

Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2002)

Pondération longitudinale : volets 1998 à 2001

Ce document est tiré en très grande partie du travail réalisé par Nathalie Plante, Marie-Ève Tremblay et Robert Courtemanche de la Direction de la méthodologie, de la démographie et des enquêtes spéciales de l'Institut de la statistique du Québec, et a été mis à jour par Nathalie Vachon et Hadi Eid de la Direction Santé Québec.

Répondants au volet 2001

Parmi les 2 120 familles répondantes au volet 1998 admissibles au suivi longitudinal, 1 950 ont répondu à nouveau à l'enquête au volet 2001. De ce nombre, 1 924 avaient également répondu aux volets 1999 et 2000. Il y a 26 répondants au volet 2001 pour lesquels aucune information n'avait été recueillie soit au volet 1999, soit au volet 2000. Parmi les familles non répondantes à la fois aux volets 1999 et 2000, aucune n'a répondu au volet 2001 (tableau I). Notons que 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	nombre de répondants
oui	oui	oui	oui	1 924
oui	non	oui	oui	9
oui	oui	non	oui	17
oui	oui	oui	non	61
oui	non	oui	non	3
oui	oui	non	non	43
oui	non	non	non	63
non	non	non	non	697

Choix de l'année de base pour l'ajustement de la pondération en EA

La pondération longitudinale doit tenir compte de la non-réponse globale aux volets 1998 à 2001. Deux stratégies sont possibles, travailler à partir des 2 120 répondants du premier volet et 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 2001. La seconde stratégie est d'examiner la non-réponse au volet 2001 par rapport aux 1 997 répondants du volet 2000; on ajusterait dans ce cas les poids longitudinaux des volets 1998 à 2000 pour la non-réponse survenue au volet 2001 seulement. Bien que cette dernière façon de faire permette d'utiliser des informations plus

à jour, elle exclut les ménages non-répondants en 2000 et qui sont répondants en 2001, ces ménages ayant un poids longitudinal nul en E3 (17 ménages). En attendant de trouver une procédure acceptable du point de vue méthodologique, qui permettrait d'assigner des poids qui proviennent de périodes différentes aux non-répondants ponctuels, (non-répondants au volet précédant le volet traité), nous avons opté pour la première solution (ajuster les poids de 1998) qui nous permet de conserver tous les ménages répondants en 2001.

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. Pour les ajustements pour la non-réponse aux instruments QAAM et QAAP, quelques variables additionnelles ont été considérées, par exemple la perception de l'état de santé. 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 2001

Parmi les 2 120 répondants au volet initial, on compte 23 familles ayant déménagé définitivement hors du Québec et une famille dont l'enfant cible est décédé entre les volets 1998 et 2001. Les 23 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 seront considérées inadmissibles. La population visée est par conséquent composée des enfants qui sont demeurés au Québec entre les volets 1998 et 2001 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 2001 est composé de 2 096 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 2001.

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

	volet 1998	volet 1999	volet 2000	volet 2001	volets 1998 à 2001 ¹
QIRI et QPRI (fichiers SOCIO, ENFAN, PCM)	2 120	2 045	1 997	1 950	1 924
QIRI (section MÈRE) ²	2 119	2 042	1 990	1 936	1 909
QIRI (section PÈRE) ³	1 949	1 855	1 773	1 723	1 598
QIRI (section FRATRIE)	s/o	1 196	s/o	1 520	1 111
QAAM (fichier MÈRE)	2 051	1 966	1 934	1 897	1 766
QAAP ⁴ (fichier QAAP)	1 775	1 673 ⁵	1 616	1 518	1 311
QAAF (fichier FRAT)	s/o	s/o	s/o	1 488	s/o
QAAPABS ⁶ (fichier QPABS)	44	69	91	112	9

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 2001. 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 quatre volets.

¹ Nombre de répondants aux quatre volets.

² Les nombres de répondants sont égaux aux nombres de mères/conjointes présentes parmi les répondants 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.

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 2000	93,7 %	70,5 %
Volets 1998 à 2001	90,6 %	68,2 %

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

	Proportion de répondants parmi les répondants au QIRI ⁷					Taux de réponse longitudinal ⁸
	volet 1998	volet 1999	volet 2000	volet 2001	volets 1998 à 2001 ⁹	volets 1998 à 2001
QAAM	96,5 %	95,7 %	96,5 %	97,6 %	91,0 %	62,1 %
QAAP	90,1 %	88,9 %	90,3 %	87,4 %	80,9 %	55,2 %

Analyse de la non-réponse au volet 2001 pour l'instrument QIRI

Pour tenir compte de la non-réponse au volet 2001, 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 2001 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 PEPCM4; elle est appropriée pour les analyses longitudinales impliquant des variables obtenues au QIRI aux volets 1998 à 2001. Plus précisément, la pondération PEPCM4 peut être utilisée pour analyser les variables du QIRI du volet 2001 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 quatre premiers volets compte 1 924 enfants, la pondération PEPCM4 porte sur les 1 950 enfants ayant participé à l'enquête à trois des volets 1998, 1999, 2000 et 2001. Les analyses impliquant des variables du QIRI des quatre volets doivent être pondérées à l'aide des poids PEPCM4, auquel cas l'absence de données pour les neuf familles non répondantes au volet 1999 ou pour les 17 familles absentes au volet 2000, est considérée comme de la non-réponse partielle. Rappelons que l'inférence faite à partir des données pondérées des volets 1998 à 2001 porte sur les enfants nés au Québec en 1997-98 et qui n'avaient pas quitté la province de façon définitive.

