avaient pas quitté la province de façon définitive.

Parmi l'ensemble des variables considérées pour l'ajustement pour la non-réponse au volet 2002, les variables suivantes ont été retenues : le statut d'immigrante de la mère, la région socio-sanitaire, le plus haut niveau de scolarité de la mère, la situation familiale de l'enfant à sa naissance, la langue parlée à la maison par les parents et le rang de naissance de l'enfant.

Analyse de la non-réponse longitudinale pour l'instrument QAAM

À partir de la pondération longitudinale du QIRI (PEPCM5), un ajustement pour la non-réponse longitudinale au QAAM a été fait pour les 1719 enfants dont la mère/conjointe a répondu au QAAM à l'ensemble des volets, peu importe qu'il s'agisse ou non de la même personne d'un volet à l'autre. On obtient ainsi une pondération longitudinale pour le QAAM permettant d'inférer aux enfants dont la mère/conjointe était présente au sein de la famille aux cinq volets. La nouvelle variable de pondération longitudinale pour le QAAM est appelée PEQAAML5. La population à laquelle on infère à partir de cette pondération représente 99 % de la population totale visée par l'enquête.

Parmi l'ensemble des variables considérées pour l'ajustement, les variables suivantes ont été retenues : revenu du ménage (revenu > 10 000 \$ aux cinq volets c. autres; revenu > 30 000 \$ aux cinq volets c. autres), langue parlée par la mère à la maison, nombre d'années dans le logement actuel et le nombre de personnes dans le ménage.

Analyse de la non-réponse longitudinale pour l'instrument QAAP

À partir de la pondération longitudinale du QIRI (PEPCM5), un ajustement pour la non-réponse longitudinale au QAAP a été fait pour les 1204 enfants dont le père/conjoint a répondu au QAAP à l'ensemble des volets, peu importe qu'il s'agisse ou non de la même personne d'un volet à l'autre. On obtient ainsi une pondération longitudinale pour le QAAP permettant d'inférer aux enfants dont le père/conjoint était présent au sein de la famille aux cinq volets. La nouvelle variable de pondération longitudinale pour le QAAP est appelée PEQAAPL5. La population à laquelle on infère à partir de cette pondération représente 81 % de la population totale visée par l'enquête.

Parmi l'ensemble des variables considérées dans l'ajustement pour la non-réponse longitudinale au QAAP, les variables suivantes ont été retenues : revenu du ménage (plus de 50 000\$ aux cinq volets c. autres), langue parlée à la maison par le père, le type de famille (biparentale intacte aux cinq volets c. autres), le statut d'immigrant du père et le plus haut niveau de scolarité du père.

Fichier de pondération

Le fichier SAS de poids longitudinaux pour les volets 1998 à 2002 est nommé POIDS501. Il contient les poids échantillonnaux suivants :

PEPCM5	poids échantillonnaux longitudinaux pour le QIRI des volets 1998 à 2002 (n=1944)
PEQAAML5	poids échantillonnaux longitudinaux pour le QAAM des volets 1998 à 2002 (n=1719)
PEQAAPL5	poids échantillonnaux longitudinaux pour le QAAP des volets 1998 à 2002 (n=1204)

Ces variables de pondération prennent une valeur nulle pour les enfants inadmissibles au suivi. Par exemple, les enfants dont le père/conjoint était absent à l'un des volets étaient inadmissibles au QAAP de ce volet et par conséquent inadmissibles au suivi longitudinal selon les variables du QAAP. Ils ont donc un poids PEQAAPL5 nul.

Choix de la pondération pour les analyses longitudinales des volets 1998 à 2002

Le choix d'une pondération appropriée dépend des variables considérées dans le modèle statistique. Le tableau V présente les poids à utiliser pour quelques situations générales qui devraient couvrir bon nombre d'analyses impliquant au moins une variable du volet 2002. La population sur laquelle porte l'inférence est également décrite. Comme il existe de nombreuses possibilités d'analyse, ce tableau ne saurait être exhaustif.

Pour une analyse longitudinale n'impliquant que des variables du QIRI (avec au moins une variable du volet 2002), les poids échantillonnaux PEPCM5 sont appropriés et ont été ajustés pour l'ensemble de la non-réponse au QIRI. Pour l'interprétation des résultats de ces analyses, il faut tenir compte de la non-réponse partielle possible à chacune des questions si elle est élevée car celle-ci pourrait entraîner certains biais dans les estimations. Il peut s'agir d'étudier dans quel sens va le biais en fonction des caractéristiques des non-répondants partiels.