⁷ 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.

⁸ Taux de réponse aux volets 1998 à 2001 relativement à l'échantillon initial.

⁹ Proportion calculée relativement aux familles répondantes au QIRI aux volets 1998 à 2001 et au sein desquelles un parent/conjoint est présent aux quatre volets.

Parmi l'ensemble des variables considérées pour l'ajustement pour la non-réponse au volet 2001, 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, le type de famille (famille monoparentale c. famille biparentale) 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 (PEPCM4), un ajustement pour la non-réponse longitudinale au QAAM a été fait pour les 1 766 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 quatre volets. La nouvelle variable de pondération longitudinale pour le QAAM est appelée PEQAAML4. 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 (revenu > 10 000 \$ aux quatre volets c. autres; revenu > 30 000 \$ aux quatre volets c. autres), perception de la mère de son état de santé (volet 2001), et présence du père biologique (volet 2000).

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

À partir de la pondération longitudinale du QIRI (PEPCM4), un ajustement pour la non-réponse longitudinale au QAAP a été fait pour les 1 311 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 quatre volets. La nouvelle variable de pondération longitudinale pour le QAAP est appelée PEQAAPL4. 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 : le niveau de suffisance de revenu (suffisant aux quatre volets c. insuffisant à au moins un volet), langue parlée à la maison par le père (volet 1998), le type de famille (biparentale intacte aux quatre volets c. autres), et le plus haut niveau de scolarité du père (volet 2001).

Fichier de pondération

Le fichier SAS de poids longitudinaux pour les volets 1998 à 2001 est nommé POIDS401. Il contient les poids échantillonnaires suivants :

PEPCM4 poids échantillonnages longitudinaux pour le QIRI des volets 1998 à 2001 (n=1950)

PEQAAML4 poids échantillonnages longitudinaux pour le QAAM des volets 1998 à 2001 (n=1766)

PEQAAPL4 poids échantillonnages longitudinaux pour le QAAP des volets 1998 à 2001 (n=1311)

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 PEQAAPL4 nul.

Choix de la pondération pour les analyses longitudinales des volets 1998 à 2001

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 2001. 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 2001), les poids échantillonnages PEPCM4 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 ajustement ¹⁰
QIRI (un ou plusieurs volets)	PEPCM4	Ensemble des enfants visés par l'enquête	nil
QAAM 4 volets avec ou sans QIRI	PEQAAML4	Enfants dont la mère/ conjointe est présente au sein de la famille aux quatre volets (99 % de la population totale)	nil
QAAP 4 volets avec ou sans QIRI	PEQAAPL4	Enfants dont le père/ conjoint est présent au sein de la famille aux quatre volets (81 % de la population totale)	nil
QAAM (1 volet) avec ou sans QIRI	PEPCM4	Ensemble des enfants visés par l'enquête	2 à 4 %
QAAP (1 volet) avec ou sans QIRI	PEPCM4	Ensemble des enfants visés par l'enquête	13 %

Une analyse longitudinale portant sur les variables du QAAM (QAAP) aux quatre volets simultanément nécessite l'utilisation des poids échantillonnaires PEQAAML4 (PEQAAPL4), 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 quatre volets à partir des instruments QAAM et QAAP des quatre volets, les poids PEQAAPL4 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 quatre volets sont considérés simultanément, situation pour laquelle la pondération PEQAAPL4 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 PEPCM4 est telle qu'une part non négligeable de la non-réponse n'est pas prise en compte. À l'opposé, la pondération PEQAAPL4 est ajustée pour l'ensemble de la non-réponse au QAAP aux quatre volets mais, puisque les enfants dont le père/conjoint n'a pas répondu au QAAP aux quatre 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 (PEPCM4 ou PEQAAML4 selon le cas).

¹⁰ Proportion de données manquantes (non-réponse globale à un instrument) pour laquelle la pondération n'a pas été ajustée, en sus de la non-réponse partielle à chacune des questions.

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 2 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 2001 puisque l'on ne s'attend pas à un changement important. 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 échantillonnaires par l'effet de plan moyen.

Pour les modèles longitudinaux, il est suggéré de tenir compte de la pondération à l'aide des poids échantillonnaires 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.