Tableau V – Variables de pondération selon les instruments impliqués

Instruments dont proviennent les variables de modélisation	Pondération à utiliser	Population visée	Non-réponse sans ajustement¹
QIRI (un ou plusieurs volets)	PEPCM5	Ensemble des enfants visés par l'enquête	nil
QAAM 5 volets avec ou sans QIRI	PEQAAML5	Enfants dont la mère/ conjointe est présente au sein de la famille aux cinq volets (99 % de la population totale)	nil
QAAP 5 volets avec ou sans QIRI	PEQAAPL5	Enfants dont le père/ conjoint est présent au sein de la famille aux cinq volets (81 % de la population totale)	nil
QAAM (1 volet) avec ou sans QIRI	PEPCM5	Ensemble des enfants visés par l'enquête	2 à 4 %
QAAP (1 volet) avec ou sans QIRI	PEPCM5	Ensemble des enfants visés par l'enquête	10 à 15 %

1. Proportion de non-réponse globale à un instrument selon le volet, pour laquelle la pondération n'est pas ajustée (en plus de la non-réponse partielle à chacune des questions).

Une analyse longitudinale portant sur les variables du QAAM (QAAP) aux cinq volets simultanément nécessite l'utilisation des poids échantillonnaires PEQAAML5 (PEQAAPL5), peu importe si des variables du QIRI sont présentes dans les modèles. Ces poids ont été ajustés pour l'ensemble de la non-réponse longitudinale au QAAM (QAAP). Reste à tenir compte de la non-réponse partielle au sein des répondants longitudinaux au QAAM (QAAP). Supposons que l'on veuille faire de l'inférence sur les enfants dont le "père" et la "mère" sont présents aux cinq volets à partir des instruments QAAM et QAAP des cinq volets, les poids PEQAAPL5 seraient appropriés bien qu'ils ne soient pas ajustés pour la non-réponse au QAAM puisque celle-ci est minime (moins de 1 %).

Les analyses impliquant des variables du QAAP posent en général problème sauf pour le cas où les cinq volets sont considérés simultanément, situation pour laquelle la pondération PEQAAPL5 a été créée. En effet, dans la plupart des cas où l'ensemble des volets ne sont pas considérés pour le QAAP, les pondérations disponibles sont plus ou moins appropriées. D'une part, la pondération PEPCM5 est telle qu'une part non négligeable de la non-réponse n'est pas prise en compte. À l'opposé, la pondération PEQAAPL5 est ajustée pour l'ensemble de la non-réponse au QAAP aux cinq volets mais, puisque les enfants dont le père/conjoint n'a pas répondu au QAAP aux cinq volets ont un poids nul, une part non négligeable d'unités d'analyse sont perdues. Les analyses impliquant des variables du QAAM ne posent quant à elles pas de problème lorsqu'une pondération appropriée est utilisée (PEPCM5 ou PEQAAML5 selon le cas).

Effet de plan

Au volet 1998, un effet de plan moyen de 1,3 avait été estimé. Cet effet de plan avait été utilisé principalement pour faire des tests du khi-deux sur des tableaux de fréquences à deux dimensions à l'aide des logiciels courants qui ne permettent pas de tenir compte du plan de sondage complexe. Au volet 2000, cet effet de plan avait été réévalué et était également de l'ordre de 1,3 puisque la non-réponse au QIRI survenue entre les volets 1998 et 2000 était faible. Cet effet de plan n'a pas été réévalué pour le volet 2002 puisque l'on ne s'attend pas à un changement important, vu le taux réponse relativement élevé et stable au QIRI. On suggère donc d'utiliser ce même effet de plan estimé pour corriger la statistique du khi-deux ou pour l'estimation de la variance d'une proportion. De façon équivalente, on peut diviser les poids échantillonaires par l'effet de plan moyen.

Pour les modèles longitudinaux, il est suggéré de tenir compte de la pondération à l'aide des poids échantillonaires mais de ne pas corriger les statistiques des tests effectués par l'effet de plan moyen. Cette façon de procéder est telle que la variance des statistiques est en général sous-estimée. Ainsi, nous suggérons d'abaisser le seuil des tests à 0,01 pour éviter de déclarer significatives des relations qui ne le seraient pas. Il s'agit toutefois d'une approche conservatrice pour cette enquête-ci. Ainsi, les résultats d'intérêt qui ne sont pas significatifs mais pour lesquels le seuil observé est entre 0,01 et 0,05 pourraient être interprétés sous forme de tendance s'ils vont dans le sens attendu ou sont appuyés par d'autres travaux publiés. On peut pour ce faire mentionner un résultat avec nuance, en indiquant qu'on ne peut démontrer de lien significatif.